RUPRECHT-KARLS-UNIVERSITÄT-HEIDELBERG
PSYCHOLOGISCHES INSTITUT

Basic anesthesia skills simulation curriculum for medical students: Development and empirical evaluation based on an instructional design model

Anja Schmitz

Dissertation
zur Erlangung des akademischen Grades eines
Dr. phil.
der Fakultät für Verhaltens- und Empirische
Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg

August 2006

Erstgutachter: Prof. Dr. Niclas Schaper

Zweitgutachter: Prof. Dr. Karlheinz Sonntag

For my grandparents

"No matter how a training device is designed, no matter what its level of fidelity, it will not be an effective trainer if it is not used properly."

Hays & Singer, 1989, p. 42

Acknowledgements

I want to take this opportunity to thank everybody who was to some extent involved in the process of this dissertation.

First of all, I thank my advisor Prof. Dr. Niclas Schaper for his constructive feedback and guidance throughout all stages of this project. I am especially indebted to him for assuring that the project could be completed even after he took a position at the University of Paderborn.

I also greatly appreciate the help of Prof. Dr. Karlheinz Sonntag, who served as the second advisor. He provided me at the Psychology Institute with the necessary resources to complete this project.

This dissertation would furthermore not have been possible without Prof. Dr. Bernhard Graf, the HANS team and their support during the data collection process. Special debts are owed to Dr. Christoph Grube and Dr. Thomas Boeker, who contributed their time and expertise.

The study would have been even less possible without the medical school students, whom I want to thank for taking the time to fill out yet another set of questionnaires.

Thanks go out to my friends and colleagues, especially Christian Unkelbach and Christina Röcke who provided helpful hints and feedback on drafts of this paper and who kept me motivated throughout the process. I also appreciate Tetiana Stadnytska and Petra Hasselbach taking the time to discuss some methodological issues with me.

I particularly want to thank Oliver Emmler who kept me focused through the final stages of the project. He also provided very useful medical and IT helpdesk services to me.

Special thanks go to Dietmar Chur, Ute Fehr, and Petra Eggensperger (University of Heidelberg Competence Center), who challenged and supported me in my professional development and provided a lot of support on a personal level. Without the opportunity to work at the Competence Center I could not have financed this project.

Most of all, I want to thank my family for their enduring love and support.

Abstract

During the last decades medical school curricula have been reformed to put more focus on building students' practical skills. In the course of these reforms new teaching methods were employed, among them simulation. By now, simulation is a widely used training method, especially in the area of anesthesia. It currently still suffers from two sets of problems, namely the lack of a learning theoretical foundation and thorough evaluation studies. The current study revised an existing basic anesthesia skills simulation curriculum based on an instructional model for the development, delivery, and evaluation of medical simulation training (Schaper, Schmitz, Graf & Grube, 2003). The original version of the training (TG 0) was compared with two instructionally revised versions (TG 1 and 2). The revised curriculum was designed following an integrated approach to learning (Reinmann-Rothmeier & Mandl, 2001). Accordingly, the didactic strategy of the simulation training's preparatory seminar for TG 1 and 2 was revised to combine constructivist (e.g., problem-based), and traditional (e.g., lecture-based) elements. The full-scale simulation sessions for TG 1 and 2 were revised based on the cognitive apprentice-ship approach (Collins, Brown & Newman, 1989) to provide a constructivist learning environment with instructional support. In addition to these changes, the sequence of the curriculum's elements was varied for TG 2. Before participating in the full-scale complete-task simulation sessions, TG 2 participated in part-scale part-task emergency simulation training.

The evaluation of the revised training programs yielded the following main results: The three groups reacted positively to the training but did overall not differ in their reactions. The revisions of the preparatory seminar did not lead to the expected increase in self-reported learning. The revisions of the full-scale simulation sessions and the change in the curriculum's elements' sequence led to more self-reported learning for TG 2. Ratings by the instructors supported this result by showing that TG 2 was better prepared for the simulation sessions. On the transfer level, assessed by the students' OSCE results, TG 2 also received the best ratings. A set of students' individual characteristics and key relations between the model's elements were also investigated: Students' readiness and expectation fulfillment were positively influenced by respective interventions. Students' readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy predicted their training reactions. The study thus corroborated that simulation training is a viable method to build students' basic anesthesia skills. It provided first indications that a simulation curriculum based on a theoretically founded model improves training effectiveness and that individual characteristics influence simulation outcomes. Stricter designs are needed to extend the study's findings.

Contents

List of tables

List of figures

1 Introduction

Medical education reform

Physicians work in dynamic high-risk environments where they have to reach decisions and take actions within fractions of minutes that can decide over patient survival or death. A question that has consequently gained increased attention in the last decades is the question of how physicians can best be educated to deal with these demands.

Traditional undergraduate medical curricula were generally aimed at conveying conceptual knowledge and understanding of basic science and clinical contents in primarily lecture-based formats. Students' knowledge was tested with periodic multiple-choice tests (Albanese & Mitchell, 1993). The basic science part of the curriculum was most often followed by clinical clerkships. These were based on an apprenticeship model in which the apprentice encountered a wide range of clinical conditions over the course of many years (Kneebone, 2003). This instructional model implied that even basic clinical skills were, and oftentimes still are, usually trained on-the-job under the supervision of experienced physicians.

It became evident over the years, however, that the traditional medical school curricula suffered from a set of problems that manifested itself in medical graduates not really knowing how to treat patients and how to deal with the above mentioned demands, even after completing their clerkships (Flanagan, Nestel, & Joseph, 2004). Final-year medical students, for example, most of the time knew theoretically what the appropriate measures for an emergency situation were. The problem was, that they lacked the experience of actually applying this knowledge and were "surprised and stressed by how difficult it is to actually do it " (Flanagan et al., 2004, p. 56). The main shortcoming seemed to be that traditional medical school programs often provided only limited education in problem solving and basic clinical skills, such as examining or intubating patients (Albanese & Mitchell, 1993; Hayden & Panacek, 1999). This shortcoming is inherent in the traditional apprenticeship model as it can only gradually build the necessary skills for independent practice because the learning needs of the apprentice are always subordinate to the clinical needs of the patient (Kneebone, 2003). Traditionally, the practical experience has thus been acquired very late by the junior doctor with a real patient in a real (emergency) situation when the patient's life was oftentimes in danger. Another drawback of the traditional approach has been pointed out by Flanagan who remarked that "for the most part, clinicians are expected to learn most responses to routine and non-routine patient care situations by a series of random experiences. There is rarely the opportunity for structured reflection and feedback" (Flanagan et al., 2004, p. 59). These two characteristics of learning situations in which the learner's needs are always subordinate to patient care and / or in which learning experiences are random incidents render the apprenticeship model inefficient.

Medical schools consequently faced increasing societal pressure and budgetary constraints to enhance the quality of medical education and the safety of medical care. 'Learning by doing' has become less acceptable, especially in the light of emerging invasive procedures and high-risk care (Vozenilek, Huff, Reznek, & Gordon, 2004).

These problems and new requirements led to wide reforms of medical school curricula. The medical education reform recognized the need for students to be able to function as medical practitioners when they leave medical school (Bradley, 2006). The traditional curricula were consequently changed into problem-based curricula in which the teaching of clinical skills became an important element (Jünger, Schafer, Roth, Schellberg, Friedman, & Nikendei, 2005). With these reforms and their new goals, new teaching and learning methods needed to be introduced in the curricula (Albanese & Mitchell, 1993). Among them was simulation.

The use of simulators in medical education

In recent years, the use of simulators has risen markedly in the medical field. Simulators are used in several medical disciplines, such as anesthesiology, surgery, and nursing (Good, 2003; Morgan & Cleave-Hogg, 2002c; Murray, Boulet, Kras, Woodhouse, Cox, & McAlliseter, 2004). Based on research in the area of human factors and its similarity to other high-risk work environments, anesthesiology took a leading role in the use of simulators for educational purposes (Gaba, Fish, & Howard, 1994). Today, simulators are used for a variety of purposes such as training, assessment and research; as well as for a variety of topics, such as anesthesia crisis resource management (ACRM), advanced life support algorithms, airway management, technical and interpersonal skills training. The resulting simulations cover a wide spectrum of sophistication, from reproductions of simple tasks on isolated body parts to performing complete tasks on high-fidelity human patient simulators replicating the whole patient and his variable physiological parameters (Bradley, 2006).

With the use of simulators instead of real patients, the focus could be shifted towards the learner and his needs, the learning experiences were not random anymore, and the insufficient skills training in the area of healthcare could be addressed (Issenberg, McGaghie, Hart, Mayer, et al., 1999).

In addition to the described problems of traditional clinical teaching and the resulting need for new methods in the reformed curricula, the rise of simulation in medical education can largely be attributed to the advantages it offers for its use in educational settings. In addition to the key advantage that in full-scale simulators acute diagnosis can be integrated with the ongoing demands of managing and stabilizing a changing medical or surgical condition (Murray et al., 2004), the reasons for the use of *simulators in anesthesia* can be summarized in five interdependent categories (cf. Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005; Kneebone, 2003): *safety*, *controllability*, *efficiency*, *focus on training goals and deliberate practice*, and the *provision of additional instructional features* (see Schmitz, 2002, for a detailed discussion).

Based on these advantages, the medical community has by now recognized that "simulation based medical education can contribute considerably to improving medical care by boosting medical professionals' performance and enhancing patient safety" (Ziv, Ben-David, & Ziv, 2005, p. 193).

Problems with the use of simulators in medical education

With the increasing use of simulation training it became evident that it still suffers from two major problems: Simulation training lacks an appropriate instructional theoretical basis (Schaper, Schmitz, Graf, & Grube, 2003) and it lacks thorough evaluation studies (e.g., Bradley, 2006; Issenberg et al., 2005; Morgan & Cleave-Hogg, 2001; Weller, 2004).

The first problem, concerning the lack of valid, theoretically based, guidelines for the development of simulator training programs is not specific to the medical setting. "Precisely why simulation and simulators work is not well known. […] There is a somewhat misleading conclusion that simulation (in and of itself) leads to learning" (Salas & Cannon-Bowers, 2001, p. 484). Too often, training design is driven by the availability of technological innovations instead of by theoretical considerations that lead to providing learners with the optimal way to acquire the needed knowledge and skills (Ivancic & Hesketh, 2000, p. 1982). This leads to simulation and simulators being used without much consideration of findings about learning theory, training design, or training effectiveness. Consequently, "there is a growing need to incorporate the recent advances in training research into simulation design and practice" (Salas & Cannon-Bowers, 2001, p. 485). Especially, since the last decades have seen a great increase in training research producing theoretical advancements that by now provide "an organized framework in which systematic research could be couched" (Salas & Cannon-Bowers, 2001, p. 474).

Kneebone (2003) pointed out that the focus of simulation research is beginning to shift from simulation technology towards the embedment of simulation in education. The medical community thus has in parts already acknowledged that simulators are only valuable and simulation training's advantages can only be fully utilized if they are embedded in a complete educational curriculum grounded in sound learning theoretical approaches (Satava, 2001). Flanagan, Nestel, and Joseph (2004, p. 57), for example, state that "regardless of the modality, the simulator is only a teaching tool, which must be coupled with an effective curriculum." Although this fact has been recognized by the medical community, specific instructional models for medical simulation training are still mostly lacking as findings from the field of training research which by now offers sound, learning theory based principles and guidelines to instructors and instructional developers, have not been utilized (Salas & Cannon-Bowers, 2001, p. 473). Research showed that effective training requires a systematic approach to the assessment of needs, the design of programs, and the evaluation of outcomes in order to build the desired level of expertise (Salas & Cannon-Bowers, 2001). Unfortunately, this research has not yet had the necessary effect on the development, delivery, and evaluation of training in applied medical education settings. A first model developed to close this gap has been developed by Schaper, Schmitz, Graf, and Grube (2003). It specifies the learning theoretical basis for simulation training, guidelines for the design, delivery, and evaluation of simulation training, as well as contextual influence factors on training effectiveness. This model served as the basis for the current project.

The lack of thorough evaluation studies is the second problem in the area of medical simulation training. Without formal training evaluations guidelines to aid practitioners in making informed choices about training cannot be developed (Wexley & Latham, 1991). Increasing fiscal constraints and published university rankings furthermore increase the pressure on educators to optimize training programs. It thus becomes necessary to improve the knowledge of fostering and inhibiting factors of training effectiveness. It is important to understand that training effectiveness is a complex phenomenon that should be viewed from a systems perspective instead of merely focusing on instruction (Cannon-Bowers, Salas, Tannenbaum, & Mathieu, 1995). Only ongoing experimentation will allow optimizing patient simulator-based educational curricula "and additional studies are needed to determine the effectiveness of patient simulators in achieving learning and performance assessment objectives" (Good, 2003, p. 20). The lack of studies is especially grave in the area of simulation training transfer (Goldstein & Ford, 2002).

The effects of this lack of research have been poignantly summarized by Cooper and Taqueti (2004, p. i16): "While many [simulators] are in use, they are still the exception, not the rule, for healthcare education and training. Technical developments may be limiting, but the greater limitation derives from the model for reimbursement for healthcare and education in industrialised countries. The lack of research showing effectiveness, transfer of training to the clinical environment, and cost-effectiveness are also barriers to diffusion of the technology. It is not yet certain, that the next historical phase of simulation will achieve the acceptance and dramatic growth that are needed to lead a radical change in healthcare education and contribute substantially to patient safety. Collaboration and greater interdisciplinary research across simulation domains than has yet occurred could be synergistic and leverage the efforts of all."

The current project and its goals

The project presented in this dissertation was conducted as part of such an interdisciplinary research effort. Based on previous collaborations between the University Clinic of Anesthesiology and the Department of Industrial and Organization Psychology at the University of Heidelberg, a new project was initiated with a two-fold purpose.

Building on previous work (Schaper et al., 2003; Schmitz, 2002) the project aimed at tackling the two problems in medical simulation research mentioned above. The project was therefore designed to systematically develop a simulation-based introductory anesthesiology training program for medical students at the University of Heidelberg and to evaluate this training program. The training program and its evaluation were based on principles deducted from learning theory and training research that have not yet been systematically applied to the design of anesthesia simulator training (Schaper et al., 2003).

Structure of the dissertation

The dissertation is structured into four parts. The first part deals with the theoretical and empirical background of simulation training. It presents an overview of basic theoretical approaches to learning as the basis for the design of learning environments and then describes Schaper et al.'s (2003) model of simulator-based training as the theoretical foundation for the development of the simulator-based introductory anesthesiology training program. The first part ends with a review of empirical findings in the area of simulator-based anesthesiology training. The second part of the dissertation describes the conducted study. The study setting is introduced and the original anesthesiology curriculum presented. The focus of this part lies on the description of the revised anesthesiology curriculum's development based on the previously described model. This part of the dissertation ends with the illustration of the study's method. Part four presents the results of the study, partitioned into chapters on the study sample, training outcomes, individual difference variables, and relations between the different elements of the training model. The discussion of the previous results makes up the last part of the dissertation. It contains a reflection of the design of the training program, the summary and discussion of the evaluation study's results, a summary of the study's limitations, future research directions, and the conclusion.

2 Theoretical and empirical background

To put this study in the appropriate context, this first part of the dissertation starts out with describing general theoretical approaches to learning and instruction (2.1). On the basis of this context, the following chapter presents a model for designing, conducting, and evaluating effective simulator-based training programs (2.2). Chapter 2.3 then provides an overview of empirical findings in the area of simulator-based medical training programs, with the focus on anesthesiology training. This part of the dissertation ends with a conclusion and the deduction of the study's research question based on the theoretical and empirical background (2.4).

2.1 Theoretical approaches to learning and instruction

The following sections provide a brief introduction to two theoretical approaches to learning and instruction that have influenced the design of learning environments in recent decades: The cognitivist approach to learning (2.1.1) and the constructivist approach to learning (2.1.2).[1] Building upon these sections, section 2.1.3 integrates the two previously introduced approaches. At the end of the chapter the conclusion (2.1.4) delineates the relevancy of these approaches for the current study.

2.1.1 Cognitivist learning theories

To provide an overview over cognitivist learning theories, the general underlying concepts will be presented first (2.1.1.1), then a specific approach will be introduced as an example of cognitivist learning theories (2.1.1.2), and to conclude, cognitivist learning theories will be critiqued (2.1.1.3).

2.1.1.1 General concepts

Traditional cognitivist learning theories have focused on the individual's mental processes and acquisition of complex knowledge and skills (Shuell, 2001). Knowledge is perceived as objective and as an entity that can be transferred from the teacher to the learner with the outcome that the latter has gained the same knowledge and understanding as the teacher. Learning is thus seen as a change in an individual's knowledge, mental processes, skills, attitudes, and/or behavior. It results from elaboration, thinking, problem solving, and/or reflection processes (see Shuell, 2001, for a more detailed discussion). The learning process is thus conceived as following general rules which can be described in detail and thus also guided or influenced (Reinmann-Rothmeier & Mandl, 2001).

In the respective cognitivist educational paradigm the focus consequently lies on *instruction*, that is, on the teacher and his expository and didactic methods and on how to best convey

[1] The two views are contrasted here for the sake of differentiation and explanation. They are, however, in many cases and resulting applications not as pure as this contrast implies.

knowledge from the teacher to the students. Learning environments based on this approach are so called object- or content-oriented environments (Reinmann-Rothmeier & Mandl, 2001). Although there is a wide range of instructional approaches derived from this theoretical basis (Lowyck & Elen, 1991), a few core concepts can be extracted that characterize most of them. Among them are the following (cf. Reinmann-Rothmeier & Mandl, 2001):

As has been mentioned above, the primary focus is on *optimal instruction* and the resulting questions of how instructional sessions need to be planned, developed, structured, and managed to enable the learner to understand the material in the same way it was presented. The resulting learning outcomes can thus be exactly determined and assessed (Kemp, Morrison, & Ross, 1998).

The *instructor* assumes the *active role* of the didactic leader (Leinhardt, 1993) in the learning process. He systematically presents and explains the material in small steps, guides the learner, and consecutively evaluates the learner's performance.

The *learner*, on the other hand, assumes a *passive role* and merely acts in a receptive way. Since the material has already been optimally structured by the instructor, the learner is not required to restructure it (Lowyck & Elen, 1991). At the end of the learning process, he is expected to possess the identical objective knowledge the teacher has transported.

Evaluation of the learning outcomes is a central task of the instructor in the cognitivist approaches. Evaluation is conceptualized as the last step of the process and thus different from the instruction itself (Reinmann-Rothmeier & Mandl, 2001).

2.1.1.2 Instructional Design as a specific cognitivist approach

Cognitivist learning theory culminated in the Instructional Design (ID) approach that is still influential in American training design (e.g., Goldstein & Ford, 2002; Kemp et al., 1998; Lowyck & Elen, 1991; Reigeluth, 1999; Reigeluth & Stein, 1983; see also Sonntag, 2006a). Instructional Design can be viewed as a technology that aims at translating empirical findings into specific practical guidelines for instruction. The result of the ID process that systematically starts with a needs analysis, then specifies the objectives, and in the third step determines the necessary measures to reach the objectives, is a comprehensive *instructional design plan* that details for the instructor which strategies to use under which conditions (Snow, 1989). Whereas the original ID models, such as Bloom's Mastery Learning concept (Bloom, 1976), were strongly influenced by behaviorism, newer ID models have incorporated cognitive and even constructivist elements, such as learner-centered variables. They nevertheless still focus on general decision rules for the instructional program and on the specific sequencing of the content. An example of such a newer model is Reigeluth's elaboration theory (1979). Aside from the aforementioned training and instruction models, instructional design plans have been very influential in the development of computer-based learning programs.

2.1.1.3 Critique of the cognitivist approach

Cognitivist learning approaches suffer from a number of problems (cf. Deci & Ryan, 1993; Duffy & Jonassen, 1991; Patrick, 2003; Reinmann-Rothmeier & Mandl, 2001; Resnick, 1987; Winn, 1993). Although they claimed to develop curricula based solely on *empirical* findings, these findings are lacking for the most part. From a *theoretical* viewpoint, it seems especially problematic to promote understanding by breaking up complex learning material into very small parts. This ignores that understanding depends on being able to represent the complete knowledge structure. As Duffy and Jonassen (1991) pointed out, it is furthermore not possible to exactly predict the effect of specific learning methods. The decision rules contained in the instructional design plans for selecting the appropriate methods are thus not as useful as assumed. Furthermore problematic is the assumption that the learner acquires the identical knowledge representation the teacher tries to convey. Individual processes do not receive enough attention in the cognitivist approach.

A very problematic aspect from a *pragmatic* point of view is the learner's passive role that often leads to a loss of intrinsic motivation and self-regulation. With a look at today's often highly complex knowledge domains it has furthermore to be criticized that the systematic reduction of complexity during the learning process does not sufficiently prepare the learner for the respective application situations, thus leaving the acquired knowledge inert. The generation of inert knowledge has to be considered the biggest problem of cognitivist learning environments. They can, however, be very useful in the training of basic skills, where a deep insight or understanding is not that important (Winn, 1996).

2.1.2 Constructivist learning theory

In reaction to traditional behavioristic and cognitivist learning theories (cf. e.g., Reinmann-Rothmeier & Mandl, 2001; Shuell, 2001), a new approach to instruction dealing with the questions of how inert knowledge can be avoided, how learners can be motivated to actively participate, and how newly acquired knowledge can be linked to relevant behaviors (Reinmann-Rothmeier & Mandl, 2001; see also Krapp & Weidenmann, 2006), became influential during the 1980's: Constructivism[2].

The following sections first describe the underlying concepts of the constructivist learning framework (2.1.2.1), then present the anchored instruction approach as one example of constructivist learning environments (2.1.2.2), and will finally conclude by providing a critique of the constructivist approach (2.1.2.3).

[2] Constructivist ideas and paradigms are found in many different scientific disciplines. The current section focuses on constructivism as it relates to the questions of learning and instruction. See Gerstenmaier and Mandl (2001) for a discussion of the broader concept.

2.1.2.1 General concepts

Constructivist approaches view learning as a constructive, cumulative, self-regulated, goal-directed, situated, collaborative, and individually different process of meaning construction and knowledge building. Knowledge is not seen as a copy of reality, but as an individual construct (De Corte, 2001, p. 7570f). The constructivist approaches' common focus thus lies on the *constructive activity* of the learner during the learning process. The approaches differ, however, in the instructional guidelines that they deduct from this core concept (Winn, 1993).

One school of thought within constructivism is the *situated cognition movement* (e.g., Resnick, 1987). *Situated* refers to the concept that the interpretation and meaning of a situation cannot definitely or objectively be determined but depend on the context, the perspective, and the person (Sonntag & Schaper, 2001): The thought processes and actions of a person can only be understood within the context of their application which provide affordances and constraints. Consequently, learning is also considered situated (Sonntag & Schaper, 2001): The learning environment has to provide experiences for the learner in which he has the opportunity to construct knowledge in a *meaningful learning context*, that is, a context that is relevant for the learner's life or work. These learning environments have been termed *situated learning environments* (Reinmann-Rothmeier & Mandl, 1999). They have been designed to enable the learner to *comprehend* new material, to *apply* the acquired knowledge and skills flexibly, and to furthermore develop *problem-solving* skills (Reinmann-Rothmeier & Mandl, 1999) by providing the learner with opportunities to construct interpretations, achieve multiple perspectives, develop and defend their own point of view while recognizing that of others, and be able to influence the knowledge construction process itself (Knuth & Cunningham, 1993, p. 164). This implies that the focus shifts from optimizing teaching to optimizing learning processes with the emphasis on *generative* rather than passive learning (Cognition and Technology Group at Vanderbilt, 1993a). Whereas the learner has a very *active role* in these environments, the trainer or teacher has a reactive role: He provides learning situations and problem-solving tools and supports the learner at his request (Leinhardt, 1993). Concerning the evaluation of the learner's progress, the focus also shifted: Constructivist approaches focus much more on the *learning process* itself than on its final outcomes. It is thus not important to compare the results of the different learners with each other. The evaluation of the learning process is furthermore integrated in the instruction. This means that the learner receives feedback during the ongoing learning process to allow for corrections or support by the instructor. The learners are furthermore asked to engage in self-evaluation processes (e.g., Jones, 1992).

The constructivist learning framework comprises a great number of instructional approaches (e.g., Cognition and Technology Group at Vanderbilt, 1993b) and, as has been pointed out by Winn (1993), the opinions of constructivists on how instruction should be designed differ. Nevertheless, the following set of *design principles* seems to have emerged which – from a constructivist perspective – should be incorporated into the training design (Honebein, Duffy,

& Fishman, 1993; Law & Wong, 1996; Reinmann-Rothmeier & Mandl, 1999; Sonntag & Schaper, 2001; Sonntag, 1997, p. 347).

Authenticity and situatedness

Authentic tasks are among the most often emphasized features of constructivist learning environments. They should be embedded in realistic, meaningful, and complex situations of application so that the learner can acquire new knowledge and at the same time, the conditions for its application (Sonntag & Schaper, 2001). Providing authentic situations with corresponding realistic levels of complexity aims at motivating learners. Ideally, they should perceive the task as a challenge to develop a variety of appropriate solutions for and assume "ownership" of this process. Authentic contexts can furthermore make learning easier as many tasks become more intuitive in their regular context (Honebein et al., 1993). The complexity of the stimulus environment should therefore correspond to the level of complexity of the environment in which the learner is expected to perform after the training.

Honebein et al. (1993) emphasize that authenticity is a relative concept: "An activity is authentic to the extent that it captures the essential characteristics of that other activity" (p. 89). A critical feature of task authenticity is that the global task is represented with all of the respective sub-tasks as it is the global task which provides meaning for the sub-tasks. Constructivist environments that confront the learner with authentic and complex problems support the acquisition of applicable knowledge and are especially useful during the phases of advanced knowledge acquisition and expertise (Jonassen, Mayes, & McAleese, 1993; Sonntag & Schaper, 2001).

Contrary to the principle of authenticity, the situatedness principle does not necessarily call for the provision of complex contexts that mirror the complexity of the real situation. It merely requires embedding the problem in a broad context, as exemplified in the anchored instruction approach (Cognition and Technology Group at Vanderbilt, 1992). The so called anchor provides an open problem in a certain context (see 2.1.2.2).

Multiple contexts and perspectives

Constructivist learning environments aim at confronting the learner with multiple contexts in which to apply his knowledge. These multiple contexts are used to prevent the fixation of the new knowledge on a specific situation and to thus enable the learner to transfer his knowledge and skills to different kinds of problems and tasks. The principle of multiple contexts serves to decontextualize the learner's knowledge or to at least provide a context in the learning environment that "would be likely to match with the context of use" (Hammond, 1993, p. 61). Related to the principle of situatedness, it is furthermore important to use contexts that are meaningful to the learner.

To further support the construction of flexible knowledge, the learning environment should be set up so that the knowledge and skills can be acquired under multiple perspectives (Sonntag, 1997). The learning content should be considered from different perspectives, as this will allow for the flexible use of the learned skills and knowledge and promote critical thinking skills. Spiro and colleagues (1992) emphasize the principle of multiple perspectives as a central concept of constructivism and a specific characteristic of experts' knowledge bases: Experts are able to view problems from multiple perspectives and to connect these problems with other sets of information.

Social context

The used learning environments should allow for the cooperation among learners in social contexts (Sonntag & Schaper, 2001) as well as for their integration into the expert community (Sonntag, 1997). Cooperative learning and problem-solving supports generative learning as it promotes discussion, reciprocal explanation, and thus deeper understanding (Cognition and Technology Group at Vanderbilt, 1993a). Another advantage of cooperative learning is that it helps learners to develop the skills of generating and evaluating alternative perspectives, as is emphasized in the cognitive flexibility theory (Spiro et al., 1992).

These design principles are reflected in the different constructivist approaches, although with varying emphasis. An example is provided in the following section.

2.1.2.2 Anchored instruction as a specific constructivist approach

A specific constructivist approach is the anchored instruction approach developed by the Cognition and Technology Group at Vanderbilt University (CTGV) (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990; CTGV, 1990). It specifically aims at preventing the generation of merely inert knowledge.

Learning environments following this approach are designed around a so-called narrative "anchor". Anchors are narrations or descriptions of problem situations used at the beginning of the learning episode. They provide a focus and emphasize the meaning of the knowledge to be acquired through its application context. They are thus aimed at preventing the learner from merely memorizing facts or procedures that turn into inert knowledge. Instead, the acquired knowledge should be perceived by the learner as a tool for problem solving.

The following characteristics and design guidelines haven been specified for anchors (Bransford et al., 1990; Reinmann-Rothmeier & Mandl, 2001):

- The used anchors should have a *narrative format* to *provide intrinsic motivation* to as many of the learners as possible, for example through interesting stories, such as planning a trip.
- All of the information needed to solve the problem should be *embedded* in the anchor.
- Anchors should contain a *general goal* that can be reached through more *specific subgoals*.

- They must allow the learner to *identify* the *main characteristics* or relevant features of the problem situation so that he can conditionalize his knowledge appropriately.
- Anchors should be *complex and authentic* to support transfer. They should present meaningful contexts for the learners in order for them to activate their prior knowledge.
- Anchors should be presented through *multi-media applications*, for example videos. Traditional case-based learning environments have often been designed around paper-based anchors that convey information only verbally. The presentation of anchors through multi-media offers the advantage of increased realism and thus richer information that allows for more opportunities to acquire necessary pattern recognition skills to identify the triggering conditions necessary for conditionalizing knowledge. These presentation formats should allow the learners to develop an appropriate mental model of the problem situation and an intrinsic interest in solving the problem.
- Anchors should always be created in *pairs* so that the learners can view the problem from different perspectives and can flexibly apply their knowledge to different application contexts.

The CTGV's (1993a) example of anchors are the "Adventures of Jasper Woodbury", videos in which the main character is confronted with a situation that requires the application of mathematical principles. Older examples of anchors are case-based approaches used in Business Schools (Reinmann-Rothmeier & Mandl, 2001).

Initial studies (see Bransford et al., 1990; Cognition and Technology Group at Vanderbilt, 1992) showed encouraging results in different educational settings. Students learning with anchors were overall more likely to transfer the acquired knowledge, they could solve complex problems faster and in a more structured fashion.

2.1.2.3 Critique of the constructivist approach

Constructivist approaches to learning also face different sets of problems (Reinmann-Rothmeier & Mandl, 2001). The first set of problems concerns the lack of *empirical* evidence for the superiority of constructivist learning environments over cognitivist ones. There are not enough studies yet to provide a basis for a sound conclusion. The results of the studies that do exist (cf. Schmidt & Moust, 2000) showed that constructivist learning environments can lead to inferior results in post-test knowledge tests. Long-term studies that also include transfer tests in which the students have to apply their knowledge indicated, however, that constructivist learning environments were superior to traditional ones.

The second set of problems is of theoretical nature. The radical approach to constructivism denies the concept of "objective reality". It thus views realism in science as inadequate and implies that no objective knowledge base exists (Maturana & Varela, 1987). However, certain, and more or less “objective”, knowledge bases do exist within subject domains and have to be acquired by learners. This does not mean, however, that the representation of this knowledge is exactly the same for every learner or that every learner can acquire this knowledge in the same way. But the radical claim that all knowledge is a purely idiosyncratic cognitive construction does not seem helpful in educational settings, since students are required to pursue goals that are set for them by others. More moderate constructivists acknowledge that knowledge can be “objectified” and learning goals can thus be set and externally initiated. A problem that has, however, not been solved by even moderate constructivists is that there are most of the time only vague guidelines or concepts for the design of the learning environments. This might lead to the design of arbitrary and ineffective environments. Elaborations of how to actually use many of the constructivist elements, for example the self-evaluations, are still lacking, as is the according empirical evidence.

Moderate constructivists have concluded that “constructivist learning approaches and environments are not equally suitable for all learning contexts” (Jonassen et al., 1993, p. 233). Existing studies showed, for example, that the lack of support for the learners often leaves them overwhelmed and thus leads to poor results (Gräsel & Mandl, 1993). This seems specifically the case with learners who have insufficient learning prerequisites. Especially high are, for example, the self-regulation demands on the learners. Learners in constructivist learning environments have many more tasks in the area of self-regulation than they have in traditional learning environments. They are, among other aspects, responsible for the selection, sequence, and depth of their learning processes. This in itself already requires a great deal of motivation and skills on part of the learner that learners with insufficient prerequisites often don’t have (Winn, 1993).

The third set of problems is a *pragmatic* one. Constructivist learning environments often require an extensive amount of time and resources that cannot be justified in regular educational settings. Another pragmatic problem is that many learners do not accept constructivist learning environments. It does not seem plausible to them why they should independently “discover” certain problems or pieces of information that could also be taught by the instructor (Perkins & Cunningham, 1992).

Overall, purely constructivist learning environments seem more appropriate for advanced learners than for beginning students (Jonassen et al., 1993; Mandl, Gruber, & Renkl, 1997; Spiro & Jehng, 1990). They are better suited than traditional learning environments for acquiring complex skills, lead to less inert knowledge and better transfer. Studies with complex constructivist learning environments (Renkl, 1997; Stark, Graf, Renkl, Gruber, & Mandl, 1995) showed that those environments did indeed lead to better transfer results, but that learners benefited most, when they received additional instructional support.

2.1.3 Integration of cognitivist and constructivist approaches to learning and instruction

The following sections present the integration of the two approaches to learning and instruction described above (2.1.3.1) and introduce the concept of cognitive apprenticeship as a specific example for an integrated approach (2.1.3.2).

2.1.3.1 Knowledge-based constructivism – an integrated approach to learning

Constructivist learning approaches have led to a necessary paradigm shift from the cognitivist view of knowledge as an entity that can be transmitted from the teacher to the learner so that the latter can conceptualize and reproduce it in the same way the teacher can, to a view of knowledge as something the learner constructs through active and individual as well as collaborative processes.

In conclusion, it can be stated, that these approaches offer different views on the acquisition of knowledge, skills, and attitudes. Although they seem incompatible with one another, they can also be viewed as merely applying to different situations and often complementing one another (Shuell, 2001). Learning environment designers are thus asked to determine which elements are appropriate in the respective learning situations.

The most substantial advantage of constructivist learning environments is the advantage of generating less inert knowledge. This advantage is especially important in knowledge domains that are complex and dynamic, such as medicine. As has been mentioned above, the danger is that without any instructional support, learners can be overwhelmed and learning will be less effective. Research has furthermore indicated that it can be helpful to provide the learners with the necessary and appropriate knowledge base before confronting them with open and complex problem situations. Without an appropriate knowledge base, the “meaningful experiences” aimed for in constructivist learning environments might not be meaningful to the learners (Reinmann-Rothmeier & Mandl, 1999; Sonntag, 1997).

Effective learning thus still requires good teaching (De Corte, 2001). Based on current knowledge in the instructional literature, learning environments should thus initiate and foster constructive and self-regulated learning processes and instructional support should be given and gradually faded to enable students to take control of their own learning (De Corte, 2001).

To capitalize on the advantages of constructivist learning environments while minimizing the aforementioned problems and to incorporate the according research findings, a third approach to learning and instruction has been developed that integrates the two puristic positions of constructing and instructing (see Reinmann-Rothmeier & Mandl, 2001, for a detailed discussion). It has been referred to as *knowledge-based constructivism* (Reinmann-Rothmeier & Mandl, 2001). This integrated approach views learning as an active, self-directed, constructive, situated, and social process. A process that is only possible, however, if the learner possesses an appropriate knowledge base. Consequently, the position of the learner is a twofold one: His position is primarily active but also includes receptive phases. Integrated learning environments are thus based on constructivist learning principles but also include instruction. This form of instruction is primarily conceptualized as providing support for the learners and coaching them. In addition, instruction also means guiding the learners, and presenting and explaining content to generate a suitable knowledge base (Resnick, Williams, & Hall, 1998).

This integrated approach has been applied in the design of problem-oriented learning environments (see Reinmann-Rothmeier & Mandl, 2001) and also in the cognitive apprenticeship approach (2.1.3.2).

2.1.3.2 Cognitive apprenticeship as a specific integrated approach

As has been elaborated above, situated cognition considers learning as naturally tied to an authentic context, activity, and culture (Brown, Collins, & Duguid, 1989). The cognitive apprenticeship approach (Collins, Brown, & Newman, 1989) proposes a model of learning and a specific type of instruction based on the situated cognition theory, providing training designers with practical guidelines for applying the theory. The explicit use of instructional techniques and specific guidelines for these techniques makes the cognitive apprenticeship approach an integrated approach in which construction and instruction both have their place.

The approach resulted from applying principles used in traditional apprenticeship models in the crafts and arts to more cognitive domains. Honebein and colleagues (1993) point out that "the apprenticeship model, with its emphasis on embedding learning in a larger, functional context, is a model for instruction that captures the constructivist epistemology of learning and understanding" (p. 88). The cognitive apprenticeship approach further builds on the Vygotskian "zones of proximal development" in which students face tasks slightly more difficult than they can manage independently, requiring the aid of their peers and instructor to succeed (Belmont, 1989). The approach focuses on providing authentic tasks for the learner and introducing him into an "expert culture" through social interaction. Learners are confronted with the authentic tasks from the beginning so that they can develop a concept of what they need to learn during their training. The learning environments further steadily increase in complexity and include different contexts for the learner so that they learn to flexibly apply their knowledge to different problems. The enculturation process is supported by continuous interaction with experts in the field. This social interaction also serves to introduce the learner to different perspectives.

In contrast to purely constructivist approaches, the cognitive apprenticeship approach also incorporates support for the learner through a set of instructional techniques. This becomes evident in the guidelines proposed for the design of cognitive apprenticeship learning environments (cf. Reinmann-Rothmeier & Mandl, 2001):

Modeling
At the beginning of the learning episode, an expert models the appropriate behaviors by solving an authentic problem. The model should thereby specifically explicate cognitive processes that are not immediately visible in the real problem situation (verbalization).

Coaching
In the next step, the learner is faced with the real and thus oftentimes complex problem. In this situation, the instructor assumes the role of a coach and supports the learner in solving the problem. The instructor gives hints, feedback, and recommendations, and can also solve parts of the problem if necessary.

Scaffolding
If the learner is not able to independently solve the problem, the instructor supports him by giving hints. Scaffolding further implies that the support of the instructor is only temporary and adjustable. In another context, Campbell (1988) called for the adaption of training to differences in trainee aptitudes and prior knowledge. In a constructivist learning environment this can be accomplished by the scaffolding technique.

Fading
In the course of the learning process, the instructor gradually removes, that is, fades, his support when the learner does not need it anymore.

Articulation
Articulation refers to verbalizing knowledge and reasoning processes (Reinmann-Rothmeier, Mandl, & Prenzel, 1994). Through this method, the normally only internal processes are externalized and can thus be accessed in the reflection phase by the group of learners. Articulation thus also promotes the development of multiple perspectives through social exchange processes. Throughout the learning process, the learner is consequently encouraged to articulate his thoughts and problem-solving strategies.

Reflection
The learner is furthermore required to reflect upon his problem-solving processes. He is asked to discuss them with others. Reflection enhances the learner's metacognition. It thereby allows him to gain knowledge about his own cognitive functioning and to acquire self-regulation skills and strategies (De Corte, 2001). Reflection thus aims at supporting the development of general concepts that are at the same time applicable to many contexts.

Exploration
When the learner does not need the instructor's support anymore, he is encouraged to actively explore further problem settings and to thus independently identify and solve problems. He is

furthermore asked to consider new ways and contexts in which his knowledge and skills can be used. These are important strategies because in authentic situations failures to identify and access potentially relevant information often result in unsuccessful problem solutions.

The cognitive apprenticeship approach furthermore offers three principles for sequencing the learning content (De Corte, 2001): Increasing complexity, increasing diversity, and performing global before local skills. The latter principle means that learners should first concentrate on the complete context and the meaning of the task and only later focus on the subtasks or concepts.

Examples for medical learning environments designed in accordance with the cognitive apprenticeship approach are, for example, PlanAlyzer (Lyon et al., 1990), Thyroidea (Fischer, Gräsel, Mandl, Gärtner, & Scriba, 1994), and Derma 2000 (Arnold, 2003). They are computer-based learning environments for medical students that incorporate different constructivist and cognitive apprenticeship design principles.

Studies (Gräsel & Mandl, 1993) showed that computer-based programs offering only experts' comments were not sufficient to teach the students the required strategies. Only the implementation of further instructional support strategies based on the cognitive apprenticeship approach proved helpful for the development of appropriate diagnostic strategies (Gräsel, Mandl, Fischer, & Gärtner, 1994).
Evaluation studies conducted on cognitive apprenticeship programs showed that the students reacted positively to the programs (e.g., Arnold, 2003; Casey, 1996; Niegemann, 1995). Instruction based on the cognitive apprenticeship approach was also shown to lead to learning (Collins et al., 1989; Niegemann, 1995) and to more learning than the control groups that were not instructed using this approach (Scardamalia, Bereiter, & Steinbach, 1984). Studies furthermore showed that the cognitive apprenticeship approach also increased students' transfer in comparison to other forms of instruction (e.g., Gräsel & Mandl, 1993). Other studies (e.g., Hendricks, 2001) did not find an advantage of cognitive apprenticeship methods over other instructional methods concerning student transfer.

2.1.4 Conclusion

Although the different approaches to learning and instruction introduced above are not new developments, they still provide the theoretical basis for present endeavors in the design of learning environments.
Accordingly, they also provide the theoretical background for the present study. First of all, the simulator training model described in the next chapter (2.2) draws from these approaches in the areas of instructional systems design and training design and can be viewed as a type of instructional design model. As has been mentioned, training models are heavily influenced by

cognitivist approaches to learning in that they, for example, contain very specific guidelines for the development of instructor-set training objectives or the design of practice opportunities (e.g., Goldstein & Ford, 2002). These guidelines will be used in the present study.

Secondly, the training design interventions in the study will build upon the described learning approaches and their differing views on the acquisition of knowledge, skills, and abilities. As has been explained, cognitivist and constructivist approaches offer a set of advantages but also suffer from several limitations. The present study takes these advantages and limitations into consideration and strives to identify elements contained in these approaches that are adequate to promote learning processes in the area of simulation training. The integrated approach to learning will serve as the starting point for the design of the interventions as it combines the advantages of the cognitivist and the constructivist approach. All three approaches to learning are thus relevant for the study and their elements will emerge again in the methods part of the dissertation.

2.2 Theoretical approach to the instructionally-based development of simulator training: A model of simulator-training effectiveness (Schaper, Schmitz, Grube, & Graf, 2003)

Based on different learning theory approaches presented above as well as on organizational research, a set of training models has been developed that provide the theoretical basis for the development and evaluation of training programs in organizational settings. The models differ in their scope, their focus, and the settings they can be applied to. General training models are, for example, Cannon-Bowers, Salas, Tannenbaum, and Mathieu's (1995) longitudinal, systems-oriented model for the effective training systems design, specifying pretraining, during-training and post-training influences on the training's effectiveness, and Goldstein and Ford's model of a systematic approach to the development of training programs (2002), explicating relevant components of training systems in organizations. Other models, such as Kozlowski and Salas' model (1997), have a strong systems focus on the organization and its levels and are suitable for personnel development programs embedded in organizations. A narrower focus was set, for example, by Colquitt (2000) who developed a new an integrative model of training motivation, or Holton (2005), who developed a training evaluation model building upon the critique of Kirkpatrick's (1994) evaluation model. All of these training models offer more or less specific guidelines for the development, research, and evaluation of training programs. In a second step, these guidelines have to be adapted to the explicit training program or setting.

The anesthesia simulation training setting exists since mid 1980's when simulators started to be widely used in medical training (Kox, 2000). As has been mentioned in the introduction, since then, the design of these simulation training programs has, however, focused too much on the simulators' technical aspects instead of on the didactic principles or the theoretical foundation to enhance training effectiveness (Schaper et al., 2003; Van Emmerik & van Rooij, 1999).

In reaction to this lack of theoretical foundation Schaper, Schmitz, Graf, and Grube (2003) developed a model for the determination of simulation training effectiveness based on the training models mentioned above. This model is introduced in the following section (2.2.1) and its impact is reflected in the conclusion (2.2.2).

2.2.1 Description of the model

Based on their research experience in the are of Anesthesia Crisis Resource Management Schaper et al. (2003) developed a simulation training model that incorporates recent theoretical advances and empirical findings in the area of instruction and training and applies these findings to the specific area of medical simulation training. The model was developed to serve multiple purposes. First of all, it outlines a theoretical basis for medical simulation training incorporating theoretical findings from, for example, the areas of learning research and organizational research. Secondly, it functions as a practical guideline for (medical) training developers when they start to design new training programs. With the help of the model, training developers can identify important aspects that influence the effectiveness of their training program. Thirdly, the training model provides the basis for thorough evaluation studies as it explains how training evaluation is to be ingrained in the training design process.

The model's purposes are reflected by its structure. The training design process is located at the core of the model (see Figure 1) and highlights that both, the design of the simulator and scenarios as well as the learning design, affect the training outcomes and are themselves influenced by or derived from instructional systems design tasks. In these areas, theoretical findings from instructional psychology research are incorporated. The outer columns of the model show that individual as well as organizational factors furthermore determine the effectiveness of simulation training programs. These factors exert their influence both, before and after the training, and broaden the view of the training designer to also include or investigate findings from organizational and other research areas.

The following sections (2.2.1.1-2.2.1.6) provide an overview of Schaper et al.'s (2003) model (Figure 1). This overview focuses on the variables that are of specific importance for the present educational context of beginning anesthesia students. In this context, certain variables contained in the model, such as organizational culture, are not as important as they would be in contexts in which medical practitioners are trained and return to their jobs after the training. As this study is a training intervention study, the description of the model furthermore focuses on the model's components and not so much on their relationships with each other.

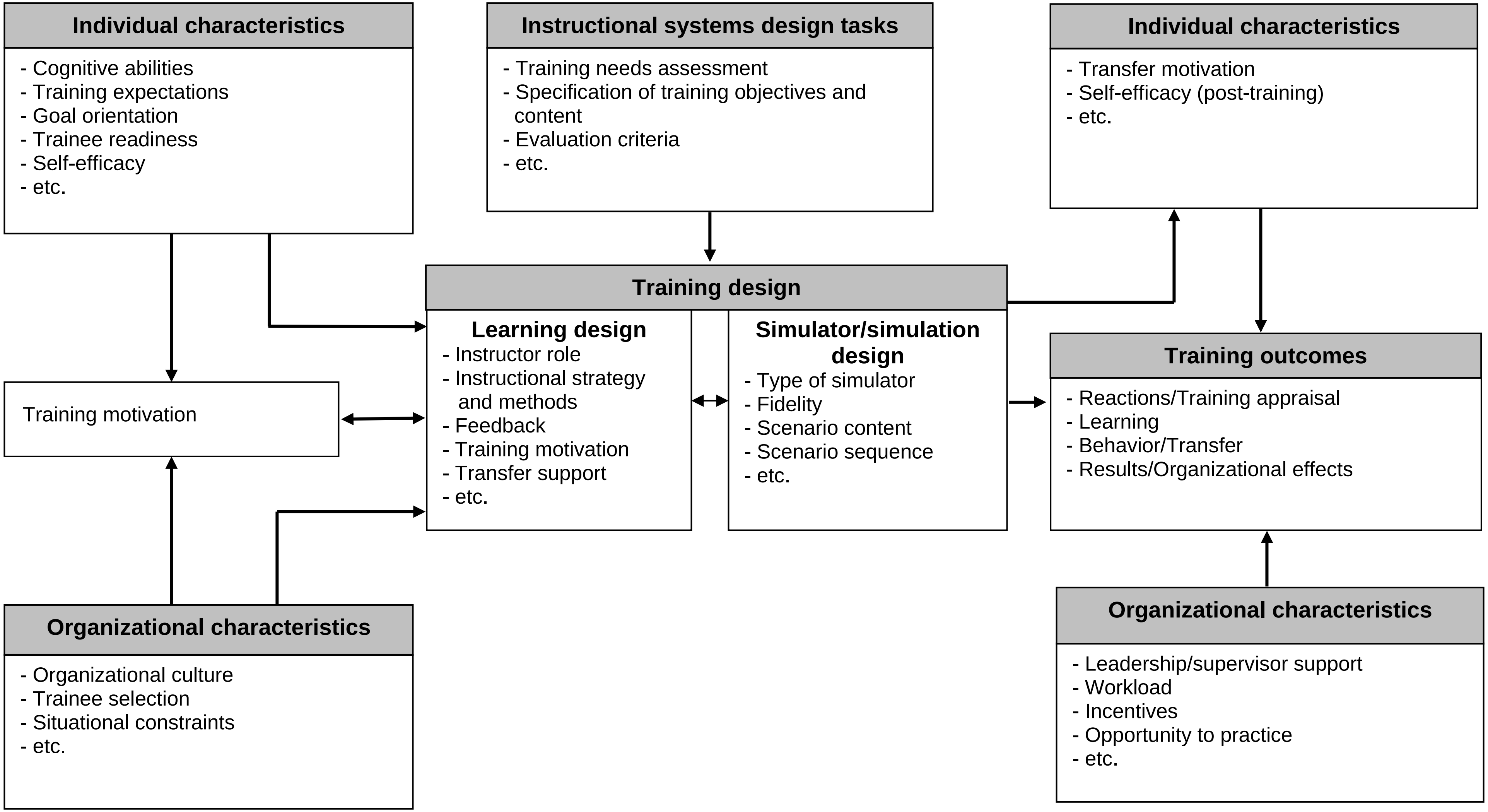

Figure 1. Model of simulator training effectiveness (after Schaper et al., 2003, p. 231).

As can be seen in Figure 1, the first column of the model specifies *individual* and *organizational characteristics* that influence the effectiveness of simulation training even before the onset of the training, such as the trainee's abilities and motivation. The center column of the model shows that the *design* of the *instructional systems* as well as the *design* of the *training* itself with *learning* and *simulator design* are of pivotal importance for the effects of the training. The right column of the model displays *individual* and *organizational characteristics* that influence the training's effectiveness after the training, namely, training transfer. The following sections provide an overview of these elements (cf. Schaper et al., 2003) focusing on the aspects relevant for the present study.

2.2.1.1 Individual characteristics

Learning and transfer as results of simulation trainings are influenced by a set of individual characteristics. Some of these individual characteristics are antecedent training conditions, others are post-training conditions (Figure 1). Among them are *cognitive ability*, personality traits, job attitudes, *goal orientation*, *pre*- and *post-training self-efficacy*, *trainee readiness*, and *training* and *transfer motivation*. *Cognitive abilities,* especially general intelligence and prior knowledge, have, for example, been shown to be important predictors of training success (see e.g., Mathieu, Tannenbaum, & Salas, 1992), although it is not yet clear how different levels of cognitive ability influence the learning process (Schmidt & Hunter, 1998; Sonnentag, Niessen, & Ohly, 2004).

The following sections provide more information on *training expectations* and their *fulfillment* (2.2.1.1.1), *goal orientation* (2.2.1.1.2), *trainee readiness* (2.2.1.1.3), *self-efficacy* (2.2.1.1.4), and *training and transfer motivation* (2.2.1.2), variables that have received a lot of attention in the training research area and have been identified as very influential in the training process. Other potential variables will not be considered in more detail as they were not included in the present study (see e.g., Colquitt et al., 2000; Goldstein & Ford, 2002; Holton, 1996; Tracey, Hinkin, Tannenbaum, & Mathieu, 2001, for a discussion beyond the scope of this dissertation).

2.2.1.1.1 Training expectations and expectation fulfillment

Training is often preceded by numerous preparatory organization events. In some cases, trainees might have spent a considerable amount of time researching and selecting a certain training program, in other cases, training might be mandatory and trainees might have received information about the training program from previous trainees or supervisors. The information trainees receive before the training has been shown to affect trainees' reactions to the training and their intention to transfer (see Tannenbaum & Yukl, 1992).

In all possible cases this implies that learners enter learning environments or training programs with differing expectations and desires regarding the learning event (Hoiberg &

Berry, 1978). These expectations can vary from emotionally toned expectations to expectations regarding the training content and the methods employed therein. Cannon-Bowers et al. (1995, p. 150) used training expectation scales assessing the aspects of a controlled learning environment, challenge, interactions with company members, and training methods.

Prior research (Cannon-Bowers et al., 1995, p. 160; Hicks & Klimoski, 1987) showed that realistic training expectations led to greater pretraining motivation and motivation to learn, greater commitment to attend, views of the workshop as more appropriate, and feelings of possible profit from the workshop.

Expectation fulfillment refers to the extent to which the training program or intervention meets or fulfills the trainees' expectations and desires (Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991, p. 759). When trainees' expectations are not met, that is, when expectation fulfillment is low, dysfunctional outcomes, such as poor training reactions are likely to occur (Tannenbaum et al., 1991). When expectation fulfillment is high, trainees perceive the training as more relevant and feel more positively about the training (Cannon-Bowers et al., 1995). Training expectation fulfillment has furthermore been assumed and shown to be positively related to post-training self-efficacy, motivation, attitudes results, and performance (Cannon-Bowers et al., 1995; Tannenbaum et al., 1991).

Based on these and similar findings, Cannon-Bowers et al. (1995, p. 161f) concluded that trainees should be led to have realistic expectations about the training.

2.2.1.1.2 Goal orientation

Goal orientation has been referred to as the individual's mental framework to interpret and behave in learning- or achievement-oriented activities (Dweck, 1986; Goldstein & Ford, 2002).

Two categories of goal orientation have been identified (Dweck, 1986; Dweck & Leggett, 1988): a) Mastery (or learning) goal orientation, whereby individuals strive to develop competence by acquiring new skills and mastering new tasks or situations. b) Performance goal orientation, whereby individuals seek assurances of their competence by looking for positive evaluations and trying to avoid negative ones.

It is currently still being discussed whether goal orientation is a disposition, a state, or both (Button, Mathieu, & Zajac, 1996); how they influence each other (e.g., Breland & Donovan, 2005; Stevens & Gist, 1997); and whether it is a multidimensional construct (e.g., VandeWalle, 1997), or whether the two categories are distinct or mutually exclusive (see Beaubien & Payne, 1999; Button et al., 1996; Lee, Tinsley, & Bobko, 2003; Steele-Johnson, Beauregard, Hoover, & Schmidt, 2000).

Aside from this conceptual debate, goal orientation has been shown to influence learning and performance (e.g., Bell & Kozlowski, 2002; Chen, Gully, Whiteman, & Kilcullen, 2000;

Diefendorff, 2004) and to provide motivation for performance (Cannon-Bowers, Rhodenizer, Salas, & Bowers, 1998). Several studies have provided evidence, for example, that learning goal orientation promotes learning, self-efficacy, persistence, and effort. They also showed that trainees with a learning goal orientation are more likely to pursue difficult tasks and persist when facing difficulties. Trainees with a performance orientation, on the other hand, are more likely to perceive the same tasks as threatening and to withdraw from them (Bell & Kozlowski, 2002; Boyle & Klimoski, 1995; Phillips & Gully, 1997; Sonnentag et al., 2004; Utman, 1997). Depending on the sample, sometimes both goal orientations were related to performance, sometimes only learning goal orientation (Steele-Johnson et al., 2000). Fisher and Ford (1998) showed, for example, that learning orientation strongly predicted knowledge-based learning outcomes. Consistent with previous research, Bell and Kozlowski (2002) found that learning orientation was positively related to knowledge and performance, and performance orientation showed negative relations with performance only. It is to be noted, however, that overall, the results are “far from unequivocal” (VandeWalle, Cron, & Slocum, 2001, p. 629). Performance goal orientation, for example, has been found to have negative (Ford, Smith, Weissbein, Gully, & Salas, 1998), nonsignificant (VandeWalle, Brown, Cron, & Slocum, 1999), and positive (Hoover, Steele-Johnson, Beauregard, & Schmidt, 1999) effects on task performance.

From a training intervention point of view, Cannon-Bowers and colleagues (1998) listed goal orientation as one of the pre-practice conditions that should be controlled or influenced to make “the time spent in practice more efficient and effective” (p. 293) and to thereby enhance training outcomes. They emphasize that a learning goal orientation is especially important during initial and intermediate phases of training (see also Brett & VandeWalle, 1999).

It has furthermore been pointed out that a learning goal orientation is especially useful when feedback-seeking behavior is needed to learn a new task or improve performance (VandeWalle & Cummings, 1997) and when transfer is required (Ford et al., 1998).

The presented research furthermore points out that goal orientation is a variable conceptualized as closely related to training motivation or as a motivational variable itself.

It can thus be concluded that a learning orientation tends to be related with more positive training outcomes, whereas a performance orientation is related with less positive training outcomes. Salas and Cannon-Bowers (2001) as well as VandeWalle and colleagues (2001), however, emphasize, that more research is needed to ascertain the specific effects of goal orientation on performance and to ascertain whether goal orientation is a relatively stable trait, or if it is a malleable state and can thus be influenced through training interventions.

2.2.1.1.3 Trainee readiness

Trainee or learner readiness refers to the “extent to which individuals are prepared to enter and participate in training” (Holton, 2005, p. 45). Trainee readiness has also been discussed as “trainability” (see e.g., Robertson & Downs, 1989) and is also often mentioned in the same line of argumentation as training expectations. The concept of trainee readiness, thus, is broad (Holton, 1996). It comprises, for example, the participants’ involvement in the needs assessment phase, in the planning of the training, their degree of choice, and the degree to which expectations are made explicit and met. Specifically the issue of clarifying and meeting trainees’ expectations has received research attention (Cannon-Bowers, Rhodenizer et al., 1998; Hicks & Klimoski, 1987; Tannenbaum et al., 1991). Trainees who do not think that the training will meet their needs and do not know what to expect will be less motivated and will learn less. The provision of preparatory information has been identified as a viable method to improve trainee readiness and was included in Cannon-Bowers and colleagues’ list of pre-practice conditions and measures (1998). Preparatory information “aids in setting the trainees’ expectations about the events and consequences of actions that are likely to occur in the training environment” (Cannon-Bowers, Rhodenizer et al., 1998, p. 305; Inzana, Driskell, Salas, & Johnston, 1996). It furthermore helps them to assess and cope with difficult tasks and is especially useful during the initial stages of skill acquisition. The concept has not yet been widely applied in the training setting, but more often in clinical settings to prepare clients for therapeutic interventions (Cannon-Bowers, Rhodenizer et al., 1998). An early study (Hilkey, Wilhelm, & Horne, 1982) showed, for example, that a pre-training video led to improved role-expectations and led to the elicitation of more adequate behavior in the first therapy session. Preparatory information thus enables trainees’ to make better use of the practice sessions by setting their expectations and informing them of what they will encounter.
Previous research has furthermore shown that the participants’ readiness to enter and to participate in the respective training program influenced their motivation to learn and led to improved training outcomes (e.g., Baldwin, Magjuka, & Loher, 1991; Hicks & Klimoski, 1987; Holton, 2005).

In conclusion, the evidence suggests that the participants’ readiness to enter and participate in the training program is an important variable in the area of training effectiveness. Its importance was corroborated by findings that highlighted the fact that practice is a complex concept and that simple task repetition may have only limited utility (Schmidt & Bjork, 1992).

2.2.1.1.4 Self-efficacy

Self-efficacy refers to the belief that one is able to mobilize one's cognitive resources, motivation, and courses of action necessary to perform specific tasks or behaviors (Bandura, 1977; Gist, Stevens, & Bavetta, 1991; Salas & Cannon-Bowers, 2001) and is a key construct in social learning theory (Bandura, 1997). The self-efficacy construct contains not only cognitive expectancies concerning future behavior, but also emotional and motivational aspects (Bandura, 1997). Bandura (1977) differentiated three dimensions of self-efficacy: magnitude, which refers to the extent of self-efficacy or the task difficulty the individual believes to master; strength, which refers to the belief's strength that the respective task difficulty can be mastered; and generality, which refers to the extent to which self-efficacy beliefs are generalized across situations.

Building upon the initial task- or situation-specific conceptualization of self-efficacy by Bandura (1977), self-efficacy has been differentiated along the lines of its generality in task, domain, and general self-efficacy (Schwarzer, 1994; Woodruff & Cashman, 1993).

Self-efficacy can be considered an individual difference variable that influences training motivation (Goldstein & Ford, 2002) but can also be viewed as a component of training motivation (see 2.2.1.2.1).

Self-efficacy beliefs result from the individual's learning experiences (personal and vicarious experiences), belief systems, and self-perceptions (Bandura, 1997).

The body of research on self-efficacy produced consistent results over the last years, among them the following. Recent meta-analyses (Multon, Brown, & Lent, 1991; Robbins et al., 2004) showed that academic self-efficacy was among the best predictors for college students' grade point average, academic performance, and persistence. In the training area, general self-efficacy has been shown to be related to training outcomes (Gist et al., 1991), as has task-specific self-efficacy (Chen et al., 2000). Self-efficacy was positively related to training expectations and reactions (Cannon-Bowers et al., 1995, p. 155; Mathieu et al., 1992). This means that trainees who believe they will perform well, want more from the training and react more favorably to it.

Self-efficacy also leads to better learning and performance, whether it is acquired before or during training (e.g., Barling & Beattie, 1983; Cannon-Bowers et al., 1995, p. 155; Gist, Schwoerer, & Rosen, 1989; Holladay & Quiñones, 2003; Mathieu, Martineau, & Tannenbaum, 1993; Mathieu et al., 1992; see also Salas & Cannon-Bowers, 2001). It has, for example, been found to be an important antecedent of training effectiveness as trainees with high levels of self-efficacy tend to learn more and perform better than trainees with lower levels of self-efficacy (Cannon-Bowers et al., 1995; Kozlowski et al., 2001; Tannenbaum & Yukl, 1992). Trainees' (post-training) *self-efficacy* also positively influences transfer motivation and transfer (Ford et al., 1998; Gist, 1997; Locke & Latham, 1990; Stevens & Gist, 1997; Tannenbaum et al., 1991). Higher self-efficacy levels after training should increase transfer because trainees should be more resilient when encountering transfer

obstacles (Marx, 1982). Fostering high levels of self-efficacy is thus especially important when training complex skills.
Self-efficacy has furthermore been shown to have motivational effects (Quinones, 1995), that is, it was positively related to pretraining motivation (Cannon-Bowers et al., 1995).

In conclusion, self-efficacy can be seen as a direct and indirect antecedent for training success, as an important process variable, and as a desirable training result (Tannenbaum & Yukl, 1992). It has furthermore been recommended as an operationalization of training motivation (Mathieu & Martineau, 1997).

2.2.1.2 Training and transfer motivation

Positive training outcomes will only occur if trainees possess both the ability and the volition to acquire and apply new knowledge, attitudes, and skills (KSAs) (Noe, 1986; Wexley & Latham, 1991). Training motivation has thus been established as a very important factor for training outcomes, i.e., reactions, learning, and transfer (Colquitt et al., 2000).
Motivation is a broad area of research. It has to be noted that motivation research often remains vague in defining the respective constructs investigated or in explicating the operationalizations of the construct (e.g., Colquitt et al., 2000). A review of the respective research indicates that training motivation is most often used as a comprehensive construct that includes other motivational processes and constructs, such as self-efficacy or training expectations.
In the training context, motivation has traditionally been conceptualized as motivation to learn and motivation to transfer (Goldstein & Ford, 2002, cf. Figure 1). Furthermore, training motivation is on the one hand a prerequisite individual characteristic necessary for learning to take place (Colquitt et al., 2000; Salas & Cannon-Bowers, 2001). On the other hand, motivation can also be fostered or hindered by the design of training programs. The latter aspect will be covered in section 2.2.1.5.1.4. This section focuses on training motivation as a prerequisite individual characteristic and transfer motivation as a post-training individual characteristic.
Several definitions of (training) motivation have been offered (e.g., Kanfer, 1991; Mitchell, 1982; Steers & Porter, 1975). According to these authors, training motivation can be defined as the direction, effort, intensity, and persistence with which trainees engage in learning activities before, during, and after training (Kanfer, 1991). It thus has arousal, energizing, directing, and maintenance components.
Motivation to learn has furthermore been defined as “a specific desire of the trainee to learn the content of the training program” (Noe, 1986, p. 743). According motivation measures investigate participants’ enthusiasm for learning and persistence under difficult conditions.
Motivation to transfer has been defined as the “trainees’ desire to use the knowledge and skills mastered in the training program on the job” (Noe, 1986, p. 743). Concerning the

conditions in which trainees are likely to be motivated to transfer, Noe (1986) mentions, among others, their confidence in using the skills, awareness of situations in which to use the new skills, and the belief that the knowledge and skills learned are useful in solving work-related problems and in satisfying job demands. This conceptualization of transfer focuses on transfer to the workplace as the only application setting. For the context of the current study, the scope of transfer has to be widened to also include settings that call for the application of the acquired KSAs but are not yet directly job-related. An example would be a performance-oriented examination setting (see also 2.2.1.6.3).

As the model shows (see Figure 1), training motivation is assumed to have various relations to other variables in the model. It is itself influenced by other individual characteristics, some of which are closely related to motivation, by organizational variables, and by training aspects, such as training method and content. Before the training program, the participants' self-efficacy beliefs, their job-involvement and organizational commitment, their career planning, personality variables, such as conscientiousness, their age, and organizational culture aspects, are all important determinants of training motivation (see Colquitt et al., 2000). The mere listing of these factors exemplifies that training motivation is a complex construct that is dependent on many of these factors and thus only partially under the control of the training designer.

Several models and predictors of training motivation have been proposed and examined (e.g., Baldwin et al., 1991; Facteau, Dobbins, Russel, Ladd, & Kudisch, 1995; Mathieu et al., 1992). The first model was presented by Noe (1986). Building on this model and the results of a meta-analysis, a recent one was proposed by Colquitt, LePine, and Noe (2000) who provided a first step toward an integrative training motivation theory. Both models specify antecedent and outcome variables of motivation in training settings. Among these variables are pretraining self-efficacy, valences, job involvement, personality variables, such as conscientiousness, as well as learning outcomes.

Concerning the influence variables of transfer motivation, Holton (1996; 2005) specified the following four categories: intervention fulfillment, learning outcomes, job attitudes, and transfer conditions. It is expected that trainees whose expectations were met and whose need for performance-related learning was fulfilled, trainees, who have learned more, trainees with more positive job attitudes, and trainees who work in more positive transfer conditions, are more motivated to transfer their skills to the job.

Research, both, based on these models and used to develop these models, provided evidence that training motivation is related to several training outcomes as well as other training variables. It could be shown that training motivation affects participants' KSA acquisition, retention, and willingness to transfer learned KSAs to the job (e.g., Mathieu et al., 1992; Noe, 1986; Quinones, 1995; Tannenbaum & Yukl, 1992). Other empirical findings show correlations between motivation to learn and learning (e.g., Hicks, 1984; Mathieu, Tannenbaum, & Salas, 1990; Tannenbaum et al., 1991) as well as relations between

motivation and training completion (Mobley, Hand, Baker, & Meglina, 1979). Pretraining motivation has been shown to be positively related to training expectations, self-efficacy, and performance (Cannon-Bowers et al., 1995, p. 156; Chiaburu & Tekleab, 2005).
Other studies (see Noe & Schmitt, 1986) didn't find relations between pretraining motivation and learning, behavior change or motivation to transfer.

Based on these general considerations, the following section (2.2.1.2.1) provides a brief overview of the valence-instrumentality-expectancy approach (Vroom, 1964) that can be used as a basis to conceptualize motivation in training research contexts (Yamnill & McLean, 2001).

2.2.1.2.1 Valence-Instrumentality-Expectancy

Among the different motivational theories (see Goldstein & Ford, 2002) an especially useful framework for investigating motivation in the training and transfer context is provided by the valence-instrumentality-expectancy theory (Lawler & Suttle, 1973; Vroom, 1964). Wexley and Latham (1991) as well as Noe (1986) and Colquitt et al. (2000) recommended the use of the valence-instrumentality-expectancy framework for studying training motivation (see also Lawler & Suttle, 1973, for an early review of findings). It is especially suited for the training context because Vroom (1964) developed it to apply to the work environment. The theory assumes that trainees' decision to attend the training, to expend learning effort, and to persist in attempting to use what they have learned, depends on the training's perceived utility for attaining desired outcomes.
Vroom's expectancy theory (1964) and its derivatives (e.g., Lawler & Suttle, 1973) are process theories of motivation linked to the question of how behavior is energized and sustained. At the center of the theory are an individual's cognitive expectancies that he can acquire a certain skill (expectancy) and that this skilled behavior will result in specific intrinsic and extrinsic outcomes (instrumentality) and his preferences among those outcomes (valence). An outcome's valence is determined by the valences of all other higher-level outcomes and the first-level outcome's instrumentality for attainting these higher-level outcomes. First-level outcomes are the direct result of a certain behavior, such as a certain degree of training performance. These outcomes attain their valence through their instrumentality for securing higher-level outcomes, such as passing exams or receiving a promotion. Higher-level outcomes may have a valence in and of themselves or they have it because they are instrumental for obtaining still other outcomes (Lawler & Suttle, 1973). The relationship between outcomes thus is one of instrumentality: "the relationship between an outcome (performing well) and another outcome (a reward such as pay) is an instrumentality that affects the valence of the original outcome" (Lawler, 1973, p. 47). It is to be noted, however, that the terminology within the expectancy theory framework often remains unclear. Acts and outcomes, for example, cannot always be clearly distinguished from each other in

the expectancy theory framework (Lawler, 1973). Expectancy has furthermore been differentiated (Lawler & Suttle, 1973) into expectations concerning the likelihood that effort will lead to mastery of the training content (expectancy I) and expectations concerning the likelihood that good training performance will lead to desirable outcomes (expectancy II). Noe (1986) pointed out that trainees' motivation to learn is likely influenced by their expectancies in regard to effort-performance and performance-outcome relationships.

Both, expectancy and valence, can differ. The degree to which an individual is motivated is thus dependent on his belief that he can achieve certain outcomes as a result of his behavior and on the value these outcomes have for him. In training settings Vroom's (1964) concepts of effort-performance and performance-outcome perceptions as causes of behavior seem especially useful operationalizations of motivation. Expectancy theory states that trainees' motivation will increase if they believe that their training effort will lead to increased performance. Training programs have a valence value for trainees if they are convinced that the training will allow them to achieve other outcomes, such as good grades, a promotion, or a job offer. An implication for the training designer is that he needs to communicate the training's value in regard to other outcomes to the trainees.

Related to the expectancy framework is also the self-efficacy concept (cf. 2.2.1.1.4). Self-efficacy conceptualizes only the expectancy I component of the expectancy-valence model. It can thus be considered as a necessary but not sufficient prerequisite of motivation in general.

Research findings over the years have provided evidence for the importance of the valence-instrumentality-expectancy perspective on training motivation (e.g., Baldwin & Ford, 1988; Mathieu et al., 1992; Tannenbaum et al., 1991). In their training model and its empirical validation Cannon-Bowers and colleagues (1995) assessed training motivation using a valence-instrumentality-expectancy approach. They found that training motivation predicted training reactions and performance. More recent studies in academic settings also document the importance of students' perception of training / education as being instrumental for their future success (Greene, Miller, Crowson, Duke, & Akey, 2004).

In conclusion it can thus be stated that training motivation is a broad concept and an important influence variable for the training process and its outcomes. Among the different theoretical approaches to training motivation, the valence-instrumentality-expectancy approach (Vroom, 1964) seems an appropriate starting point for motivation research in the training setting.

2.2.1.3 Organizational characteristics

Organizational factors are also important aspects to consider in the design of effective training programs. Some of them are important before the training (e.g., trainee selection), some of them become important after the training (e.g., workload), and some of them are important at all times (e.g., learning culture). They directly influence learning but also indirectly influence it through their influence on training motivation, expectancy, and attitudes (Cannon-Bowers et al., 1995; Salas & Cannon-Bowers, 2001).

Among those factors are the *organization's culture* concerning learning in general and the specific training objective in particular. Only recently has the concept of learning culture and its influence on trainee learning and transfer become a research focus. Conducted studies showed that a positive learning culture can have positive influences on learning and training outcomes (Schaper, Friebe, & Sonntag, 2003; Tracey, Tannenbaum, & Kavanagh, 1995). Simulation-based education furthermore needs to be viewed in the big scheme of organizational quality improvement and patient safety. Research suggests that simulation-based training can help foster a "culture of safety" where the analysis of error is emphasized instead of silence and blame. Yet the latter still existing culture oftentimes threatens the implementation and acceptance of simulation-based training (Helmreich & Merritt, 1998; Savoldelli, Naik, Hamstra, & Morgan, 2005). While this cultural aspect is particularly important for practicing physicians, the foundations of such a safety culture must be laid in the education of medical students.

The process of *trainee selection*, as a characteristic of the pretraining context, has also been shown to affect learning. Quinones (1995; see also Martocchio, 1992), for example, found that the way the training program was advertised or announced (i.e., training as reward or punishment, as "advanced" or "remedial", or as an "opportunity", etc.) influenced training outcomes. Motivation to learn has also been shown to be influenced by contextual pretraining factors, such as mandatory or voluntary attendance (Baldwin & Magjuka, 1997).

Situational constraints, such as financial and time limitations, can also have a detrimental effect on training outcomes, as they can, for example, lead to lesser quality training programs.

Especially influential are organizational factors after the training (Ford & Weissbein, 1997; Salas & Cannon-Bowers, 2001).

Learning transfer largely depends on the *supervisor's support*, as well as on peer, subordinate, and social support (Facteau et al., 1995; Rouillier & Goldstein, 1993; Smith-Jentsch, Salas, & Brannick, 2001; Tracey et al., 1995). Transfer is, for example, enhanced if supervisors show their interest in the newly learned skills, if they motivate employees to apply their new skills, and if they set transfer goals (Lemke, 1995).

Trainees' workload upon their return from training further determines if the new skills get transferred. Many times, the new skills may initially take more time or might still have to be practiced before they can be efficiently applied. This implies that trainees might not be able to

transfer the skills unless their workload gets temporarily reduced upon their return from training (Ford, Quinones, Sego, & Sorra, 1992; Salas & Cannon-Bowers, 2001).
Transfer can furthermore be supported by providing material or immaterial *incentives* for participating in the training program and transferring the learned skills and by creating opportunities to apply their new knowledge and skills (Ford et al., 1992; Quinones, Ford, Sego, & Smith, 1995).
Among the reasons for insufficient transfer is oftentimes a lack of *opportunity to practice or perform* the new tasks or skills once the trainee is back on the job. This lack is oftentimes related to too big of a workload for the trainees (Ford et al., 1992; Quinones et al., 1995).

In conclusion it can be stated that organizational factors are important determinants of training effectiveness. In the current setting they are, however, only of minor importance as the participants are medical students who are not required to immediately apply the learned skills in the organization. Factors such as mandatory attendance or opportunity to practice are nevertheless likely to be influential.

2.2.1.4 Instructional Systems Design (ISD)

ISD refers to the systematic planning and development of instruction based on an overall view of the training process and its components (cf. Briggs, 1977; see also section 2.1.1.2). ISD-tasks are consequently started before the training.
In the *training needs assessment* (2.2.1.4.1), the training developer identifies, for example, the tasks to be trained, the trainees' competence level, and resulting training needs. Based on the outcome of the needs analysis, *training objectives* can be *specified* together with the appropriate *content* to be taught (2.2.1.4.2) and the derivation of *evaluation criteria* (Goldstein & Ford, 2002; Kraiger & Jung, 1996) which can be used to assess performance as well as to guide feedback processes (2.2.1.4.3).

2.2.1.4.1 Training needs assessment

The *needs assessment* process has been described in detail by Goldstein and Ford (2002). It stands at the beginning of any training program providing the basic data necessary to develop the training. The following phases of *training and development* and *outcome evaluation* build upon these data. The *needs assessment* is used to determine where training is needed, what needs to be taught, who needs to be trained and whether training is the appropriate measure to address the organization's needs, objectives, and problems (Arthur, Bennett, Edens, & Bell, 2003, p. 235).
The *needs assessment* comprises several components (Goldstein & Ford, 2002, p. 34ff), the first of which is securing *organizational support*. Since the *needs assessment* already has to be considered an intervention into the organization, it is pivotal to secure support in the

organization. Only with the appropriate amount of trust and support can the organizational interruptions caused by the needs assessment be minimized. The formation of liaison teams or work groups is a recommended measure to improve organizational support.

The next component is the *organizational analysis*. It is conducted to find out where and when training is needed and to examine the systemwide components or organizational characteristics (cf. 2.2.1.3) that determine whether a training program can lead to behavior change on the job (Sonntag, Stegmaier, Schaper, 2006). It consists of the analysis of the organization's short- and longterm-goals, selection procedures, the determination of the training and transfer climate and the identification of external and legal constraints. It should focus on the congruence between training objectives and organizational factors, such as organizational goals, available resources, constraints, and support for transfer. Regarding the organization's goals, aspects such as employee recruitment, motivation, succession planning, and retention have to be investigated. Only by matching it to the organization's goals and strategy, can training have a long-term positive effect (Sonntag, Stegmaier, Schaper, 2006).

The *requirements analysis* is the third component. It includes the definition of the target job, the choice of methods for the needs assessment (such as observations, interviews, surveys), the selection of needs assessment participants, the choice of contact members in the organization, the anticipation of possible problems, and the development of a protocol for the needs assessment.

The *task and KSA* (knowledge, skills, abilities[3]) *analysis* requires the analysis of the tasks to be performed on the job by the participant after the training program. The first step of this analysis is oftentimes the specification of job tasks which are used in the second step to extract the KSAs needed to perform the job.

Person analysis refers to the assessment of the employees' current level of KSAs and their comparison to set standards for performing the job. It thus aims to identify who should be trained and what kind of training is needed by the individual. Pre-training individual characteristics, such as motivation and attitudes, can also be included in the assessment. This analysis is very important to prevent wasting time on training KSAs the employees already possess.

Overall, although the importance of the training needs assessment phase has been recognized, there is not yet enough research to develop a systematic methodology (Salas & Cannon-Bowers, 2001). Sonntag, Stegmaier, and Schaper (2006), however, provide a recent overview of the process and instruments that can be used in the organization analysis step.

Although a variety of training needs analysis methods has been developed, they have not yet been fully absorbed and adapted to the area of medical simulation training (Buerschaper, Hofinger, & Harms, 2003; Flechter, McGeorge, Flin, Glavin, & Maran, 2002; Goldstein & Ford, 2002; Schaper & Sonntag, 1998).

[3] Depending on the source, the A in KSA refers to either abilities or attitudes.

The needs assessment process leads to the following outcomes: Instructional objectives, from which guidelines for the design and delivery of training can be deducted (2.2.1.4.2), and evaluation criteria (2.2.1.4.3).

It can thus be concluded that a systematic needs analysis should guide the design, development, delivery, and evaluation of the training program (Arthur et al., 2003).

2.2.1.4.2 Training objectives and content

Based upon the results of the *needs assessment*, the *instructional objectives* can be defined which are then used to specify the skills to be trained. The *objectives* thus inform trainer and trainees of the training goals.

Several typologies have been developed for the categorization of skills (e.g., Gagné, Briggs, & Wagner, 1992; Rasmussen, 1986; Schneider & Shiffrin, 1977). Based on the overlap in these models (Goldstein & Ford, 2002), a general typology of skills seems to emerge that differentiates between cognitive (e.g., thinking, generating ideas, problem solving), interpersonal (e.g., teamwork, communication), and psychomotor skills (e.g., the use of the musculoskeletal system to perform behavioral tasks). For all of these different sets of skills, training objectives can be specified that indicate what the learner can accomplish upon successful completion of the training program.

Mager (1994) formulated guidelines for the definition of *behavioral objectives* that aim at communicating the training goals unambiguously. The behavioral objectives should indicate what the participants are able to achieve after the training, the conditions under which the performance must be maintained, and the criteria used for the evaluation of the participants, that is, the standards or levels of the expected performance (Mager, 1994).

Based on these formulated objectives, the appropriate *training content* can be selected. It should be noted that even small changes in the objectives might require rather large changes in the training content or methods (Patrick, 2003).

This implicates that the training objectives need not only be consistent with the identified training needs but also with the evaluation instruments administered after the training. In the following steps of the training development, the *objectives* are thus used for the *selection and design of the instructional programs* as well as for the definition of evaluation *criteria.*

2.2.1.4.3 Evaluation criteria

Based on the training objectives, the *evaluation criteria* can be specified. Evaluation criteria serve as measures of the training program's success by which the training program's value can be judged.

The selection of appropriate evaluation criteria is a complex process that goes beyond matching the criteria to the training objectives. The criteria can be evaluated along their relevancy, contamination, and deficiency (see Goldstein & Ford, 2002, for a detailed discussion). *Criterion relevancy* means that the KSAs needed to succeed in the training program are the KSAs needed to succeed on the job. Criteria are thus relevant to the degree that the items identified in the needs assessment phase are represented in the items chosen to assess the training program. *Criterion deficiency* refers to the degree to which elements identified in the needs assessment phase are not represented in the evaluation criteria. They moreover might or might not be contained in the training program itself. *Criterion contamination* is present when the evaluation criteria contain extraneous elements that have not been identified in the needs assessment process.

Beyond these considerations, different evaluation models have come up with categories or levels of training criteria. From this perspective, they are discussed in the section on training outcomes (2.2.1.6).

2.2.1.5 Training Design

The core of the training model is the *training design*. The goal of training design is to create a learning environment that allows for the attainment of the formulated *objectives*. Trainees need to be able to acquire the respective KSAs as easily as possible and transfer them to the according application setting (Patrick, 2003). As Goldstein and Ford (2002) point out, it is important to choose the environment based on the objectives and learning theory considerations and not to merely choose the training environment or method most readily available, newest or most interesting. As has been pointed out in the introduction, the latter is often the case in the area of medical simulation training.

Training design consequently concerns how the training content should be structured, sequenced, and organized into an effective and efficient learning program and by which methods the content should be conveyed to the learners. Accordingly, the next two sections cover learning design issues (2.2.1.5.1) and simulator/simulation design issues (2.2.1.5.2).

2.2.1.5.1 Training design: Learning design

Although historically the focus of simulator training research has been on the design of the simulator itself as well as on the design of the scenarios (cf. section 2.2.1.5.2), the focus needs to be broadened to also include the following elements of learning design: The role of the instructor (2.2.1.5.1.1), instructional strategy and methods (2.2.1.5.1.2), feedback (2.2.1.5.1.3), training motivation (2.2.1.5.1.4), and transfer support (2.2.1.5.1.5).

2.2.1.5.1.1 Instructor role

The *instructor* can assume different *roles* in training, depending on the training objectives and the chosen instructional strategy including the theoretical approach to learning and the employed training methods. In support of this line of thought, Good emphasized that "faculty that are going to 'teach with the simulator' must learn and hone new instructional skills and techniques" (2003, p. 20) because the requirements in the simulation setting often differ from requirements in other medical school teaching settings.

Simulation as an instructional method requires an active role of the learner and thus a new role of the instructor. How this role is specified largely stems from the underlying theoretical approach to learning (Leinhardt, 1993; see also 2.1).

Following the c*ognitivist educational* paradigm, the focus lies on *instruction*, that is, on the instructor and his expository and didactic methods and on how to best convey knowledge from the instructor to the students. The instructor thus assumes the *active role* of the didactic leader and is thus primarily didactically directive (Leinhardt, 1993). Broken down into actual teaching behaviors, this means that he systematically presents and explains the material in small steps, guides the learner, and consecutively evaluates the learners' performance (cf. Reinmann-Rothmeier & Mandl, 2001).

Although the focus is on teaching in this paradigm, Goldstein (2002) points out that there has not yet been enough research on instructors' characteristics in regards to fostering learning. Most of the available studies have collected data on the reaction level (Kirkpatrick, 1994) only. Bartlett (1982) provided a collection of instructor characteristics assumed to improve learning. The following table (Table 1) contains examples.

Table 1. *Example items of effective instructors' characteristics*

Items: The instructor …
• is well-organized
• presents an outline of the lesson
• designs the sequence of materials for maximum learning
• emphasizes conceptual understanding
• relates lectures to other aspects of course
• answers questions clearly and thoroughly
• uses examples
• sets difficult but attainable goals
• points out how materials the students are learning can be useful
• encourages discussions
• explains how topics in the course are related to each other
• is well-prepared
• encourages students to learn the material
• allows students to express problems related to the course
• effectively uses audiovisual aids
• accomplishes the course objectives
• shows enthusiasm for the subject
• stimulates interest in the subject

Note. From Bartlett (1982) and Goldstein and Ford (2002, p. 216f).

Research on instructional effectiveness was reviewed by Borich (1989; 1992) resulting in the following five categories of instructor behaviors. *Clarity* describes the teacher's ability to present the content in an understandable manner. *Variety* concerns the teacher's flexibility during the lesson regarding, for example, the use of different media and instructional methods. Instructors who value *task orientation* strive to cover all of the material and focus on students' learning instead of on their enjoyment. Effective instructors furthermore focus on the learners' *engagement in the learning process*, meaning they adapt their materials to the learners' skill levels and provide feedback. *Moderate to high rates of success* of learners are also emphasized by effective instructors. To reach these rates, the instructors structure their lessons in a coherent way that builds on prior learning and considers the amount of new material that can be absorbed by learners.

Although faced with many disadvantages, such as learner passivity and the lack of learner responsibility, this role of the instructor as the didactic leader can be helpful and advantageous in certain contexts. The advantages are transparency, focus, and predictability (Leinhardt, 1993). In well-designed, strong didactic models, knowledge is still constructed by the student but in response to the instructor and the presented content. What is missing, however, is that exploration and the merging of multiple intuitive meanings are not attended to systematically under these models (Leinhardt, 1993).

In the constructivist paradigm, the focus consequently shifted from optimizing teaching to optimizing *generative* learning processes (Cognition and Technology Group at Vanderbilt, 1993a). This changes both, the learner's and the instructor's role. The learner has a very active role in these constructivist learning environments, the instructor a *reactive role*. The instructor's role becomes that of a facilitator or coach (Leinhardt, 1993). He now arranges meaningful learning environments and poses problems or questions, provides problem-solving tools and supports the learner at his request (Leinhardt, 1993). The instructor is responsible for initiating, guiding, and supporting the learner's constructive processes. He challenges students and focuses their attention on relevant features of the task (Leinhardt, 1993). Errors are still corrected by the instructor through private and public reflection of results and effects (Brown & Palinscar, 1984; Leinhardt, 1993). However, the instructor focuses more on supporting the learners' self-evaluation processes than on evaluating the students himself (e.g., Jones, 1992).

A lot of research on instructor qualities has also been conducted in the area of problem-oriented learning as one form of a constructivist learning approach (see Albanese & Mitchell, 1993; Mayo, Donnelly, Nash, & Schwartz, 1993; Moust & Schmidt, 1995; Schmidt, Van der Arend, Kokx, & Boon, 1995). Important aspects and tasks of the instructor and his role that have been identified in this area are guiding the work of the students by questioning and probing, encouraging critical appraisal, and balancing foci; and promoting interactions by supporting interpersonal relationships and encouraging students to direct the tutorial (Wilkerson, 1992). Another study identified three underlying aspects of the tutor role: facilitative expertise (ability to facilitate group work), knowledge expertise (medical knowledge), and clinical reasoning expertise (knowledge of medical problem-solving and critical reasoning) (Caplow, Donaldson, Kardash, & Hosokawa, 1997).

In this area of research it was also pointed out, that the notion of facilitation or facilitating discussions instead of directing them is "foreign to many faculty" (Albanese & Mitchell, 1993, p. 74).

In conclusion, these elaborations of the instructors' role show that in order to be effective as instructors, they need knowledge about the subject area but they also need to gain knowledge about effective teaching and according learning paradigms. It needs to be emphasized that the definition of the instructor's role is an important, yet often neglected, aspect of training design.

As has been pointed out in the section on the theoretical approaches to learning (2.1), an integrated position incorporating cognitivist and constructivist elements seems to be most suited to foster learning. This position has also been adopted by the cognitive apprenticeship approach (Brown et al., 1989). In this approach the role of the instructor was furthermore specified into modeling-, coaching-, scaffolding-, fading-behaviors, as well as articulation- and reflection-support, as was detailed in section 2.1.2.2.

The integrated position seems especially useful when using simulation as a central training method as it is a very complex and variable method. It requires instructors to adapt their teaching style and role to the setting's differing requirements. The instructor might function as a teacher/presenter of material, when his main objective is to convey information (e.g., declarative or procedural knowledge about the simulator and its functions); as a coach when he primarily provides help and feedback during the learning process (e.g., when the students conduct a standard anesthesia); and as a moderator/facilitator when he focuses on guiding group processes (e.g., during the reflection/feedback phase).
It is important to specify these different roles and to provide training for less experienced instructors in these different roles.

2.2.1.5.1.2 Instructional strategy and methods

Instructional strategies have been defined as "a set of tools (e.g., task analysis), methods (e.g., simulation), and content (i.e. required competencies) that, when combined, create an instructional approach" (Salas & Cannon-Bowers, 2001, p. 481; Salas & Cannon-Bowers, 1997). Within the framework of this model, further tools are the ones mentioned in the ISD section (2.2.1.4), other methods are cognitive modeling or practice conditions, as will be elaborated below, and the respective content are basic anesthesia skills (see 2.2.1.5.2.2.1). The instructional strategy's aim is the progression from declarative knowledge through knowledge compilation (integration of facts) to the establishment of procedural knowledge (c.f. Ackermann & Kyllonen, 1991; Anderson, 1987).
Salas and Cannon-Bowers (1997) describe four basic principles as the basis of most effective training strategies: a) *presentation* of the relevant information; b) *demonstration* of the knowledge, skills, and abilities the trainees should learn; c) creation of *practice* opportunities for the skills to be learned; and d) provision of *feedback* to trainees during and after training (see 2.2.1.5.1.3). The enumeration of these principles shows that elaborations on instructional strategies are primarily based on instructional design or cognitivist learning theories. Nevertheless could some of these principles or elements also be embedded in constructivist approaches, such as the creation of practice opportunities in authentic contexts. Therefore, the inspection of the different strategy elements is especially useful when striving to create an integrated learning approach (cf. 2.1.3.1). The following paragraphs will thus inspect the above mentioned principles, except for the feedback principle, which is covered in section 2.2.1.5.1.3.

Presentation of information. How information should be presented has been of great research interest in the instructional design area (Goldstein & Ford, 2002). Since the presentation of information is mainly instructor-centered, the according guidelines can be derived from the teaching behaviors specified in cognitivist approaches (see 2.2.1.5.1.1).
It is to be noted, however, that - even within presentation phases - the learners' activation is of pivotal importance for the learning process. The more active the learners are, the greater the retention and transfer of the learned KSAs (Campbell, 1988).

Demonstration of KSAs. For the demonstration of the necessary KSAs in the anesthesia area, cognitive and behavior role modeling have been recommended as training methods (cf. Gräsel, 1997; Schaper, Schmitz et al., 2003; Tannenbaum & Yukl, 1992). As has been mentioned above, modeling is also a central element of the cognitive apprenticeship approach (see 2.1.3.2) and originated in Bandura's social learning theory (1986).
Research on modeling indicated that it is an effective training method (see Tannenbaum & Yukl, 1992). Behavior role modeling emphasizing practice and performance feedback, for example, was shown to lead to more learning than lecture-only or lecture with demonstration formats (Smith-Jentsch, Salas, & Baker, 1996). Concerning the type of used role models, Baldwin (1992) found that the best learning outcomes resulted after trainees had only been exposed to positive role models, displaying the correct behaviors. Transfer and generalization however, were better when trainees had been exposed to both, positive and negative, models.

Practice. Practice is necessary for skill acquisition (Anderson, 1982) but its complexity has long been ignored (Ehrenstein, Walker, Czerwinski, & Feldman, 1997; Salas & Cannon-Bowers, 2001). Research on practice has looked at practice intervals, reinforcement schedules, and other conditions to maximize learning and transfer (Baldwin & Ford, 1988). Cannon-Bowers and colleagues (1998) delineate conditions, concepts, and interventions to enhance the utility and effectiveness of practice. They mention goal orientation and preparatory information among the pre-practice conditions needed to improve the efficiency and effectiveness of practice (see 2.2.1.1.2, 2.2.1.1.3). Among the practice-conditions they mention practice schedules, as well as learner-driven practice and guided/cued practice. The latter conditions resemble the cognitive apprenticeship approach's elements of scaffolding and fading (cf. 2.1.3.2).
Concerning the practice schedules, research has found that *distributed practice* led to improved results as compared to *massed practice* (Baldwin & Ford, 1988; Goettl, Yadrick, Connolly-Gomez, Regian, & Shebilske, 1996): Learned material was generally retained longer. Baldwin and Ford (1988) further mention evidence that the learning of complex tasks benefits from initially massed practice sessions followed by briefer sessions.
Concerning the type of the practiced units, practice can further be differentiated into *whole- versus part-task learning* (Goldstein & Ford, 2002). In whole-task learning, the learner practices the task as a single unit, in part-task learning, he practices the task broken down into

separate components. The choice of whole- vs. part-task practice should be based on the task's complexity and on the relationship between its components. For tasks with high organization an increase in task complexity renders whole methods more efficient than part methods. Fredericksen and White (1989), however, showed that even in highly organized tasks part-task training can be superior to whole-task training if the parts are psychologically meaningful. For tasks with low organization, an increase in complexity renders part methods more efficient (Goldstein & Ford, 2002). Furthermore, whole-task training has been found superior to part-task training for high ability trainees, distributed practice, low task complexity, and high task organization (Baldwin & Ford, 1988). Care should be taken in part-task training to provide methods that foster the combination of the different subtasks into the whole task (Goldstein & Ford, 2002).

Concerning *training methods*, it has furthermore been pointed out that the effectiveness of the training is determined by the match between the training delivery method and the skill to be learned. Thus, the most effective training method can only be found if skill and task characteristics are taken into consideration (Wexley & Latham, 2002). Depending on the training objectives and the training needs, simulation as the main *training method* thus might have to be enhanced by further methods (Gräsel, 1997). This might be the case, if the participants' skill level is low but the simulations' demands are high. In their review, Tannenbaum and Yukl (1992) recommended that simulations should include the presentation of an effective model, clear descriptions of relevant skills, debriefing segments, opportunities for practice, coaching, feedback, and reinforcement. An instructional approach that includes these aspects is, for example, the cognitive apprenticeship approach described above (see 2.1.3.2).

In conclusion it can be stated that a simulator-based anesthesiology curriculum can thus be conceptualized as an instructional strategy aimed at building medical students' basic anesthesiology competencies by applying well-tested training tools (e.g., exercises, feedback loops) and appropriate training methods (e.g., simulation, lectures, case practices) targeted at specific content (i.e., basic anesthesiology knowledge and skills) (see Davies, 2001).

2.2.1.5.1.3 Feedback

Practice alone does not improve performance if appropriate feedback is lacking (Patrick, 2003). Appropriate feedback guides and enhances the learning process in simulation training. Since not every kind of feedback is effectively enhancing the learning process (Kluger & DeNisi, 1996), it is important to specify which type of feedback shall be given in which situations and by whom.

Feedback interventions have been defined as "actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task performance" (Kluger & DeNisi, 1996, p. 255). Excluded from this definition are, for example, task-generated feedback that is received without an intervention, or intrinsic feedback, or personal feedback that does not relate to the task (see Kluger & DeNisi, 1996, for a detailed differentiation).

The use of feedback is among the earliest variables that were investigated as an intervention to enhance learning (cf. Thorndike, 1927). Yet research has produced inconsistent results over the years and has thus indicated that feedback has highly variable effects on performance (Goldstein & Ford, 2002). Whereas many studies have shown positive effects of feedback (e.g., Komaki, Heinzmann, & Lawson, 1980; Richardson & Waller, 2005; Youmans & Stone, 2005), others showed that feedback also produced negative results (see Balcazar, Hopkins, & Suarez, 1985; Kluger & DeNisi, 1996; Latham & Locke, 1991; Salmoni, Schmidt, & Walter, 1984). Improved performance as a result of feedback can be attributed to informational as well as motivational effects (Goldstein & Ford, 2002).

Feedback thus needs to meet a number of conditions in order to be truly effective. Kluger and DeNisi (1996) tried to identify and investigate these conditions in a comprehensive Feedback Intervention Theory (FIT) to integrate the inconsistent results concerning the effects of feedback on performance. FIT suggests that learners perform better after the feedback if the feedback focuses trainees' attention on the task instead of on themselves as a person. Kluger and DeNisi (1996) investigated a large number of feedback cues and task characteristics that can play a role in the effects of feedback interventions. They showed, for example, that praise, feedback designed to discourage (i.e., to threaten self-esteem), and verbal feedback attenuated feedback intervention effects. These types of feedback focus the recipients' attention more on the self than on the task. Frequent feedback, feedback providing the correct solution, and feedback concerning velocity, on the other hand, augmented feedback effects. These types of feedback focus the recipients' attention more on the task at hand. It can thus be concluded that feedback cues affect performance by changing the recipients' focus of attention. The more the focus of attention is shifted to the task, the stronger the benefit of feedback for task performance. Feedback Intervention Theory further proposes that feedback effects depend on task characteristics. The specific task properties that moderate feedback effects are, however, still only poorly understood. There is first evidence that simpler-task performance benefits more from feedback, as does task performance in more advanced skill acquisition stages. The latter point was also underlined by Schmidt and Wulf (1997). They showed that the timing of feedback is an important determinant of its effect as they found that continuous feedback during skill acquisition is detrimental to learning. Ilgen, Fisher, and Taylor (1979) found that high frequency of feedback is often detrimental to performance. This can be attributed to a perceived loss of control by the learners or to their sole reliance on external feedback which prevents them from developing their internal feedback skills. Undeveloped internal feedback skills then decrease the chance of successful transfer to the application setting where this feedback is missing. It has thus been recommended by different authors and in different instructional approaches to gradually reduce the feedback as the learners proceed in their skill acquisition (cf. Collins et al., 1989, "fading"; Winstein & Schmidt, 1990).

In Kluger and DeNisi's study (1996), the sign of the feedback (positive vs. negative) did not have an effect. There is, however, evidence that negative feedback in combination with attention directed to the self might lead to an internal attribution for the cause of failure and result in a strategy of disengagement from the task. Feedback cues, on the other hand, that foster self-efficacy and focus the attention on the task at hand should cause learners to increase their effort.

A recent study (Ellis & Davidi, 2005) has shown that performance of trainees improved significantly when they received feedback on their failures and successes, compared with other trainees who received feedback on their failures only. It is to be noted that the most critical issue for conveying negative feedback is the balance between making it possible for performers to accept responsibility for their substandard performance and, at the same time, not lower their self-concept (Ilgen & Davis, 2000).

Studies on the impact of feedback sequence (Stone, Gueutal, & McIntosh, 1984) showed that a positive-negative sequence led to more positive reactions concerning feedback accuracy than the other way round. Concerning feedback delivery in groups, Davies and Jacobs (1985) found that a sandwich order of feedback (positive – negative – positive) resulted in the most positive reactions by the group.

Kluger and DeNisi (1996) conclude that feedback can produce a large positive effect on performance when it is provided for a familiar task, contains cues that support learning, focuses attention to the discrepancy between actual performance and performance standards at the task level, and does not contain cues that direct attention to the self.

Another line of research investigated the recipients' reactions towards the rater (i.e., the person giving feedback) (e.g., Waung & Jones, 2005). Ilgen, Fisher, and Taylor (1979) emphasized that individual characteristics determine how recipients perceive feedback. High performers who are growth-oriented require feedback that emphasizes competency and leaves room for their personal initiative. Poor performers might need closer monitoring and more specific feedback. Waung and Jones (2005) found that the recipients react more favorably to feedback when the rater used legitimizing statements, that is, when he claimed to adhere to established guidelines or standards in performance evaluation. This effect might be attributed to the changes it causes in respect to the credibility of the source. Ilgen, Fisher, and Taylor concluded in their earlier studies (1979) that the accuracy of feedback might be affected by the credibility of the source. The person giving feedback must thus work to establish credibility, for example by using legitimizing statements.

Waung and Jones (2005) furthermore found evidence that recipients reacted more favorably to the rater when he used less personal language, that is unassuming, acquiescent, informative language, and third person pronouns. The latter finding, concerning the kind of language used in feedback sessions, still warrants further research, though.

In conclusion it can be said that the provision of feedback is helpful in learning processes but carries many dangers. Feedback information should be accurately, clearly, and simply related to the task performance, the discrepancy between what has been accomplished and what is required should be clearly indicated, the trainee should receive at least cues of the results of his actions that are inherent in the task and that he can thus also use as a feedback source after completion of the training. The feedback should furthermore be credible, timely, and constructive (Campbell, 1988; Wheaton, Rose, Fingerman, Karotkin, & Holding, 1976). At the current stage of research it does not seem possible to create feedback interventions that incorporate all of the different and often contradictory findings. Kluger and DeNisi's theory (1996) along with some older and more practical guidelines seem to provide the best basis for the design of feedback interventions.

2.2.1.5.1.4 Training motivation

As Noe (1986) pointed out, training motivation can be viewed from a selection angle as well as from a design or training development angle. This is reflected in the training model (see Figure 1) as motivation is included in several places. Whereas the selection angle looks at the trainees' motivation they bring to the training as an individual characteristic (as elaborated in section 2.2.1.2) and how motivated trainees can be selected, the design angle looks at how trainees' motivation can be supported during the training. The latter angle will be the focus of this section.

A number of strategies to improve training motivation have been developed. Some of them depend on the underlying operationalization of motivation. If motivation is operationalized as goal orientation, for example, strategies to foster a learning goal orientation are used to improve motivation (see e.g., VandeWalle et al., 2001). Aside from these operationalization-specific strategies, sets of more general strategies to improve motivation have been developed from an instructional viewpoint. It is to be noted, however, that these two kinds of strategies cannot always be clearly distinguished from each other and oftentimes merely depend on the focus of the intervention or study.

Strategies to improve training motivation from an instructional viewpoint are, for example, the provision of an advance organizer (Ausubel, 1960) or training goals (Hesketh, 1997). Both of these strategies also have cognitive effects in that they better prepare the trainees for the new material and enable them to direct their attention. Knowledge of how the training fits into the trainees' future job and career paths has furthermore been shown to have positive emotional effects (Patrick, 2003). Smith-Jentsch and colleagues (1996) found that negative experiences prior to training increased trainees' motivation to learn. Negative experiences were in this case related to the training content. Trainees had thus previously experienced a lack of KSAs that were subsequently taught in training. Tannenbaum and Yukl (1992) further showed that trainees' attitudes and perceptions of the training program are determinants of their training motivation.

A useful overview of different motivational design strategies was provided by Keller and Kopp (1987). They developed the ARCS theory of motivational design from an Instructional Design viewpoint. The model identifies twelve strategies subsumed under four categories for improving motivation during training. The strategies are presented in the following table (Table 2).

Table 2. *Motivational design strategies from the ACRS model of motivational design*

A. Attention	**C. Confidence**
1. *Perceptual arousal*. Gain and maintain student attention by the use of novel, surprising, or uncertain events in instruction. 2. *Inquiry arousal*. Stimulate information-seeking behavior by posing or having the learner generate questions or a problem to solve. 3. *Variability*. Maintain student interest by varying the elements of instruction.	7. *Expectancy for success*. Make learners aware of performance requirements and evaluation criteria. 8. *Challenge setting*. Provide multiple achievement levels that allow learners to set personal standards of accomplishment and opportunities that allow them to experience performance success. 9. *Attribution molding*. Provide feedback that supports student ability and effort as determinants of success.
B. Relevance	**D. Satisfaction**
4. *Familiarity*. Use concrete language and use examples and concepts that are related to the learner's experience and values. 5. *Goal orientation*. Provide statements or examples that present the objectives and utility of instruction, and either present goals for accomplishment or have the learner define them. 6. *Motive matching*. Use teaching strategies that match the motive profiles of the students.	10. *Natural consequences*. Provide opportunities to use newly acquired knowledge or skill in a real or simulated setting. 11. *Positive consequences*. Provide feedback and reinforcements that will sustain the desired behavior. 12. *Equity*. Maintain consistent standards and consequences for task accomplishment.

Note. From Keller and Kopp (1987).

Table 2 shows that students' in-training motivation can be enhanced by getting and maintaining their attention, by establishing the relevance of the training material to their goals, by giving them sufficient confidence to participate in the training, and by increasing their satisfaction through rewarding new skills and appropriate performance.

In the light of a plethora of training motivation research, these guidelines provide a pragmatic starting point for training designers. Although these design principles are grounded in Instructional Design theory and thus in traditional approaches to training, most of these strategies are also incorporated in constructivist learning environments. The only strategies that are not as present in the more radical constructivist learning environments are *attribution molding* and *equity*.

Concerning the notion of *relevance*, further guidelines could be derived from Vroom's (1964) expectancy theory (cf. 2.2.1.2.1) that implies that trainees need to be informed of the future outcomes that can be obtained as a result of the training.

In conclusion it can be remarked that the presented research offers several techniques for training designers to foster trainees' motivation. A lot more research on training motivation has been conducted, focusing for example on trainee choice, or framing of the training program (see Goldstein & Ford, 2002, for an overview). This research is however, not applicable to the context of this training in a mandatory university curriculum.

2.2.1.5.1.5 Transfer support

A lot of research has dealt with transfer and its support (see e.g., Baldwin & Ford, 1988; Ford & Weissbein, 1997). However, there is still a lack of transfer studies relevant to simulation training. This implies that the empirical base to draw simulation training design principles from is limited. Until respective findings become available, the design of simulators and the according simulations has to be based on general findings in the transfer literature (Goldstein & Ford, 2002).

Holton (1996) pointed out that one reason why transfer might not occur is that training design seldom provides for transfer of training. He consequently called for interventions that aim at directly influencing transfer of training.

Such interventions for supporting transfer can be distinguished in interventions that focus on the learning environment/training itself and interventions that focus on the application environment or organizational post-training context, such as supervisor support or situational consequences (see Baldwin & Ford, 1988; Salas & Cannon-Bowers, 2001, for information on these aspects, as well as section 2.2.1.6.3). This section will focus on the interventions concerning the learning environment since medical school curriculum designers will not be in the position to influence possible future work environments of their students.

Depending on the basic approach to learning and teaching, for example behaviorist, cognitivist, or constructivist, different answers were found as to the necessary training interventions to support transfer (see Bergmann & Sonntag, 2006, for an overview). In behaviorism, "identical elements" (Thorndike, 1914; Thorndike & Woodworth, 1901) are considered the underlying transfer mechanism. The theory states that transfer can be enhanced by increasing the degree of correspondence between the stimuli, responses, and conditions employed in the learning setting and those in the application setting.

Cognitivism focuses on teaching learners general problem-solving strategies or principles to enhance transfer. To support transfer, learning environments should thus focus on the general principles necessary to master a task so that the learner can then use these principles to solve the task in the transfer setting (Goldstein, 1993).

The constructivist approach was developed to a large extent in reaction to the problem of inert knowledge and the support of transfer is thus interwoven in the design of constructivist learning environments. Nevertheless, it has to be stated that even within constructivist learning environments, transfer doesn't occur spontaneously (Hendricks, 2001). Learning environments thus have to be designed to provide authentic, situated, and social multiple contexts and perspectives (see above, section 2.1.2) to support transfer. In the anchored instruction approach, anchors are furthermore used to support knowledge application and thus transfer (cf. section 2.1.2.2).

From an instructional design point of view, Patrick (2003) concludes that the notion of "identical elements" is still at the core of current transfer support strategies and is also reflected in the simulator fidelity discussion (see 2.2.1.5.2.1.2). The only difference lies in the conceptualization of these elements. Gick and Holyoak (1987) emphasize that shared identical elements can nevertheless only be considered a necessary but not sufficient condition for transfer. The learner furthermore needs to perceive and be aware of the similarity between the learning and the application setting in order for transfer to take place. For the training design they thus recommend to inform learners of the range of tasks and situations in which their new KSAs are relevant. Trainees furthermore need to be taught how to apply their new KSAs on the job.

Besides needing to know how to apply their KSAs on the job, these KSAs also need to be stable enough to be transferred. Transfer often does not occur because the new KSAs are still too fragile and may thus break down in application settings that are distracting and stressful (Patrick, 2003).

Based on the plethora of research (e.g., Baldwin & Ford, 1988; Clark & Voogel, 1985; Holton, 1996; Salas & Cannon-Bowers, 2001), the following table (Table 3) contains an overview of the most widely recommended principles and methods for supporting transfer. Following Leifer (1980), these principles and methods to support transfer have been categorized into three segments: before, during, and after training. The table furthermore shows if the principles and methods were specifically aimed at fostering near or far transfer in the respective literature.

Table 3. *Learning design principles and methods to support transfer*

Principles/methods (with exemplary sources)	Time of method use			Type of transfer supported	
	before training	during training	after training	near	far
Advance organizer (Ausubel, 1960)	x				
Goal setting (assigned and participative) (Werner, O'Leary-Kelly, Baldwin, & Wexley, 1994; Wexley & Baldwin, 1986)		x	x		
Goal setting with self-management training (Gist, Bavetta, & Stevens, 1990)		x			
Participant activation (Campbell, 1988)		x			
Opportunity to perform / to practice the training content in a job context (Holton, 1996; Quinones et al., 1995)		x	x		
Teaching how to apply the learned KSAs to the job (Clark & Voogel, 1985; Holton, 1996)		x			
Identical elements in the learning and transfer setting (Thorndike & Woodworth, 1901)		x			
Stimulus variability (Ellis, 1965)		x			
Teaching general principles (McGehee & Thayer, 1961)		x			
Feedback (Baldwin & Ford, 1988; Goldstein & Ford, 2002)		x			
Relapse-prevention modules (Tziner, Haccoun, & Kadish, 1991)			x		
Refresher courses (Wexley & Baldwin, 1986)			x		

Table 3 continued. *Learning design principles and methods to support transfer*

	Time of method use			Type of transfer supported	
Principles/methods (with exemplary sources)	**before training**	**during training**	**after training**	**near**	**far**
Overlearning (Noe, 1986; Wexley & Latham, 1981)		x		x	
Similarity between learning and application setting (Baldwin & Ford, 1988)		x	x	x	
Teaching trainees underlying principles, concepts, and assumptions of the KSAs to be learned (Goldstein, 1993)		x			x
Practicing in different contexts and using novel practice exercises (Baldwin & Ford, 1988)		x			x
Encouraging trainees to discuss and apply the learned KSAs in application settings (Noe, 1986), among them application settings they were not trained for (Goldstein, 1993)		x	x		x
Introduction of difficulties (Schmidt & Bjork, 1992)		x			

Note. This list is not exhaustive but focuses on the most widely recommended principles that could be of use in a medical simulation training setting.

Principles to foster near and far transfer (cf. 2.2.1.6.3) have been extracted by Clark and Voogel (1985). Whereas recommendations to support near transfer are primarily derived from the identical elements theory (Thorndike & Woodworth, 1901), recommendations to foster far transfer (cf. 2.2.1.6.3) are based on principles theory (Yamnill & McLean, 2001).

In conclusion it can be stated, that over the years, research has shown that the design of intervention strategies does influence the transfer probability (e.g., Collins et al., 1989; Salas & Cannon-Bowers, 2001). Of further importance is the training context because it affects motivations, expectations, and attitudes for transfer (e.g., Quinones, 1997; Salas & Cannon-Bowers, 2001).

As this section has shown, many transfer-enhancing strategies have been extracted in the training literature. For the training designer, the question remains, which methods or principles to use. Answering this question becomes even more difficult when considering the so called transfer dilemma (Hesketh, 1997). Hesketh (1997) pointed out that methods that enhance transfer often do not lead to an increase in immediate learning, such as variation of training tasks or contexts, or error training. The question of which strategy to employ can only be answered within the framework of the complete training program with its requirements and goals and thus points out the importance of taking a systems approach to the development of training.

Holton concludes with stating that transfer design remains nevertheless difficult to measure due to the lack of definite guidelines as to what constitutes appropriate transfer designs (Holton, 1996, p. 15). Adding to this measurement difficulty is the multidimensionality of the transfer construct (Yelon & Ford, 1999).

2.2.1.5.2 Training design: Simulator and simulation design

After considering learning design issues that are applicable to training programs in general, this part of the training design deals with specific choices and considerations that have to be made in the area of simulators (2.2.1.5.2.1) and simulations (2.2.1.5.2.2). Whereas the training designer has to primarily select the appropriate simulator with the desired design features, his design influence is much larger in the area of the simulations or scenarios conducted with the help of the simulator.

2.2.1.5.2.1 Simulator design

Simulators have always been used for a variety of purposes, such as research (e.g., to determine the effects of working conditions, ergonomic design variations), education (for very different target groups, e.g., laymen, college students, or students of specific professions), training (of professionals), performance assessment, testing of new equipment before its use, and engineering design (Boulet et al., 2003; Bradley, 2006; or Chopra, 1996, for detailed discussions of possible applications; Dalley, Robinson, Weller, & Caldwell, 2004; Flexman & Stark, 1987; Gaba, 2000; Gaba & DeAnda, 1988; Grevenik & Schaefer, 2004; Murray et al., 2004). As a result, a wide variety of simulators have been developed. But even in the training setting, different kinds of simulators are available (see Schmitz, 2002). The following sections will primarily deal with simulators in the anesthesia training setting.

In the process of the training's development, an appropriate simulator must be chosen based on the training objectives set for the specific training program. To be able to choose among the different simulators, the following aspects of simulator definition and classification (2.2.1.5.2.1.1), as well as of simulator fidelity (2.2.1.5.2.1.2) have to be considered.

2.2.1.5.2.1.1 Types of simulators

In the training context, simulators are generally viewed as a training device. “A training device is an arrangement of equipment, or materials that simulates the actual task environment and provides functional interaction by the trainer” (Hays & Singer, 1989, p. 13). Training devices include part-task and whole-task trainers. While part-task trainers only represent parts of the operational task, the latter ones teach the complete operational task (Hays & Singer, 1989). Although some researchers limit the term “simulator” to technologies that replicate the complete task environment (Cooper & Taqueti, 2004), different kinds of part-task and whole-task trainers are many times categorized as simulators in a more generic sense. Hays and Singer (1989) point out that the “distinction between simulators and other training devices is really a matter of degree, since any training device is a form of simulation of the operational situation and all support some level of practice” (p.13). They suggest the following definition of *training simulators*:

> “A simulator is a complex device that provides a highly realistic simulation of the operational situation and provides a situation adequate for practicing and maintaining previously acquired skills” (p. 13-14).

Flexman and Stark (1987, p. 1014) offered a functional definition of training simulators. Emphasizing the two primary functions of training simulators, they further specify the above definition: A training simulator’s first function is

> “to present information like that associated with some real system for which training is required. The simulator stores, processes, and displays information reflecting the functional characteristics of the system, the effects of relevant environmental events, and the effects of control inputs made by the operator.”

Its second function is to

> “incorporate special features that facilitate and enhance their ability to support practice and learning for the express purpose of influencing operator performance in the real system which is simulated.”

These definitions emphasize the two-fold function of simulators: The representation of the real system and the provision of opportunities for the trainee to practice relevant behaviors and thus receive feedback on them.

The second function implies that simulators are designed to simulate only the relevant elements of the real-world system (Hays & Singer, 1989) and often contain additional elements which are not a part of the real system to improve their training effectiveness. These instructional features are supposed to enhance the learning process, but also imply a departure from reality. Examples are video cameras and the possibility to restart or stop any scenario to provide enhanced feedback to the trainees (Hays & Singer, 1989). Simulation environments hence allow for the learner to receive feedback in various forms and at any point during or after the task, thus providing an excellent basis for process-oriented feedback (Leplat, 1989).

A definition specifically for *anesthesia training simulators* was offered by Chopra (1996, p. 299):

> "A simulator mimics the environment and phenomena as they appear in the real world. Thus, an anesthesia simulator should provide a learning experience that has the look and feel of a real operating theatre and real patient."

This definition points out that the actual mannequin or patient simulator is only a part of the complete anesthesia simulator. It does not, however, apply to every anesthesia training simulator. Different classes of simulators and training devices need to be distinguished, as not every simulator will, for example, represent the complete system including the environment. As Hays and Singer (1989) pointed out, a simulator is a kind of training device, but the distinction is not always clear and different authors use different categorizations and terminologies. In general, simulators are more sophisticated and complex training devices that thus provide more realistic simulations.

The following classification of anesthesia training devices and simulators (Table 4) was offered by Chopra in line with Gaba (see also Bradley, 2006; Chopra, 1996, p. 300; Ziv, Wolpe, Small, & Glick, 2003).

Table 4. *Classification of anesthesia simulators and training devices*

Anesthesia training devices
• Computer assisted instruction programs, e.g., Gas Man, TIV-Sim
• Part-task trainers, e.g., intubation and resuscitation manikins
Anesthesia simulators
• Computer screen-only simulators, e.g., ASC, BODY
• Full-scale or realistic simulators, e.g., CASE, GAS, LAS, Sophus
• Virtual reality simulators

Note. From Chopra (1996, p. 300).

The category of *training devices* comprises *computer assisted instruction programs* which can, for example, simulate the distribution process of inhalation agents (Chopra, 1998). The trainee interacts with a computer program that presents relevant information on the computer screen but does not represent the complete anesthesia work environment. *Part-task trainers* are also anesthesia training devices. They replicate only critical aspects of the system, without providing the complete system or its environment (Goldstein & Ford, 2002), and allow for the practice of subsets of tasks (Helmreich & Schäfer, 1998). For this reason, they are most of the time not classified as simulators but as training devices (Chopra, 1998; Steininger, 1995) and are used to teach specific skills or subtasks. Examples from the medical domain are mannequins that consist of only a head and lungs on which airway management skills, such as intubation can be practiced (Ashton, Kennedy, & Kirkland, 2004), the use of anesthesia machines can be learned (Helmreich & Schäfer, 1998), or the application of cricoid pressure can be practiced (Owen, Follows, Reynolds, Burgess, & Plummer, 2002). Further examples are mannequins that allow for the practice of emergency procedures, such as resuscitation.

The category of *anesthesia simulators* comprises three kinds of simulators. The first are *computer screen-only simulators*, such as the Anesthesia Simulator Consultant (ASC) (Schwid & O'Donnell, 1990) or the simulation of pediatric basic life support (Romalho de Assis, Sigulem, & de Carvalho, 2005). They represent the entire anesthesia work environment on the computer screen and can be adapted to the trainee's level of expertise. They do not simulate reality and are not able to represent certain important system characteristics, such as interactions between operating room (OR) personnel or human-machine interactions (Chopra, 1998), nor are they able to provide the trainee with sensomotoric practice (Steininger, 1995), such as hands-on intubation. Nevertheless, they can be used to teach technical skills (Nyssen, Larbuisson, Janssens, Pendeville, & Mayne, 2002).

More realistic simulations are provided by f*ull-scale simulators* that "recreate the anesthesia work environment in which the mock patient and equipment look, feel and behave as they do in real life" (Chopra, 1998, p. 42). A training environment is thus established in which participants can administer anesthesia under highly realistic conditions. One example is the METI PatientSimulator^{TM} that will be described below as it is used in the HANS full-scale simulation sessions. It consists of the mannequin, anesthesia and ventilation machine, monitoring equipment, and computer hard- and software, and is installed in fully functional operating room. Through the representation of the complete system with its environmental context these full-scale simulators can thus reproduce the dynamism of the simulated tasks (Steininger, 1995).

Virtual reality simulators are just now becoming established in medicine (McCloy & Stone, 2001). Their application is so far mainly seen in minimally invasive surgery, laparoscopy, and endoscopy (e.g., Berg, 2001; Blyth, Stott, & Anderson, 2006; Michel, 2002) but has also been extended to airway management (Myers, Mayrose, Ellis, Anand, & Kesavadas, 2005; Rowe & Cohen, 2002). These simulators are based on high-definition monitor screens and tools that enable the user to interact with three-dimensional computerized data-bases in real time, using their natural senses (see Chopra, 1996, for a more detailed description; Kneebone, 2003; McCloy & Stone, 2001).

2.2.1.5.2.1.2 Fidelity of simulators

Different classes of simulators and also the different simulators within these classes, vary along the lines of their *fidelity*. The concept of fidelity refers to the question of "how similar to the actual task situation must a training situation be to provide effective training?" (Hays & Singer, 1989, p. vi-vii). As Hays and Singer (1989) pointed out, an agreement on the meaning of fidelity has not yet been attained. They offer the following working definition (p. 50):

> "Simulation fidelity is the degree of similarity between the training situation and the operational situation which is simulated. It is a two-dimensional measurement of this similarity in terms of: (1) the physical characteristics, for example, visual, spatial, kinesthetic, etc.; and (2) the functional characteristics, for example the informational, and stimulus and response options of the training situation."

Physical fidelity thus refers to the simulator's resembling the equipment's appearance in its visual, spatial, and kinesthetic aspects. It thus has a large number of dimensions, such as size, shape, location, etc. The informational characteristics of the *functional fidelity* refer to the operational knowledge the trainee receives from physical stimuli. The functional fidelity's stimulus-response characteristics refer to the trainee's opportunities to use this information for actions in the simulator.

Another concept that is frequently mentioned in the training literature is *psychological fidelity,* which is related to the above mentioned concept of functional fidelity. *Psychological fidelity* means that participants perceive the simulator to behave in the same way as the equipment in terms of perceptual, sensory-motor, or cognitive cues (Morrison, 1991, p. 242), which in turn leads to the reproduction of those behavioral processes in the simulator that are required in the real setting, that is, on the job (Goldstein & Ford, 2002).

Although most researchers emphasize psychological/functional fidelity as the objective of simulator design, the question of how realistic the physical representation needs to be in order to attain the psychological/functional fidelity needed to lead to learning and transfer is still not answered satisfactorily (Goldstein & Ford, 2002). "The physical and functional characteristics incorporated in the training system media must be based in the actual situation, but do not have to exactly replicate that situation" (Hays & Singer, 1989, p. viii). Although some level of fidelity is essential, it is only one of many elements in the training setting that influence training effectiveness (Hays & Singer, 1989), as can be seen in the training model (Figure 1).

The effectiveness of part-scale simulators has been shown where specific behaviors or subskills are the focus of the training (Goldstein & Ford, 2002). Research in aviation has furthermore suggested that part-scale simulators can be of training value equal to that of full-scale simulators (e.g., Prophet & Boyd, 1970). Trainees acquire skills that can be transferred to full-scale simulators as well as to the job (Gopher, Weil, & Bareket, 1994; Jentsch & Bowers, 1998; Koonce & Bramble, 1998; Prince & Jentsch, 2001). Dennis and Harris (1998) showed that low fidelity simulators are specifically beneficial in early phases of the learning process.

It is thus not necessary for simulators to have complete physical fidelity or even the highest (and most expensive) degree possible, since practice, learning, and transfer are not proportionate to the degree of fidelity (see Flexman & Stark, 1987). It is more important that the simulators provide adequate training in the underlying cognitive principles of the task rather than that they possess a high physical fidelity.

As has been mentioned above, human patient simulators will never reach the fidelity levels of simulators that replicate systems comprised of only technical elements. The patient mannequin can never re-create the subtleties of the human body, which can be considered as the fundamental limitation of medical simulation (Gaba, 1998; Good, 2003; Hesselfeldt, Kristensen, & Rasmussen, 2005). The pharmacological models driving the simulator cannot, for example, incorporate the interactions among different administered drugs (Lussi, Grapengeter, & Schüttler, 1999). Another major shortcoming of current mannequins is their lack of certain important clinical cues. They cannot yet simulate skin pallor, cyanosis, sweating, or smell. This limitation can be somewhat reduced by using a confederate role-player or instructor who, for example, verbally states the current skin color of the "patient" (Lussi et al., 1999). The mannequin must thus be considered the least realistic part of the simulator.
Although it has been pointed out that participants may have problems suspending disbelief of the artificial situation (Barach, 2000), the limited physical fidelity might not be as grave a limitation as could be expected when looking at the lack of clinical signs or the overall physical appearance of the patient simulator. First of all, the reactions of participants seem to overall indicate that they find the simulations sufficiently realistic (Helmreich & Merritt, 1998). This might however, have to be attributed primarily to other elements of the simulator, such as the role-players or the operating room in which the patient simulator is set up. Secondly, as has already been mentioned above, research in aviation has suggested that effective training can also be conducted with simulators of low fidelity (Flexman & Stark, 1987). Supporting this view, Hays and Singer (1989) point out that the key to simulation is that it allows participants to "experience realities in scaled-down versions of those realities" (p. 209). The important aspect of these scaled-down versions is their functional fidelity which needs to be high enough for the participants to experience the results of their actions "realistically enough to learn from them" (p. 209). For tasks that require interpersonal interactions, such as ACRM training, physical fidelity is not deemed as important as for those which mainly require physical interaction with machines (Hays & Singer, 1989).

It can be concluded that there is no "best" type of simulator for all training purposes. Depending on the *training objectives*, an appropriate *type of simulator* from the categories of part-task, computer-screen-only, and full-scale simulators has to be chosen (cf. Chopra, 1996). In this choice it has to be remembered that different studies (see Hays & Singer, 1989) indicated that the highest possible degree of fidelity is not always necessary to design the most effective training environment. Training designers thus have to consider what degree of fidelity is most conducive to their training objectives. It is to be noted, however, that in spite of the research efforts mentioned above, the findings are still limited and their generalizability questionable (Goldstein & Ford, 2002). Clear guidelines for the simulator choice are thus not yet available.

2.2.1.5.2.2 Simulation design

Simulators as training devices are the media through which simulations can be experienced by the trainees (Morris & Thomas, 1976). *Simulation* can thus be defined as

> "the imitation of the operation of a real-world process or system over time. Simulation involves the generation of an artificial history of the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented" (Banks, 1998, p. 3).

A simulation session covering one specific topic or patient case is often referred to as a (simulation) scenario (Gaba et al., 1994). In training simulations only those features are imitated or represented, that serve to achieve the training goal (Morris & Thomas, 1976). Training simulations have a set of features, that make them extremely useful tools in the area of medical training: controllability, efficiency, safety, and instructional features (see Schmitz, 2002).

In the area of simulation training[4], simulations are used as a training method to allow trainees to practice a certain set of tasks. Among the aspects that influence the success of the simulation method are its complexity, the scheduling, type of preliminary preparation, and type of feedback (Tannenbaum & Yukl, 1992). Within the framework of this training model, two aspects of the design and selection of simulation scenarios will be further elaborated: the scenarios' content (2.2.1.5.2.2.1) and their sequence (2.2.1.5.2.2.2).

2.2.1.5.2.2.1 Scenario content

Scenarios contain different types of cases or problems to be solved. In the area of anesthesia simulation training they contain some type of case in which the physician has to administer anesthesia to a patient. Scenarios differ in their complexity, which is determined by the type and degree of tasks and/or problems contained in the scenario.

The scenarios' content has to be derived from the defined training objectives and needs assessment (cf. 2.2.1.4). Training designers thus have to script scenarios which enable the trainees to achieve the defined objectives.

Among the suggestions for designing scenarios and their content has been the event based approach to training (EBAT) (Fowlkes, Dwyer, Oser, & Salas, 1998). In EBAT, specific events are introduced in the scenarios to trigger the behaviors that are to be learned. These events must be identified in the training needs assessment phase and then scripted into the

[4] Note that based on the literature, simulator training and simulation training are used synonymously.

scenario. This approach supports the training designer in focusing on anticipating trainees' behavior in the scenarios and thus including the necessary triggers for these behaviors. Care has to be taken, that the scenarios' complexity matches the objectives.
The scenarios should furthermore allow for trainees' differences in skill levels. As Patrick (2003) emphasized, training should be adaptive to differences between trainees and scenarios should provide a range of difficulties. This implies that the different trainees should be able to focus their practice on mastering the parts of the task that are difficult for them. In a simulator environment this could, for example, be done by letting trainees assume different roles in the scenario.

In conclusion it can be stated the content of the scenarios must be derived from the defined learning objectives and must provide the respective affordances to reach these objectives.

2.2.1.5.2.2.2 Scenario sequence

If trainees participate in more than one scenario, the sequence of the used scenarios has to be considered as well. Which type of sequencing is used largely depends on the general learning theory approach. As has been mentioned there (2.1), traditional learning approaches would call for a steady increase in the difficulty level of the scenarios. They would furthermore consider how much systematic reduction of complexity and variation is useful. Constructivist approaches, on the other hand, would not call for this steady increase. But would call for scenarios that each reflect a realistic level of complexity. Nevertheless, even in the real problem setting, cases vary in their complexity and it thus seems advisable to start with cases of lower complexity so as not to overwhelm the trainees.
The cognitive apprenticeship approach has provided guidelines for sequencing the learning content and thus in the case of simulation training, the scenarios (De Corte, 2001): Increasing complexity, increasing diversity, and performing global before local skills (cf. 2.1.3.2).
From an instructional design point of view, the mentioned sequencing aspects are also related to research on practice schedules: distributed versus massed practice, whole- versus part-task training (see 2.2.1.5.1.2).
Again, these considerations need to be set in the frame of training objectives and considerations about the learners' individual characteristics. As a general guideline it can be concluded that the *sequence* of the used *scenarios* should allow for increasing complexity and variability without overwhelming the learners.

2.2.1.6 Training outcomes

As reflected in the right column of the training model (Figure 1), the goal of all training programs is the generation of a set of *training outcomes* or effects. The measurement of these outcomes is the focus of the training evaluation process. Measurement criteria must be based on the training needs analysis and linked to the training objectives (cf. 2.2.1.4).

The variety of different outcomes training programs can lead to, has been categorized by Kirkpatrick (1959; 1994) in his training evaluation model. Based on its clarity and simplicity, Kirkpatrick's four-level model is still the most widely used and influential one (Arthur et al., 2003; Salas & Cannon-Bowers, 2001) and was thus taken as the basis for this training effectiveness model.

Nevertheless, Kirkpatrick's model has also been heavily criticized. Research pointed out the model's flaws and implicit assumptions that were most likely not intended by Kirkpatrick himself when he presented the model and which might have led to the misuse of the model (Alliger & Janak, 1989). The model has furthermore been extended by several authors (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Holton, 1996, 2005; Kraiger, Ford, & Salas, 1993; see also Thierau-Brunner, Wottawa, & Stangel-Meseke, 2006).

As in Figure 1, the following sections are structured along the four training outcome levels introduced by Kirkpatrick (1994): Reaction/training appraisal (2.2.1.6.1), learning (2.2.1.6.2), behavior/transfer (2.2.1.6.3), results/organizational effects (2.2.1.6.4). The most important critical issues will be briefly described within the respective section (see Alliger & Janak, 1989; Goldstein & Ford, 2002; Holton, 1996; Kraiger & Jung, 1996, for detailed critiques of the model; see Schmitz, 2002, for a more detailed discussion of Kirkpatrick's model).

2.2.1.6.1 Reaction/Training appraisal

The first level of training outcomes is referred to as *reaction*. Kirkpatrick describes this level as the assessment of participants' affective and attitudinal responses to the training which can be operationalized by administering self-report measures directly after the training program. The reaction level thus includes the subjective dimension of training evaluation and is considered the most basic level of evaluation. Kirkpatrick also refers to it as "a measure of customer satisfaction" (1994, p. 21).

He emphasizes that the degree of satisfaction can often not only determine if the individual trainee continues to attend training but also if the whole training program is continued. Reaction measures should thus definitely be collected as it makes little sense to design and run training programs that result in unfavorable responses. Research has shown that the most commonly used criterion measures in organizations are reaction measures (Goldstein & Ford, 2002; Twitchell, Holton, & Trott, 2001).

More recent research using *reaction measures* has shown that there are multiple aspects to what is being examined. In Cannon-Bowers et al.'s (1995) model trainee *reactions* are conceptualized as reactions pertaining to *training relevance* and *value* and reactions pertaining to *affect* and *happiness*.

Based on their meta-analysis, Alliger et al. (1997) came to the conclusion that the reaction level should be divided into *affective reactions*, referring to the participants' liking of the training, and *utility judgments*, indicating if the participants found the training relevant and valuable for their job. Similar dimensions have also been found by Warr and Bruce (1995), although they additionally included *training difficulty* as a third dimension.

Morgan and Casper (2000) found six different factors of training reactions: *Satisfaction with the instructor, satisfaction with the training management/administration process, satisfaction with the course testing process, utility of the training program, satisfaction with the course materials, satisfaction with the course structure*. Morgan and Casper's utility factor is very similar to Alliger and Janak's utility factor (Alliger et al., 1997; Morgan & Casper, 2000).

In conclusion, it can be said, that trainees' reactions should be assessed as the first evaluation criterion. This conclusion is underscored by Kraiger (2002) who identified three primary reasons for evaluating training programs: decision making, feedback, and marketing. Assessing trainees' reactions can help to decide which courses should be continued and by which instructors. From the viewpoint of marketing, prior trainees' reactions can create program reputations that then influence rankings, enrollment rates or pretraining motivation of future trainees.

The problem of the reaction level's dimensionality can currently not be answered satisfactorily. More research is clearly needed to determine the underlying dimensions of reaction measures.

2.2.1.6.2 Learning

On the second level, Kirkpatrick proposed to assess the learning of knowledge, skills, and attitudes as specified in the training objectives. Learning is commonly defined as "a relatively permanent change in knowledge, skill, or attitude produced by some type of experience"[5] (Goldstein & Ford, 2002, p. 28). Depending on the training program and its objectives, one or more of these components need to be assessed. It has to be noted, that the learning level does not refer to changes in performance or behavior on the job. It is rather conceptualized as the precursor of possible overt behavior in the transfer setting.

Learning is usually operationalized by administering paper-and-pencil or skill-based performance tests and can be assessed by a pre-posttest control group design.

[5] Definitions vary according to the underlying learning theory. See Shuell (2001) for a discussion.

Newer research has further specified training effects on the learning and transfer (cf. 2.2.1.6.3) level. The problem of the learning level's unidimensionality has been approached by the research of several authors, for example by Kraiger, Ford, and Salas' classification scheme (1993) and by Alliger et al.'s taxonomy (1997). Kraiger et al. (1993) emphasize that "learning" is a much richer concept and consequently distinguish between cognitive, affective, and skill-based learning outcomes (see Kraiger et al., 1993, for a detailed review of theory and research concerning these learning outcomes, as well as for a discussion of relevant measurement issues). It needs to be pointed out, however, that although Kraiger et al.'s (1993) typology helps researchers and practitioners to link evaluation measures to certain learning outcomes, it still does not aid them in linking instructional objectives to these learning outcomes (see Kraiger & Jung, 1996).

Alliger (1997) differentiated the learning level in *immediate knowledge*, *knowledge retention*, and *behavior/skill demonstration*. This proposal could again be specified if behavioral/skill demonstration was differentiated in immediate and retention (Goldstein & Ford, 2002).

It can thus be concluded that newer research concepts highlight the need to further differentiate Kirkpatrick's (1994) learning level. However, these newer taxonomies and models are still waiting to be combined and validated by further studies.

2.2.1.6.3 Behavior/Transfer

In Kirkpatrick's model (1994) behavior refers to the trainees' performance or observable behavior on the job. It thus measures training transfer in the area of personnel development. In this area, transfer is generally defined as the extent to which KSAs learned in training are applied and generalized to other than the trained tasks, as well as the extent to which they maintained over time in the transfer or job environment (Baldwin & Ford, 1988; Klauer, 1993; Wexley & Latham, 1981). Transfer thus takes place when KSAs have been acquired in the learning environment (or "source") and are applied by the learner in the transfer environment (or "target") to a different set of problems (Holyoak & Thargard, 1989; Singley & Anderson, 1989). The transfer environment has also been referred to as the "functional environment" containing the problems or tasks to which the skills need to be transferred (Bergmann & Sonntag, 2006). Transfer can be *positive*, meaning that the learners can apply their learned KSAs to new tasks, *negative*, when the training interferes with other tasks or skills in the application setting or when the learners try to apply the new skills to tasks for which they cannot be used, *zero*, when the training does not exert any influence on task performance (Bergmann & Sonntag, 2006).

When assessing job performance, the training objectives must again be considered. This can prove difficult since most training programs are not set up to influence all job components[6].

[6] Cf. the discussion of criterion relevancy, deficiency, and contamination in section 2.2.1.4.3.

The behavioral criterion level is typically operationalized by using supervisor ratings or objective performance indicators.

Since the initial introduction of Kirkpatrick's model (Kirkpatrick, 1959), transfer of training has received a lot of research attention, the concept of transfer has been specified, and important aspects of transfer have been discussed (see e.g., Bendorf, 2001; Bergmann & Sonntag, 2006; Ford & Weissbein, 1997; Ivancic & Hesketh, 2000; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992). The following paragraphs elaborate aspects that are important in the context of this study.

In their training literature review, Baldwin and Ford (1988) presented a transfer process model. Transfer in this model specifically refers to the application of the learned KSAs to the job (Baldwin & Ford, 1988; Wexley & Latham, 1981). The model (Baldwin & Ford, 1988, p. 65) explicates that *training inputs* (trainee characteristics, training design, and the work environment), *training outputs* (learning and retention), and *conditions of transfer* (generalization and maintenance) influence transfer.

Considering the definition of transfer and thus the outcomes of training, Cannon-Bowers et al. (1995) extended Kirkpatrick's (1994) evaluation model and specified Baldwin and Ford's model (1988). They further differentiated behavior change or transfer into *performance after training* and *performance on the job*, thus specifying the function environment (Bergmann & Sonntag, 2006). This concept of performance after training might extend the notion of learning and retention as training outputs (Baldwin & Ford, 1988) because it emphasizes that trainees might be able to apply their knowledge to different tasks and settings directly after the training. And this application requires more than mere learning and retention. Nevertheless, trainees might not display the according behaviors on the job, due to a plethora of factors in the work or transfer environment (Baldwin & Ford, 1988; see also organizational characteristics, section 2.2.1.3).

Although Baldwin and Ford (1988) explicitly refer to the job setting (*performance on the job*), their two dimensions of transfer are also of great value when looking at *performance after training*: a) *generalization* that refers to the extent to which trained KSAs are exhibited in the transfer setting and b) *maintenance* that refers to the length of time the KSAs are kept up.

Another critical issue was emphasized by Laker (1990) with the concept of *transfer distance*, referring to the transfer situation's similarity to the training situation. *Near transfer* is conceptualized as the application of the learned KSAs to situations or tasks similar to those in which learning has taken place. Near transfer has been identified as the primary objective of short-term skill development that can be applied immediately in the current job (Spitzer, 1984) and of technical training (Laker, 1990). *Far transfer* is conceptualized as the application of the learned KSAs to situations or tasks dissimilar to those in which learning has taken place. Far transfer is the primary objective of training programs that are directed toward long-term goals and positions the trainees might hold in the future (Laker, 1990).

In conclusion, it can be said that Cannon-Bowers et al. (1995) contributed a much-needed clarification of Kirkpatrick's (1994) terminology by specifying more clearly the behavior level. Baldwin and Ford's model (1988) furthermore shows that transfer is a complex process that is influenced by many factors. The influence of these factors varies depending on the specific training program and training context. Concerning the different specifications of transfer it is to be noted that they often exist next to each other without having been brought in relation (cf. Tannenbaum & Yukl, 1992). It is therefore necessary to closely examine the respective training setting with its requirements and to select the appropriate transfer conceptualizations on this basis. Especially since some of the conceptualizations seem to evoke questions, such as the dichotomy of the transfer distance concept.

2.2.1.6.4 Organizational effects/Results

Kirkpatrick's (1994) highest level is called "results" and refers to the organization's objectives for the implementation of the training program. It investigates the final results for the organization that occur due to the trainees' participation in the course (Kirkpatrick, 1994). Some examples of results that could be evaluated are company profits, absenteeism, higher productivity, lower turnover rates, and an improved morale. Results are the most difficult to assess, as they are the most distal criteria, and are most susceptible to issues of criterion deficiency and contamination. Nevertheless, they are also considered the most important (Kirkpatrick, 1994).

Results can be operationalized by utility analysis methods, providing a methodology to assess the company gains due to training.

In conclusion, it can be observed that results are only rarely assessed in the training evaluation literature due to the mentioned problems of deficiency and contamination (Goldstein & Ford, 2002).

2.2.1.6.5 Conclusion

Overall, Kirkpatrick's model (1994) is valuable because it provides a framework to investigate training results on different outcome levels. Still lacking is a model that aids practitioners in deciding what to evaluate or how to link evaluation results to the organization's strategic considerations and how to operationalize the different measurement levels, or which assessment techniques to use.

Three more cautionary notes have to be made for the use of Kirkpatrick's model (1994). The first pertains to the *value of the training criteria*. One of the emerged assumptions is that a higher evaluation level always provides more valuable information. But, as follows from the section on instructional systems design (2.2.1.4), the problematic aspect of this assumption is the fact that not all training programs are aimed at influencing all four levels. There are, for example, training programs that solely aim at creating a sense of company pride (Alliger &

Janak, 1989, p. 332). These programs might best be assessed by using reaction measures only. There are other programs, for example in the educational domain, that might not be aimed at achieving behavior or results effects in Kirkpatrick's sense.

The second note concerns the *causality among the training criteria.* A critical assumption is that the four evaluation levels are causally linked. Kirkpatrick (1994) himself is not clear about causal linkages in his model. Although the question of causality is difficult to answer, the time of measurement provides helpful information. Alliger and Janak remark that when reactions and learning are assessed at the same time, it is not reasonable to assume that one caused the other. They conclude that reactions "may not be expected to cause learning" (1989, p. 334) and propose different causal links (see Alliger & Janak, 1989, p. 335f).

Holton (1996; 2005) further discusses the need to incorporate intervening variables that affect learning and transfer in any training evaluation model.

The third note refers to *interrelations among the training criteria*. The assumption of causal links leads to the assumption of positive correlations among the different levels. Research that would allow testing the intercorrelations among the criteria by collecting all four types is unfortunately still scarce. The relationship between *reaction criteria and the higher criterion levels* has been investigated in several studies and led to inconclusive results (e.g., Alliger & Janak, 1989; Alliger et al., 1997; Arthur, Tubre, Paul, & Edens, 2003; Colquitt et al., 2000). Some studies investigating correlations between reaction and higher-level criteria did not show substantial relationships (see Alliger & Janak, 1989; Dixon, 1990; Warr & Bunce, 1995). Other studies (see Alliger et al., 1997) revealed that affective reaction measures did not show significant relationships with learning measures. Utility reaction measures, on the other hand, correlated significantly with learning measures. It can be concluded that until further research has clarified the relationships, reaction measures cannot be used as a "suitable surrogate for other indexes of training effectiveness" (Tannenbaum & Yukl, 1992, p. 425). Goldstein and Ford (2002) support the latter conclusion by emphasizing that, although a good training atmosphere might be beneficial for learning processes, it does not necessarily cause trainees to learn more. One of the reasons for this phenomenon is that reaction and learning measures do not necessarily reflect the same behavioral domains. If the reaction measures assess how much the participants enjoyed the training they are not related to learning measures. Similar problems are encountered trying to empirically establish the relationship between *learning and behavior*, although these levels are conceptually linked (Alliger et al., 1997; cf. Colquitt et al., 2000). Several studies (Mathieu et al., 1990, as cited in Alliger et al., 1997; Tannenbaum & Yukl, 1992) found significant but often modest relationships between learning and subsequent performance, indicating that the used learning measures are not very predictive of skill transfer. Tannenbaum and Yukl (1992, p. 425) state that „trainee learning appears to be a necessary but not sufficient prerequisite for behavior change." This insufficiency is rooted in the behavioral criteria's susceptibility to environmental variables that can influence training transfer (Quinones, 1997). Nevertheless, learning is conceptualized to have direct effects on transfer (Baldwin & Ford, 1988). In Alliger et al.'s (1997) meta-analysis only two studies were identified that correlated learning with results finding a rather large correlation of r = .52. Since this result is only based on two studies, the nature of the relationship cannot yet definitely be determined.

It has to be concluded, that the issue of interrelations among the different training criteria has not been satisfactorily solved yet and is much more complex than suggested in Kirkpatrick's (1994) model (see Holton, 1996, for a more detailed discussion). Newer models, such as Cannon-Bowers et al.'s model (1995), do not, for example, assume a causal link between trainee reactions and learning anymore, but propose that reactions are influenced by other factors than learning, behavior, and results are: Reactions are hypothesized to be influenced by *pretraining motivation* and by the extent to which the training fulfills trainees' *training expectations*. *Learning*, on the other hand, is assumed to be affected by *individual characteristics*, such as *cognitive ability, training variables*, such as *method* and *content, training motivation*, and *expectancy fulfillment*.

Kirkpatrick's model (1994) can thus still serve as a guideline for training evaluation studies but should be extended depending on the study's goals.

2.2.2 Conclusion

The review of different approaches to learning and the description of a training model tailored to the simulation settings, showed that research in instructional psychology and, more specifically, research in the training area provide useful frameworks and findings that can be adapted to and employed in the development of training programs in the medical area. The theoretical findings provide guidelines for training designers for developing theoretically based training programs. Nevertheless, as has been mentioned in the introduction, these theoretical advances have not yet been adequately adopted by medical practitioners and teachers. This will become evident looking at the empirical findings in anesthesia simulation training.

2.3 Empirical findings: The use of simulators in anesthesia training

During the last years, the use of simulators has increased rapidly in anesthesia and other medical areas. Simulators have been used as tools in training, evaluation, and research (Wong, 2004). The following sections describe current research findings concerning the use of simulators as training tools (2.3.1) and, briefly, their use as assessment tools (2.3.2). The focus lies on the results in the field of full-scale anesthesia simulation, but findings from other medical fields and with other types of simulators will be reported where appropriate.

2.3.1 Simulators as training tools

Issenberg and colleagues (2005) recently conducted a review of the use of high-fidelity simulation in medical education, investigating medical learning in the categories of clinical skills, practical procedures, patient investigation, patient management, health promotion, communication, information skills, integrating basic sciences, and attitudes and decision-making. Their literature search yielded 109 articles on comparative studies that used simulation as an educational intervention. They found that the majority of articles were published in surgical and biomedical engineering journals, followed by anesthesia journals with approximately 16% of the publications. The articles were focused on learning of practical procedures. Three target groups can be identified in the literature of medical simulation training. Accordingly, the following sections briefly describe findings for the continuing education for physicians (2.3.1.1), for resident training (2.3.1.2), and finally for medical students (2.3.1.3).

2.3.1.1 Simulators as training tools for physicians

Most full-scale simulation training programs were originally developed in the wake of human factors research and were thus aimed at the continuing education of advanced learners (i.e., physicians) in the management of critical incidents in the OR (see Flechter et al., 2002; Schmitz, 2002; St.Pierre et al., 2004).

Evaluation efforts of these programs have focused on examining trainee *reactions* (e.g., trainee interest, perceived usefulness, and level of realism) (Chopra et al., 1994; Helmreich & Schäfer, 1994; Holzman, Cooper, & Gaba, 1993; Holzmann et al., 1995; Howard, Gaba, Fish, Yang, & Sarnquist, 1992; McDonald, Flanagan, Keast, & Gray, 2001; Tarshis, DeSousa, Brown, Iglesias, & Kohlhammer, 2005). The reaction findings showed that anesthesiologists and other physicians consider the simulation sessions rewarding and valuable, and expect to benefit from the training in their daily clinical work (e.g., Berkenstadt et al., 2003; Garden, Robinson, Weller, Wilson, & Crone, 2002; Reznek et al., 2003; cf. Schmitz, 2002; Weller, Wilson, & Robinson, 2003). Trainee *learning* and *behavior* have only been assessed to a limited degree (St.Pierre et al., 2004), but a few findings are nevertheless available from anesthesia and other disciplines, and various types of simulators (e.g., Blum et al., 2004; Flanagan et al., 2004; Good et al., 1992; Howard et al., 1992). The existing studies provide first indications that simulation-based training leads to learning (e.g., Berkenstadt et al., 2003; Flanagan et al., 2004; Gaba, Howard, Fish, Smith, & Sowb, 2001) and transfer (e.g., Chopra et al., 1994; Gaba, 1992; Gaba et al., 2001). A study with anesthesiologists showed, for example, that simulation scenarios led to improvements in managing simulated crisis situations (Nyssen et al., 2002). A recent study in the area of emergency team training showed that multidisciplinary teams performed better on team tasks after the simulation sessions and that they had higher simulated patient survival rates. The transfer of simulation-based teamwork training results to the actual emergency medicine work environment has recently been studied by Shapiro, Morey, Small, Langford, Jagminas, Suner, Salisbury, Simon, and Jay (2004). Although they found no significant interaction between group and time of measurement, the simulator-trained group showed a trend towards the improvement of quality of teamwork behaviors in the OR from pre- to posttest.

Results, such as improved patient outcomes, have not been assessed yet (Cooper & Taqueti, 2004; see Schmitz, 2002, for a discussion).

2.3.1.2 Simulators as training tools for residents

Rather common is the use of simulators in anesthesiology residents' training (Murray, Schneider, & Robbins, 1998). Issenberg and colleagues' recent review (2005) showed that simulation study participants were most often residents in the area of surgery or anesthesiology, followed by medical students. The content of these resident training programs varies from technical anesthesia skills to crisis management skills. The evaluation studies investigated outcomes on different levels as well as comparisons of various training aspects.

The first study reported with a small number of anesthesia residents was published in 1969 (Abrahamson, Denson, & Wolf, 1969). The full-scale simulator SimOne was used to train the residents in endotracheal intubation, that is, a subskill of inducing anesthesia. The study collected OR data on the *behavior* level and could show a tendency of the simulation-trained residents to achieve the accepted performance levels in fewer days and also in a smaller number of trials in the OR than the residents with traditional training.

First studies have also already provided evidence for the simulation training's effectiveness and its superiority over didactic sessions without simulators (Good et al., 1992; Ostergaard et al., 1997). An early study by Good and colleagues (1992), for example, showed that simulation training could accelerate residents' *learning* of basic anesthesia skills, such as checking and operating the anesthesia machine, inducing general anesthesia, assessing anesthetic emergence and extubating the patient, as compared to lecture formats. Comparing the use of simulation as a training method with a lecture-based method, the study of 16 anesthesia residents indicated that in the first three months, the simulator-trained residents had better clinical evaluation scores than the lecture-trained residents. Although both groups

reached the same level of proficiency after three months, the simulation-trained group had a steeper learning curve than the lecture-trained group (Good et al., 1992).
Concerning the best amount of simulation training, a recent study (Yee et al., 2005) found that anesthesia residents' nontechnical crisis management skills improved due to a single simulation session. No further significant improvements could be found after an additional session. The authors conclude that single exposure to anesthesia crises using a simulator are beneficial but that any additional sessions may not add any additional benefit.
Simulators have furthermore successfully been used in orienting and training new residents to conduct solo anesthesia in the OR (Loyd, 2004). The newly established 8-day simulation training curriculum proved superior to a traditional mentoring and lecture-based course in shortening the time the residents needed until they were able to conduct anesthesia by themselves.

Simulation training for residents in other medical areas has also produced encouraging results. Simulation scenarios have been used, for example, in emergency medicine to teach metacognitive strategies (Bond et al., 2004). The results showed that residents *reacted* positively to the training and that they had gained knowledge.
On the *learning* level, the use of simulators has furthermore been reported to enhance the development of trauma management skills in residents (Marshall et al., 2001). In the area of obstetrics and gynecology simulators have also been shown to have positive effects on resident *learning* and *performance* (Deering, Poggi, Macedonia, Gherman, & Satin, 2004).
Another study on the behavior level, a deliberate practice study, showed that residents trained with a simulator on advanced cardiac life support performed better on a clinical skills evaluation than a wait-list control group (Wayne et al., 2005).

2.3.1.3 Simulators as training tools for students

Within the last twenty years, the use of simulation technology has also steeply increased in medical student education. It has been used to improve student learning, to provide controlled and safe practice opportunities, and to improve the acquisition of the students' clinical skills (Issenberg et al., 2005). In their survey, Morgan and Cleave-Hogg (2001; 2002c) identified 158 simulation centers worldwide. Of the 60 centers that returned the questionnaire, 77% used the simulator in undergraduate education, 85% for postgraduate education. The most often taught areas in undergraduate medical training were crisis management, rare events, anesthesia induction and airway management. The specific use of simulators to educate medical students in basic anesthesia skills is thus also starting to spread throughout medical curricula (Goodrow, Rosen, & Wood, 2005; Lake, 2005; Liu & Lee, 2005).
An early example of the incorporation of a simulator in the medical student elective curriculum was reported from Stanford University (Fish & Flanagan, 1996). If the students chose the elective, they spent two half-day sessions in the simulation center. They were first

oriented to the anesthesia equipment and simulator and then had the opportunity to administer different drugs and observe their effects on the simulated patient. At the end of the first session, the students performed an anesthetic induction as a team. In the second session, the students worked on two cases. For each case, they prepared an anesthetic plan, discussed it, and then administered the anesthetic of their choice. After the induction of anesthesia, a critical event was simulated. The sessions were evaluated using a pre-posttest questionnaire design. The students' *reactions* were assessed in the questionnaire. The results showed that the students perceived the simulation sessions as suitable for their stage of training, thought their time was well spent in the simulator, and found that the format aided their understanding of individual drugs and their interaction. The students specifically appreciated the opportunity to take responsibility for all aspects of the anesthetic.

McIvor (2004) also reports the use of simulators in anesthesiology training for medical students at the University of Pittsburgh School of Medicine since 1994. Since 2002 the Medical school has incorporated the anesthesiology simulation training in their anesthesiology elective curriculum for senior medical students. Unfortunately, he reports only anecdotal data in regard to the students' *reactions* to this training. He states that the students perceived the scenarios as relevant, realistic, appropriately difficult, and adequately debriefed. The studies also indicated that they had received enough instruction and opportunity to practice.

Positive *reaction data* have furthermore been reported early on by Freid (1998) and more recently by Cleave-Hogg and Morgan (2002). The latter study found that the students considered the session a positive experience which provided opportunities to apply their knowledge in a realistic environment. A few students indicated that they were not comfortable in the simulation environment, but this did not seem to affect their performance.

More recently, Flanagan, Nestel, and Joseph (2004) described an undergraduate curriculum with the focus on patient safety in Australia. The courses follow the classic outline of ACRM courses (see Schmitz, 2002) and comprise didactic sessions on ACRM principles, a videotape of a simulator re-enactment of an aviation crash, simulation sessions followed by video-based debriefing sessions, and a review, summary, and evaluation of the course. The reported *reaction* data were collected from 132 final-year medical students. The post-test only qualitative results indicated raised awareness and understanding of ACRM issues and high levels of satisfaction with the course. The students found the learning experience valuable, an effective way to learn, and relevant.

To investigate the superiority of simulation training over other training methods, Morgan and colleagues compared simulator-based with video-based anesthesia training (2002a). The students' *performance* in the simulator was assessed in a pre- and posttest with the help of behavioral checklists. Furthermore, students' *learning* was assessed in a written final examination. The results showed an improvement in scores from the pre- to the posttest. No significant differences were found between the two different kinds of training in the posttest and the final examination. The only significant difference was found in the collected reaction measures: Students in the simulator-based training group found training more enjoyable and valuable than students in the video-based training group. The authors concluded that both, simulator and video faculty-facilitated training, offered valuable learning experiences.

Studies in other medical areas have also reported the use of simulators for medical student education and provided evidence for the simulators' usefulness as a teaching tool (Sinz, 2005).

First-year medical students were, for example, introduced to (respiratory) physiology with the help of simulations (Euliano, 2000; Good, 2003; Gordon, Oriol, & Cooper, 2004; Lampotang, Öhrn, & van Meurs, 1996; Lee, 1998; Tan, Ti, Suresh, Ho, & Lee, 2002), second-year students encountered a PBL case through a simulator (Winston & Szarek, 2005), and fourth-year students were instructed in cardiovascular haemodynamics (Good, 2003; Öhrn, van Meurs, & Good, 1995).

Positive student *reactions* were obtained, for example, in studies by Gordon (2000) to pathophysiology simulation training and by Gordon and colleagues (2004), as well as by Tan and colleagues (2002) to respiratory physiology training. Another study (Gordon, Wilkerson, Shaffer, & Armstrong, 2001) assessed the *reactions of students and educators* to high fidelity patient simulation in emergency medicine and also found overall very positive reactions on parts of both, the students and the educators.

An example of how full-scale simulators can be combined with problem-based curricula was provided by Winston and Szarek (2005): In their second year, students had to use their communication and physical examination skills to solve a medical problem presented with the help of the simulator. The collected *reaction* data showed that the simulator was helpful in enhancing the students' understanding of the problem, that students found the sessions useful for integrating and applying their previously acquired knowledge, and for identifying gaps in their knowledge.

Objective learning data on physiology was collected by Tan and colleagues (Tan et al., 2002). They administered a physiology test before and after a simulation session and could show an increase in student knowledge after the simulation session.

A course in the management of medical emergencies using a medium fidelity simulator was described by Weller (2004). Students were trained in a realistic medical emergency where they could apply their knowledge of resuscitation. The study collected student *reactions* and *self-reported learning* data (n = 33). The results indicated overall positive reactions to the course. The students found the course to be a valuable learning experience, felt that they had learned how to approach a problem better, and had learned how to apply their theoretical knowledge. Concerning student learning, the study found a significant increase in students' self-rated competencies regarding the taught material. The study further showed that even medium fidelity simulators are well received and accepted as learning tools by medical students.

In the area of surgical clinical basic skills training, a study by Schmidts and colleagues (1998) collected data on the *reaction*, *learning*, and *transfer* level. Data showed that the students reacted positively to the training, that there was an increase in self-reported learning due to the course, and that the students performed well in a following OSCE (Objective Standardized Clinical Examination). In the posttest, there were no differences between students with and without clinical internship experience prior to the course on either of the criteria.

Comparing simulation training to other methods, Mueller and colleagues (2005) reported positive student reactions to the use of simulators in teaching antiarrhythmic therapy within the "basics of drug therapy" block. Students in the simulator group rated the course's ability to link theory and practice higher than students in a comparable lecture course.
Preliminary data on the learning level showed that students trained in the simulator retained key clinical knowledge better than students who received traditional training (Gordon, 2000).
In a randomized controlled study, Steadman and colleagues (2006) provided evidence that simulation-based training in acute care critical assessment and management skills was superior to problem-based learning. They tested students' performance in the simulator following simulation training or problem-based learning training. Not only did the simulation-trained group perform better in the assessment scenario, they also had a higher increase (pre-posttest) in skills than the problem-based learning group.
Comparing simulation-trained and video-trained emergency medicine students, Spillane, Spencer, and Maddow (2004) found no difference in performance (stabilizing and managing a patient in a chest pain scenario).
A study with paramedic students trained in endotracheal intubation compared OR training on live patients with simulation training (Hall, Plant, Bands, Kang, & Hall, 2004; Hall et al., 2005). The students first received didactic and mannequin (that is, part-task) training and were than randomized to receive either human patient OR training or full-scale simulator training. In the following learning and transfer test of patient intubations in the OR, the simulator-trained students achieved comparable results to the students trained on live patients. The results thus show that simulators can be successfully used to train endotracheal intubation (Hall et al., 2004; Hall et al., 2005).

2.3.2 Simulators as assessment tools

Simulators have also been used not primarily as training tools but as assessment tools to rate students' and physicians' diagnostic and therapeutic skills in anesthesia as well as in other areas. Their reliability and validity as assessment tools could be shown in anesthesia and other medical fields (e.g, Boulet et al., 2003; Byrne & Jones, 1997; Devitt, Kurrek, & Cohen, 1998; Devitt, Kurrek, Cohen et al., 1998; Forrest, Taylor, Postlethwaite, & Aspinall, 2002; Girzadas, Harwood, Clay, Rzechula, & Caris, 2006; Morgan & Cleave-Hogg, 2000; Morgan, Cleave-Hogg, Byrick, & Devitt, 1998; Morgan, Cleave-Hogg, DeSousa, & Tarshis, 2003; Morgan, Cleave-Hogg, Guest, & Herold, 2001; Murray et al., 2002; Murray, Boulet, Kras, McAllister, & Cox, 2005; Rogers, Jacob, Rashwan, & Pinsky, 2001; Schwid et al., 2002; Weller et al., 2003).

Devitt and colleagues (2001), for example, provided evidence for the simulators' construct validity as an assessment tool: A simulation scenario was able to distinguish between the practice categories, that is, between medical students, final year residents, and anesthesiologists. Overall, examinees with more advanced training obtained higher scores. This finding was corroborated by other studies in anesthesia and other medical fields.

Assessment studies furthermore showed that students reacted positively to the simulator, and reported that they had learned something from this assessment experience (Morgan et al., 2001). Savoldelli and colleagues (2006) found the assessment of senior anesthesiology residents' performance in the simulator a useful adjunct to the traditional oral examination.

Conclusion

The mere number of published articles (Issenberg et al., 2005) and the above reported findings show that simulators are by now being widely used in different medical domains.

Concerning the training content and the training target groups, the literature shows that simulation training can be used for a variety of training topics, from basic skills training to human factors training, and for a variety of target groups with differing skill levels. For training medical students in anesthesia, simulation training is able to provide the students with clinical applicability of their knowledge, to teach topics that are difficult to teach with other methods, and to provide a natural and safe bridge between didactic and clinical work (Lee, 1998).

Concerning the evaluation of these training programs, there is only a limited amount of data available. The data that is published, and thus available, is mainly positive. Evidence has been provided that simulation training leads to positive training outcomes on Kirkpatrick's (1994) reaction, learning, and behavior levels. There is also evidence that simulation training offers advantages over other training methods, such as video-based or problem-based lecture formats. Simulators have furthermore been shown to be reliable and valid assessment tools for diagnostic and therapeutic skills.

With a closer look at the research, it needs to be stated, however, that these evaluation efforts can only be seen as preliminary evidence because strenuous efforts are still largely missing. This statement is in line with Bradley's (2006) review of the history of simulation in medical education, who concluded that: "at the present time the quantity and quality of research in this area of medical education is limited" (p. 254). Especially the paucity of randomized, controlled experiments in the simulation-based medical education literature has to be lamented (Morgan & Cleave-Hogg, 2001; Weller, 2004). In the same vein, Issenberg and colleagues (2005) concluded that although simulations are now widely used in medical education, the literature on the use and effectiveness of simulation technology is scarce, inconsistent, and varies in regard to the methodological rigor. They found less than 20% of the reported results to be clear and "likely to be true" (p. 19). Also, they concluded that more rigorous and high-quality research is needed in the field but that the existing research nevertheless provides first indications that simulation training in medical education is effective and complements medical education with real patients, if appropriate instructional principles are observed.

From an instructional point of view, it has furthermore to be emphasized that although anecdotal papers describe the inclusion of simulation-based training into existing curricula, these papers do not report the design of these courses based on instructional models. Studies investigating the usefulness of instructionally different simulation sessions are also lacking. Neither did the conducted literature search yield any studies that investigated trainee charateristics and their relation with training outcomes.

2.4 Conclusion and research question of the present study

The presented theoretical advances in instructional psychology and the presented training model show that the training literature provides valid guidelines for the design and evaluation of training programs.

The review of existing training literature in the area of medical simulation training and, more specifically, introductory simulation-based anesthesia skills training for medical students, revealed that training programs developed and evaluated based on principles of instructional psychology are still missing. This is especially problematic for the German simulation community, as the German Anesthetic Association has provided teaching hospitals with full-scale simulators to be used in anesthesia medical education on the condition that the teaching hospitals would evaluate their use.

Internationally, the need for basing simulation curricula on sounder theoretical foundations has, however, been recently acknowledged (Bradley, 2006). Based on their review Issenberg and colleagues (2005) identified features and uses of high-fidelity medical simulations that lead to effective learning. The simulations' features they found to provide for effective learning were the following: Provision of feedback, repetitive practice, curriculum integration, range of difficulty level, adaptability to multiple learning strategies, capturing a wide span of clinical conditions, controlled environment, individualized learning with learners being active, defined objectives and outcomes, and simulator validity. The suggestions and features that have been shown to affect learning in this area (Bradley, 2006; Issenberg et al., 2005) are thus all contained in the model described above (sections 2.1 and 2.2).

Based on the identified gaps, the following study thus aimed at providing a practical model for the design and evaluation of simulation-based anesthesia training for medical students. The general theoretical approaches to learning as well as the presented training effectiveness model will serve as a theoretical basis for the study. The research question was how the mentioned principles that have been investigated in prior research can be included in a training program for anesthesiology students, if the incorporation of these principles leads to better training outcomes, and whether specified individual characteristics affect the anesthesia simulation training outcomes. The focus of the project was thus twofold: On the one hand, the focus was on the development of a pragmatic yet theoretically based curriculum design model that could be used by medical practitioners without detailed theoretical knowledge. On the other hand, this curriculum should also be evaluated based on current psychological training research and its advantages over an instructionally not optimized curriculum should be demonstrated. Furthermore, the potential influence of a set of individual characteristics identified in the training model should be explored since the existing literature has not yet investigated them. Again, the presented theoretical findings served as the underlying guidelines. In contrast to other studies, it was not intended to investigate the validity of the theoretical model per se or test all of its elements and their relations.

3 The present study: Instructionally based design and evaluation of a basic anesthesia skills simulation curriculum

The following part of the dissertation describes how the present project was conducted. The part begins with Chapter 3.1 describing the setting of the current study at the University of Heidelberg's medical school. Chapter 3.2 continues with detailing the anesthesiology simulation curriculum as it had been implemented before the start of the project. The development of the instructionally revised anesthesiology simulation curriculum is presented in chapter 3.3, followed by a description of the evaluation study's methodology (3.4).

3.1 The study setting

The project was undertaken as a cooperation between the University of Heidelberg's industrial and organizational psychology department and the clinic of anesthesiology and could build on earlier projects that had been conducted together (Schaper, Schmitz, Grube, Graf, & Dieckmann, 2002).

The following sections describe the general setting of this project. Section 3.1.1 describes the reformed curriculum at the Heidelberg Medical School, section 3.1.2 describes the clinic of anesthesiology's simulation center, and section 3.1.3 presents the conclusion of this setting for the current study.

3.1.1 The reformed medical school curriculum at the University of Heidelberg

As has been mentioned in the introduction, medical school curricula saw widespread reforms throughout the last decades. The medical school curriculum at the University of Heidelberg was reformed in October 2001. It was changed from a traditional curriculum in which the focus lay on conveying theoretical and detailed factual knowledge to students in a mainly lecture-based form and separately teaching the different medical subdisciplines, to a problem-based curriculum: The new clinical curriculum, called HEICUMED (Heidelberger Curriculum Medicinale). About 150 students each semester start HEICUMED in their third year of medical school. HEICUMED focuses on patient problems as a context for students to acquire problem-solving and clinical skills as well as theoretical knowledge about basic and clinical sciences (see Albanese & Mitchell, 1993, for a detailed discussion). The curriculum is divided into five thematic blocks, which are themselves divided again into four or five modules that last two to four weeks (see H.-G. Sonntag, 2003, for a further description). The students rotate through these modules in groups of differing sizes. With the implementation of the new curriculum, anesthesiology was for the first time included in the undergraduate medical curriculum, namely in the curriculum's second block. This block comprises the subjects of surgery, orthopedics, urology, and anesthesia/emergency medicine. It lasts fourteen weeks and ends with objective structured clinical examinations (OSCE) in the last week. The anesthesia and emergency medicine content is contained in two weeks of the block. The OSCE cover the complete content of the block in different stations.

Since this curriculum is relatively new in the medical school, many different evaluation efforts are in existence to document the students' reactions to the curriculum's elements. The students of the medical school participate in the curriculum's elements and their evaluation on a mandatory basis. For the current study, this implied that the participants had already filled out many questionnaires by the time they participated in our study.

3.1.2 The HANS Simulation Center

The clinic of anesthesiology runs a full-scale simulation center, the HANS Simulation Center. The HANS (Heidelberg ANesthesia Simulation) facilities center around a fully instrumented mannequin, the METI (Medical Education Technologies, Inc.) Human Patient Simulator™, on which anesthetic procedures can be conducted and which consists of hard- and software components. The mannequin is set up in a fully equipped operating room (OR) which is connected to a control room. All of the clinic of anesthesiology's full-scale simulation sessions take place at the HANS center. The HANS center is staffed with anesthesiologists and student workers. The physicians work there in addition to their clinical work or as part of their teaching duties for the anesthesiology department.

Figure 2. HANS full-scale simulator with instructor and trainees

The mannequin physically represents an adult male (see Figure 2) and features clinical signs, such as chest wall motion, a full set of palpable pulses, heart and breath sounds that are broadcast over appropriate areas of the chest, a thumb that can be stimulated by clinical nerve stimulators, eyelid and pupillary responses, carbon dioxide exhalation, and the transmission of vocal information from the control room, allowing the mannequin to talk. The METI™ mannequin does not allow for the simulation of surgical procedures, and of certain clinical cues, such as skin color and sweat, and is restricted in its mobility.

As a full-scale simulator, the HANS simulator aims at realistically replicating the complete anesthesia process (Chopra, 1998). For this purpose, the mannequin is set up in a fully equipped operating room with authentic equipment whenever possible, such as standard monitors and anesthesia equipment (see Figure 2). This allows the students to give anesthesia under maximally authentic conditions and to thereby practice a variety of tasks, such as complete airway management including mask ventilation, tracheal and bronchial intubation. The mannequin lies on an OR table and is hooked up to the anesthesia machine, ventilator, and the sensors for blood pressure, heart rate, and blood oxygen that are attached to a regular anesthesia monitor showing the patient's heart rate, blood pressure, electrocardiogram, blood oxygen saturation, and the amount of carbon dioxide exhaled.

The mannequin's reactions are controlled by a microprocessor computer located in an adjacent control room. The software determines the simulator's cardiac and respiratory physiological and pharmacological responses on the basis of mathematical models. Among the physiological parameters that can be generated are the following: Electrocardiographic information, invasive pressures, non-invasive blood pressure, cardiac output, and temperature. The gas-exchange lungs are controlled by the computer and allow respiratory functions to be simulated, such as spontaneous breathing, manual (with a breathing bag) or mechanical ventilation. Patients of either sex, any adult age, and varying degrees of illness can be modeled. The software contains several preprogrammed default patients with different physiological and pharmacodynamic settings (Chopra, 1998).

Responsible for monitoring the simulator's functions is the simulator operator. He is located in the adjacent control room and operates the computer which controls the patient simulator. The operator has full control over the default settings of the computer software running the scenarios and can consequently tailor the reactions of the simulator to the input of the students. The course of the scenarios is thus regulated depending on the treatments initiated by the students.

To be able to observe the simulation, the operator is provided with a camera overview of the OR and a view into the OR through the one-way mirror (see Figure 2). The simulator operator is connected to the instructor in the OR via intercom. The combination of these features allows the operator to closely monitor the scenario and react immediately to occurring problems.[7]

[7] The interested reader is referred to Good (2003) for a more detailed description of patient simulators.

3.1.3 Conclusion

The study setting provided advantages and disadvantages for the present study. Among the advantages were a project team who was experienced in the cooperation due to previous studies undertaken together, a fully functional and well-established simulation center, and a large pool of participants due to the mandatory anesthesiology curriculum. On the other hand, this mandatory curriculum implied the disadvantage that the study could not be designed based on research considerations only, but had to allow for compromises due to the restrictions of the mandatory curriculum.

The study started in the summer semester of 2003. An original anesthesiology curriculum was already in use. It had been taught since the clinical curriculum had been changed into the problem-based HEICUMED. During the cooperation project this existing curriculum was revised.

3.2 Description of the original anesthesiology curriculum

Within the anesthesia and emergency medicine's two-week module, the curriculum comprised two elements that the cooperating anesthesiologists were responsible for. These elements were a preparatory seminar introducing the students to the subject of anesthesiology (3.2.1) and two kinds of simulation sessions (3.2.2). To assess these original anesthesiology elements as they were taught in the first semester of the study, two anesthesiology instructors were interviewed, existing training materials were reviewed and the curricular elements were observed. These assessments were part of the needs assessment phase described later in section 3.3.1.1.

3.2.1 Preparatory introductory anesthesiology seminar

The original preparatory seminar was a three-hour lecture-based seminar that was placed in week one of the anesthesiology curriculum. It was taught on Tuesday mornings. The covered contents were focused on the three functions of anesthesia (hypnosis, analgesia, and relaxation) and introducing the different drugs that are used for these functions. For each drug, the effects, the side effects, the contra indications, and the dosage were covered. The number of drugs covered in the seminar varied depending on the instructor and the respective seminar.

The seminar showed characteristics of cognitivist learning settings: The teacher had the active role of presenting the material; the students were in a receptive role and mainly took notes on the presented material.

Overall, the focus lay clearly on presenting the different drugs. Learning goals had not been clearly identified, there was no standardized lesson plan, and, consequently, the seminars varied depending on the instructor teaching them.

3.2.2 Simulation sessions

The curriculum contained two types of simulation sessions: Full-scale complete-task simulation sessions (3.2.2.1) introducing the students to anesthesiology, and part-scale part-task simulation sessions (3.2.2.2) covering emergency medicine topics.

3.2.2.1 Anesthesia full-scale simulation sessions

The anesthesiology curriculum originally contained two full-scale complete-task simulation sessions. The simulation sessions were conducted at the HANS simulation center and used the METI Human Patient Simulator™. Each simulation session lasted one hour. They were scheduled in week one of the anesthesiology curriculum, on Tuesday afternoon and on Friday morning. The students rotated through the simulation sessions in groups of 8.

Each simulation session was in itself partitioned into two scenarios. Groups of four students participated actively in one scenario and followed the other scenario as observers from the control room.

Other than these formal standards, no teaching or learning standards had been defined. The team of instructors was composed of experienced simulation personnel. They ran the simulation sessions based on their experience and chose different foci. Some of the instructors focused on applying the theoretical knowledge of the preparatory seminar, other instructors focused more on teaching basic skills, such as intubation and mask ventilation. The scenario content was furthermore not standardized. Some instructors covered only the induction of basic anesthesia to healthy patients, others also included critical incidents during the anesthesia induction. The amount of support the students received during the scenarios also varied. Some instructors stopped the scenarios frequently to discuss the steps taken by the students, others focused on letting the students complete the whole induction process and then reflected on the scenario.

3.2.2.2 Emergency medicine part-scale simulation sessions

The original curriculum furthermore contained four three-hour part-scale part-task simulation sessions covering emergency medicine contents. These sessions were also conducted in groups of eight students.

The simulators used for these sessions were Laerdal's SimMan™ and Mega-Code Trainers. These part-scale simulators are down-scale versions of the full-scale Human Patient Simulator™. They offer the great advantage that they can easily be moved. They feature a realistic head and airway and thus allow for airway management exercises. They can furthermore be auscultated, ECGs can be taken, and defibrillators used. They do not possess computer-driven physiological models that generate the vital parameters. These parameters can either be used as they are contained in preprogrammed scenarios or can be entered into the computer by the simulator operator.

Session 1 covered primary diagnostics, electrocardiograms, and defibrillation; session 2 covered reanimation cases; session 3 covered specific emergencies, such as multiple trauma; and session 4 covered megacode training[8]. These part-task sessions thus introduced the students to skills that are also needed in the area of anesthesia, such as mask ventilation, intubation, reading electrocardiograms, and formulating diagnoses. As part-task training sessions, they did, however, not simulate the complete task process of inducing and maintaining anesthesia. Nor did they specifically cover anesthesia topics, such as the necessary drugs, or the procedures of inducing anesthesia in an OR. Due to the part-scale simulators, the students could furthermore not communicate with the simulator.

These part-scale simulation sessions could not be influenced as part of the study. The only changes that could be made were changes regarding their place in the curriculum.

3.2.3 Conclusion

Overall, the original curriculum was not standardized. The preparatory seminar was mainly teacher-centered and tried to convey a lot of factual knowledge in a lecture-based format. The simulation sessions did not reflect a specific instructional strategy. Together with the medical school setting and the simulation center, the existing curricular elements thus offered a good starting point for the development of a revised anesthesiology curriculum that was based on instructional psychology considerations and on the training model presented in the theory part of this dissertation.

3.3 Development of the revised anesthesiology curriculum

The following sections describe the first focus of the project, namely the development of the revised simulator-based anesthesiology training program. The development of the program was conducted according to the principles specified in the instructional model (Schaper, Schmitz et al., 2003) that has been presented in the theory part of this dissertation. Section 3.3.1 describes the instructional systems design tasks undertaken in the beginning of the redesign process. This step was followed by the design of the training interventions (3.3.2). The developed new training elements then had to be taught to the medical instructors. The respective instructor training is presented in section 3.3.3. The description of the revised anesthesiology curriculum's development ends with a conclusion for the evaluation of this curriculum (3.3.4).

3.3.1 Instructional systems design tasks

As described in the previously presented model (Schaper, Schmitz et al., 2003), the training designer first has to complete a set of instructional systems design tasks. The tasks relevant

[8] Megacode training covers resuscitation algorithms following the guidelines published by the International Liasion Comitee of Rescuscitation (ILCOR) or the American Heart Association (AHA).

for this study are described in the following sections: The first task was a needs assessment including an organizational as well as a person and task analysis (3.3.1.1) that resulted in the definition of the learning objectives for the revised training program (3.3.1.2). The next task in the instructional systems design process was the explication of the evaluation criteria (3.3.1.3). Based on the explicated evaluation criteria, preliminary general evaluation considerations are presented (3.3.1.4). These preliminary considerations provide the source for the final design of the evaluation study which will be described after the training design has been introduced (3.4.2).

3.3.1.1 Needs assessment

As the first phase in the training development process, a needs assessment was conducted to gather the basic data necessary for the design of the training program. Two elements of the needs assessment were considered appropriate for the current context: In the first step, data concerning the training's organizational context (see also section 3.1) were collected in an organization analysis to gather information on the constraints and allowances for the training development and evaluation. In a second step, a task and person analysis was performed. In a third step, the existing anesthesia curriculum was investigated and yielded the results described above (see section 3.2).

The first two needs assessment steps were conducted in form of structured interviews with two anesthesiology instructors who have worked in the area of simulator-based training in anesthesiology for several years. These two instructors served as the subject matter experts (SMEs) who were involved in the complete design, development, delivery, and evaluation process.

The following table (Table 5) shows the topics that were covered in the needs analysis and provided the structure for the conducted interviews.

Table 5. *Guideline for the needs assessment aspects investigated in the structured interviews*

Organization analysis
• Description of the existing anesthesiology curriculum with its components and interdependencies, allowances, and restrictions,…
• Possible intervention points for the current study (e.g., which components will be under the control of the study, …)
• Background of the participating faculty/instructors
• Securing of support and participation of faculty/instructors
Task and person analysis
• Target group of the intervention
• Tasks to be included in the training program
• Target group's knowledge, skills, abilities

Based on this structural guideline, the interviews yielded the following results.

Organization analysis:

- The organization analysis revealed that the existing anesthesiology curriculum had already been in place since October 2001 when the medical school curriculum was reformed. This curriculum was embedded in the surgery block (as described above).
- Overall, two curricular elements in three curricular slots were available for the redesign process as they were conducted by the anesthesia department. The first slot was a three-hour seminar offering an introduction to anesthesiology. The second element in the second and third slot available for redesign were two one-hour full-scale simulation sessions.
- A change in the sequence of some of the curricular elements would furthermore be possible. Since the sequence of the different curricular elements is partly dependent on the availability of medical personnel who is also involved in clinical work, it was not possible to determine the scope of possible changes at the beginning of the study. It became obvious that the study would have to be planned in two phases.
- The different curricular elements were taught by anesthesiologists who had for the most part not received any prior training in the area of didactics.
- The SMEs would secure the support and participation of all of the instructors teaching the mentioned curricular elements.

Task and person analysis:

- Third-year medical school students need to be trained in basic anesthesiology skills. Basic anesthesiology skills refer to the induction of a standard anesthesia, for which theoretical knowledge concerning the used drugs as well as basic science knowledge is necessary, and for which psychomotor skills are necessary, including ventilation and intubation skills.
- Based on experiences with previous student cohorts, it was assumed that, on average, these learners have no prior experience in the field of anesthesiology. Concerning their theoretical knowledge, the students should be able to build on the basic science knowledge they acquired during the first two years of the medical school curriculum.
- The SMEs reported that the students' knowledge upon entering the first full-scale simulation sessions was not adequate. The students did not seem to transfer enough of the preparatory seminar's content.

3.3.1.2 Learning objectives

Based on the results of the needs assessment, the following learning objectives for the simulation sessions were derived together with the SMEs. They are listed in the following table (Table 6).

Table 6. *Learning objectives for the full-scale simulation sessions*

Having participated in the full-scale simulation sessions, the students are able to …
1. independently induce anesthesia (standard and rapid sequence), that is, • they know which drugs to use, • they know the appropriate dosage of the drugs, • they can ventilate the simulator with the mask, • they can determine the right point of time for the endotracheal intubation, • they are able to perform an endotracheal intubation.
2. assess the patient's medical history, to clinically examine the patient, and to interpret the monitoring data. Based on this information, they are able to deduct and explain the following preliminary diagnoses: • Cardiac infarction, • Asthma attack, • Anaphylaxis, • Pulmonary embolism, • Bleeding.
3. develop the above mentioned differential diagnoses.
4. develop therapeutic plans for the respective differential diagnosis, that is, they • know the respective drugs to be administered, • know their respective dosage, • are able to name the drugs' side effects, • are able to check for a possible contraindication by actively collecting the required information.
5. use their interdisciplinary knowledge for the deduction of the differential diagnoses.
6. extract the relevant information through communicating with the patient.
7. efficiently distribute tasks among the team members.

As the list above shows, the learning objectives contain cognitive, interpersonal, and psychomotor goals (Goldstein & Ford, 2002).

The learning objectives are to be achieved at the end of the full-scale simulation sessions. The preparatory seminar was supposed to provide the theoretical knowledge necessary to later achieve the learning objectives in the simulation sessions.

Together with the needs assessment, these learning objectives served as the basis for the definition of the evaluation criteria, the development of the evaluation design, and the development of the interventions, as will be explained below.

3.3.1.3 Evaluation criteria

Based on the learning objectives and the theoretical considerations of the model's evaluation part (cf. 2.2.1.6), it was decided to assess the training program's effectiveness by measuring students' reaction to the training, their learning, and their behavior, that is, the transfer due to the training. The fourth evaluation level, results (Kirkpatrick, 1994), could not be employed in the current context because data could only be collected while the students were participating in the surgery and anesthesiology block of the curriculum.

Concerning the measurement of the learning and transfer evaluation criteria, the following arguments have to be considered. Kraiger and colleagues (1993) suggested behavioral observations, hands-on performance measurement or structured situational interviews to measure the development of skills. Sonntag and Schäfer-Rauser (1993) point out, however, that self-assessment of vocational skills can also be used for the purpose of training evaluation. According to their studies, it is possible to measure skill-based learning outcomes by using self-report questionnaires that assess the self-concept of vocational competencies.

Based on these findings, learning as a training outcome of the current study should be assessed by using self-report data and, where possible, more objective data sources as well.

More objective data sources seem especially helpful when trying to assess training transfer. Ford and Weissbein (1997) pointed out that self-report measures might be inadequate for drawing conclusions about training transfer. It was therefore decided to use more objective measures in form of exam grades. It has to kept in mind, however, that these grades carry higher dangers of criterion deficiency and contamination.

Taking into account the previous research in the area of anesthesia simulation training with its lack of the investigation of personal characteristics, the current study should also exploratively assess a set of individual characteristics and determine their influence on the training outcomes. Theses individual characteristics to be assessed should be derived from the presented training model.

To determine the possibilities of assessing these evaluation criteria, the following training evaluation considerations were undertaken.

3.3.1.4 Training evaluation considerations

This part of the dissertation explicates the general design considerations (3.3.1.4.1) that were going to serve as the basis for the final evaluation design of the study. Within the process of these general considerations, and specifically when considering the use of self-report data, the potential threat of the response-shift bias became evident, as will be described in section 3.3.1.4.2. Finally, the conclusion from these considerations is presented in section 3.3.1.4.3.

3.3.1.4.1 General evaluation design considerations

Several designs are available to evaluate training programs (e.g., Goldstein & Ford, 2002; Wottawa & Thierau, 1998). Fully controlled, randomized experimental designs offer the highest degree of internal validity. In applied settings, the use of experimental designs is oftentimes not possible due to the encountered organizational restrictions (Wottawa & Thierau, 1998). As the description of the general study setting and the medical school curriculum (3.1.1) showed, this was also true for the present study.

The following restrictions became evident when the study context was analyzed:

First of all, no control group without any intervention could be used. The group that was used as a comparison group instead, was the first group of students who participated in the original mandatory curriculum (see 3.2) that had previously been developed without instructional design considerations (Training group 0).

Secondly, the only elements that could be influenced for the purpose of this study were the full-scale simulation sessions, the preparatory introductory anesthesiology seminar, and to a limited extent the sequence of the curricular elements. The part-scale simulation sessions in emergency medicine and the OSCE at the end of the block could not be experimentally manipulated.

Thirdly, at the beginning of the cooperation it could not be finally determined which elements of the curriculum could be included in the study. These uncertainties were due to curriculum planning considerations. The study thus had to be designed in two phases because the sequence of the different curricular elements could only be determined shortly before the start of the new semester. Plans could not be made earlier because of the restricted availability of the teaching staff whose primary tasks are clinical. Their teaching tasks thus have to be scheduled with regard to their clinical responsibilities.

Fourthly, the frequency of the data collection would have to be restricted because the students already had to fill out evaluation questionnaires as part of a general evaluation of the reformed curriculum. This implied that maximally three points of measurement could be used: A pretest, a posttest, and a transfer-test.

Fifthly, to measure the evaluation criteria, self-report data would have to be the primary source of information for the pre- and posttest, due to limited personnel and time resources. It was, however, possible, to also use the students' OSCEs as a transfer-test, thus yielding more objective ratings by observers.

3.3.1.4.2 Response-shift bias

In general, training research relies heavily on the use of self-reports to measures change in knowledge, skills, and abilities. This change is most of the time assessed in pre-/posttest designs. The assumption underlying the use of these designs is the stability of intra-individual standards of measurement from one point of measurement to the other(s). This assumption may, however, not be valid, since one of the problems associated with the use of self-report measures for evaluation purposes in pre-posttest designs is the occurrence of the so-called *response-shift bias* (Howard & Dailey, 1979). Sprangers and Hoogstraten (1989) define response-shift bias as "a change in a subject's internal standard for determining his or her level of functioning on a given dimension" (p. 265), that is, a scale recalibration[9]. As a result of this change of standards, conventional pre-/posttest designs using self-report measures as outcome criteria might not assess change appropriately. If subjects actually changed their internal frame of reference during the time between the ratings, the posttest scores would be a combination of the change in standard and of any change in the rated dimension. Posttest scores would thus be confounded and could not clearly assess intervention effects.

The challenge of detecting an occurring response-shift can be confronted by imposing design changes. Howard and Dailey (1979) suggested to include, as another set of ratings, a *retrospective* or *then test*. For this purpose, subjects are asked for two ratings per item in the posttest questionnaire. For the first rating, subjects are instructed to state their agreement with the item as it is at the moment ("*post"*). For the second rating, they are instructed to rate their agreement as it was before the intervention ("*retrospectively"* or *"then"*). The actual response-shift is assessed by investigating the difference between the pre-test ratings and the retrospective ratings within the experimental group. If the intervention changed the subjects' standard of measurement, they will retrospectively rate the items differently than in the pretest – usually more conservatively. It is assumed that subjects use the same standard for the retrospective pretest ratings and the posttest ratings, thus eliminating the response-shift. A comparison of retrospective and posttest ratings then gives an unbiased estimation of the training effects.

Sprangers and Hoogstraten (1989) examined possible causal determinants for the response-shift phenomenon. They came to the conclusion that subjects might lack sufficient information to adequately rate their level of functioning before the intervention, that is, during the pretest. The intervention oftentimes influences the participants' awareness of the target concept (Sprangers & Hoogstraten, 1989) and of the dimensions comprising the concept. This might be of special importance in educational settings when participants have to rate themselves on behaviors or skills they have not yet performed. In those cases, the students might not be able to give valid self-representations on the pretest. As a remedy for this problem, Howard, Dailey, and Gulanick (1979) suggested to provide students with information on the given subject before the pretest. With the help of this information, students should thus be able to give more adequate estimates of their performance on the investigated dimensions and no response-shift should occur. Findings haven been equivocal, as for example, Goedhart and Hoogstraten (1992) could not show that the provision of specific information helped to prevent a response-shift, although the used information might not have been adequate.

---

[9] For further definitional aspects of the response-shift phenomenon see, for example, Schwartz and Sprangers (2000).

Whereas studies have in general supported the validity of the retrospective pretest-/posttest-design and could also show that the retrospective scores were more in agreement with non-self-report measures (e.g., Hoogstraten, 1982, 1985; Howard, 1980; Levinson, Gordon, & Skeff, 1990; Sprangers & Hoogstraten, 1989), others did not find significant correlations between self-reports and objective criteria (Sprangers & Hoogstraten, 1987). There is furthermore some question of the response-shift's being confounded with recall bias and of the optimal time interval between pretest and retrospective pretest (Howard & Dailey, 1979; Schwartz & Sprangers, 2000; Schwartz, Sprangers, Carey, & Reed, 2004). As Schwartz and colleagues (2004) point out, there is "a substantial amount of noise in the thentest that can reduce one's statistical power to detect recalibration response shifts. Furthermore, thentest results are somewhat ambiguous in that the actual meaning of the scores has never been validated" (p. 65). The relative validity of pretest/posttest and retrospective pretest/posttest scores is thus not unequivocally established (Koele & Hoogstraten, 1988).
To assess the possible occurrence of a response-shift, a retrospective pretest will be included in the posttest questionnaire because the study could only employ self-report measures to assess student learning.

3.3.1.4.3 Conclusion

The general evaluation design considerations and the specific response-shift considerations provided the basis for the final training evaluation design study. During these considerations it already became evident, that the design would have to incorporate several limitations due to the restrictions of the applied and thus natural setting in which the study was to take place. These restrictions will be explicated below, when the final study design is described (3.4.2).

3.3.2 Training design tasks: Development of the interventions

In the course of the development of training programs, the step following the instructional design tasks is the design of the actual training program (cf. 2.2.1.5). The design of the present training program was based on the needs assessment, on the derived learning objectives, and on the previously existing training program. As the underlying general learning theoretical approach, an integrated approach of cognitivist and constructivist elements was chosen (cf. 2.1.3). As elaborated in the theory part (2.1.3), this approach combines the advantages of constructivist and cognitivist learning environments.
Based on the allowances the curriculum afforded, two training elements could be re-designed as the first two instructional interventions and thus provided the main elements of the instructional strategy: The preparatory introductory anesthesiology seminar (3.3.2.1) and the full-scale simulation sessions (3.3.2.2). Furthermore, as a third instructional intervention, the curricular sequence could be revised (3.3.2.3). The latter revision could not be planned at the beginning of the study because it was not clear if the respective instructors' clinical schedules

would allow any changes. This revision was thus conducted in phase two of the redesign process.

3.3.2.1 Design of the revised preparatory seminar

Based on the learning theoretical considerations, the analysis of the original preparatory seminar (see 3.2.1) showed that the seminar’s instructional strategy should be redesigned, as it was primarily teacher-centered and lecture-based, conveying a lot of factual information. The following sections first present the overall redesigned instructional strategy and resulting structure of the seminar (3.3.2.1.1). Different aspects that received special consideration in the seminar development are subsequently described in more detail (3.3.2.1.1.1-3.3.2.1.1.6) and rounded up with the conclusion (3.3.2.1.2).

3.3.2.1.1 An integrated approach to learning as the didactic basis for the seminar’s instructional strategy

An integrated approach to learning (Reinmann-Rothmeier & Mandl, 2001, cf. 2.1.3.1) was used to restructure the seminar. This led to a new structure of the seminar with four main components or phases (Table 7).

Table 7. *Seminar structure of the revised preparatory introductory anesthesiology seminar*

Seg-ment	Content	Method	Time
Introductory phase: Warming-up and activation of prior knowledge			
1	Start of the seminar, greeting of the participants, introduction of the instructor	Instructor presentation	3’
2	Clarification of learning objectives for the seminar and full-scale simulation sessions	Instructor presentation	5’
3	Learning contract and learning orientation, overview of seminar	Instructor presentation	8’
4	Introduction to the content: Problem-based anchor: “HANS needs anesthesia”	Video presentation, group discussion	10’
Main phase: Presentation of new content			
5	Drugs in anesthesiology I (hypnosis): Types of drugs, effects, side effects, contraindication, dosage	Group discussion, structuring and visualization by instructor, instructor presentation	20’
6	Drugs in anesthesiology II (muscle relaxation): Types of drugs, effects, side effects, contraindication, dosage	Group discussion, structuring and visualization by instructor, instructor presentation	20’
7	Drugs in anesthesiology III (analgesia): Types of drugs, effects, side effects, contraindication, dosage	Group discussion, structuring and visualization by instructor, instructor presentation	20’
Main phase: Application of new content			
8	Case 1: Healthy patient (Guided application)	Group discussion, structuring and visualization by instructor	10’
9	Standard anesthesia procedure	Group discussion, structuring and visualization by instructor	5’
10	Case 2: Varying patients (Application)	Small group work (6 groups, working on a different case each)	15’
11	Solutions for case 2	Student presentations of group work results, group discussion of the presented results	20’
Conclusion phase: Transfer support			
12	Transfer-anchor “HANS receives anesthesia”	Video	20’
13	Summary of the video and the seminar content	Group discussion led by instructor	
14	Clarification of further questions	Group discussion led by instructor	
15	Preparation for the full-scale simulation sessions – homework (further patient cases)	Instructor presentation	

In the *introductory phase*, the instructor first greeted the students and introduced himself to them (segment 1). Following this introduction, he presented the learning objectives of the seminar and the full-scale simulation sessions (as listed in 3.3.1.2), as well as the seminar's agenda to the students. The instructor emphasized that the seminar functioned as the preparation for the full-scale simulation sessions. The students should thus realize that the simulation sessions would be the application context of the material they were going to learn in the seminar. This was done in an effort to provide students with an understanding that they first needed to acquire a sufficient knowledge base before they could encounter "meaningful experiences" in which to apply their knowledge (Reinmann-Rothmeier & Mandl, 1999; Sonntag, 1997). The presentation of the learning objectives furthermore served to motivate the students as will be further explicated below (3.3.2.1.1.2)

In segment 3, the instructor then presented the learning contract to the students, clarifying what was expected of them in their role as learners in the seminar and the simulation sessions and what they could expect from the instructors in their alternating roles as presenters and facilitators (see appendix A). The learning contract further served to influence the students' learning orientation as will be elaborated below (3.3.2.1.1.3).

The introductory phase concluded with segment 4 in which the students received a problem-based anchor presented in video format (see below 3.3.2.1.1.6). Following the short video presenting a physician who modeled examining the patient and preparing him for the anesthesia induction, the instructor asked the students to identify the subsequent tasks that needed to be completed to anesthetize the patient. This was done to activate students' prior knowledge in form of a group discussion.

The first part of the main phase (presentation of new content) was further divided into three segments (5-7) based on the different groups of drugs needed to anesthetize patients. Each segment began with a group discussion which the instructor facilitated by asking the students which drugs are needed, in which order they are needed, in which dosage, and so on. The instructor visualized the students' answers using a flip chart or chalkboard. Building on the students' answers he then presented the information about the respective drugs.

Following this knowledge-building part of the main phase, was the application part of the main phase. In this part, the students were confronted with a patient case (segment 8) and were asked to apply what they had previously heard to solve the case and identify the necessary drugs and steps needed to anesthetize the presented patient. In this exercise, the instructor still guided the students where necessary. He facilitated the group discussion, structured and visualized the students' answers. The result of this exercise was the development of a standard anesthesia procedure (segment 9) that served as a guideline for the students to approach future cases in the seminar and then also in the simulator.

In the next step (segment 10), the students received cases to solve in small groups on their own. The seminar participants were divided into six small groups of about five students each. Each group then received a patient case on paper. The groups had to come up with an anesthetic plan for the patient. Following the group work, each of the groups presented their

case and their solution to the group. The solutions were discussed among all of the students and evaluated by the instructor.
The conclusion phase of the seminar served to summarize the contents and to support the transfer of the learned material. This phase started with the second part of the transfer-anchor video in which HANS is actually anesthetized by the physician. The physician thus modeled in the simulator all of the steps the students had previously worked out during the seminar.
The video contents were then summarized by the instructor. He again pointed out what would be expected of the students in the simulator, how their work today had prepared them for the simulation sessions, and which additional skills they would acquire or further practice in the emergency medicine part-task simulations, such as mask ventilation and intubation skills.
The instructor then asked if the students had any further questions and answered them if necessary.
At the end of the seminar, the instructor discussed with the students what they thought they still needed to do at home to be prepared for the simulation sessions. He then handed out further patient cases the students should solve in preparation for the simulation sessions.

On a content level, the preparatory seminar was restructured to achieve the explicated learning objectives of knowing the respective drugs available to induce and maintain anesthesia and the according algorithms for inducing it. The restructuring also implied that the content was focused on the most important drugs instead of presenting information on all of the available drugs. This means that a reduction of presented individual facts was accepted in exchange for the focus on teaching the students problem-solving skills.

The following sections describe in further detail, how the above presented seminar components were designed to foster the students' activity (3.3.2.1.1.1), their motivation (3.3.2.1.1.2), their learning-goal orientation (3.3.2.1.1.3), their readiness for the simulation sessions (3.3.2.1.1.4), the adequacy of their training expectations (3.3.2.1.1.5), and the transfer of the KSAs acquired in the seminar to the simulation sessions (3.3.2.1.1.6).

3.3.2.1.1.1 Fostering student activity

The seminar in its original version followed a traditional instructional approach, focusing on the active role of the teacher. As has been explicated in the section on constructivist learning environments (cf. 2.1.2), students' active involvement in seminars is, however, critical for fostering their learning. Campbell (1988), for example, already noted that the activation of the trainees is of pivotal importance for the learning process. The more active the learners' production, the greater the retention and transfer of the learned KSAs. An integrated approach to learning recognizes that learning also needs to include receptive phases (cf. 2.1.3.1) to provide students with a sufficient knowledge base before exposing them to complex problem contexts.

Since the original seminar was purely lecture-based, a new format was developed aiming at incorporating more problem-based features to foster students' generative activity and at the same time utilizing the advantages of a lecture-based format to provide them with the appropriate knowledge base (cf. Reinmann-Rothmeier & Mandl, 2001). This was accomplished by restructuring the seminar. The students' prior knowledge was activated through the use of a transfer-anchor (segment 4, cf. 3.3.2.1.1.6) and they were then encouraged to verbalize the prior knowledge they had gained in other subject areas and their pre-clinical studies. After these active and activating parts, they received the respective new input by the instructor (segments 5-7). Even in the presentation phase the instructor involved the students by asking them questions and encouraging them to comment on the content. The second part of the main phase then focused on the students' constructive activity in form of patient cases the students received on paper. These cases were intended to enable the learner to comprehend the new material, apply the newly acquired knowledge in a meaningful learning context, and to allow for the development of problem-solving skills (cf. Reinmann-Rothmeier & Mandl, 1999). Through the provision of different patient cases the students could solve together (segments 8 and 10), the development of a flexible knowledge base was furthermore supported. These different patient cases thus incorporated the constructivist learning principles of authenticity, situatedness, multiple contexts, and a social context (cf. 2.1.2.1). Since the first case was solved by all of the students together with the help of the instructor, the students also received the instructional support called for in the integrated approach to learning (cf. 2.1.3.1).

At the conclusion of the seminar, the students were further encouraged to continue working with the material by the provision of more patient cases they could independently solve. This allowed the students to take full control of their learning process (De Corte, 2001).

The first aspect changed in the revision process of the seminar was thus the fostering of the students' activity through the inclusion of several constructivist elements as well as features of an integrated approach to learning. The second aspect that was tried to be influenced was the students' motivation.

3.3.2.1.1.2 Fostering student motivation

In addition to employing constructivist methods, that is patient cases as authentic tasks and anchors to improve student motivation (Honebein et al., 1993), the seminar aimed at improving student motivation through varying the elements of instruction (group work, discussion, videos, etc.) (Keller & Kopp, 1987).

The next measure employed to improve motivation, was the presentation of the learning objectives in segment 2 of the seminar. They were used for informative as well as for motivational purposes (Hesketh, 1997). As has been explained in the presentation of the model (cf. 2.2.1.5.1.4), the measures taken to improve motivation largely depend on the operationalization of the construct. The chosen operationalization of motivation was the instrumentality-expectancy approach (Lawler & Suttle, 1973; Vroom, 1964). Based on this operationalization, the instructor's elaborations after the presentation of the learning objectives were designed to influence the training's instrumental value as perceived by the students. The instructor informed the students that the simulation sessions were a very important preparation for the students' exams, the OSCE at the end of the block. He pointed out that the students would have a better chance to succeed in the exam if they learned a lot during the simulation sessions and that the simulation sessions offered a unique learning opportunity no other method offered. Based on his clinical experience, he further explained to them that they would have the opportunity to acquire skills in the simulation sessions that would be very useful for them in their future clinical work.

The second aspect targeted in the revision process was thus the redesigned introductory phase of the seminar aiming at improving the students' outcome-expectancies regarding their expectancies that doing well in the simulation training would lead to better outcomes, such as exam success and improved clinical skills. The next aspect to be influenced was their situational goal orientation, as this has been identified as an important variable in the area of motivation and training effectiveness.

3.3.2.1.1.3 Fostering learning goal orientation

Among the challenges for training designers is the creation of a learning environment that trainees do not primarily perceive as an examination situation as this might foster their performance goal orientation (Schaper, Schmitz et al., 2003). This is of special concern for the development of simulation training, as simulation training has been found to be a stressful experience for medical students (Shimoda & Ikuta, 2005). Other research indicated that participants perceive simulation training as stressful as well as intimidating and as causing discomfort (Cleave-Hogg & Morgan, 2002; Kurrek & Fish, 1996; Savoldelli et al., 2005). They fear, for example, the instructors' and peers' judgments and anticipate anxiety. Furthermore, previous studies in the field of computer-based constructivist medical learning environments (Gräsel, 1997; Reinmann-Rothmeier & Mandl, 2001) provided first indications that the students used the learning cases (or scenarios) to test their knowledge (reflecting performance orientation) rather than to acquire further knowledge (reflecting learning orientation).

Experiencing high levels of stress and working through learning cases thus might lead students to become more performance oriented, trying to avoid mistakes and negative evaluations instead of trying to maximize their learning and acquire new skills.

Since respective research has shown that trainees with a performance goal orientation are more likely to perceive tasks as threatening and that a performance goal orientation can interfere with learning (see 2.2.1.1.2), specific measures were taken to foster a learning goal orientation in the seminar and for the simulation sessions (see Cannon-Bowers, Rhodenizer et al., 1998; VandeWalle et al., 2001).

Following Cannon-Bowers and colleagues' suggestion (1998) students' situational goal orientation was fostered by information presented by the instructor in lecture format during the revised seminar. The intervention was placed in the presentation of the learning contract (segment 3). The instructor was provided with a learning contract slide as a basis for his elaborations (see appendix A). He encouraged the students to suspend disbelief during the simulation sessions, to view them as a challenge rather than a threatening situation, to use these sessions to build their skills, and to remember that mistakes are useful for their learning process. The instructor emphasized viewing the simulation training as a non-threatening opportunity to develop skills in which the students should concentrate on their own learning process instead of on comparing themselves to their peers. Students were further informed that their performance in the simulator was confidential and would not count towards their course grade. They were encouraged to focus on the lesson content instead of on possible reasons for their success or failure in completing the tasks (Dweck, 1986; Kraiger et al., 1993). Based on findings that performance goal orientation is related to personal beliefs about ability and effort, namely, the belief that successful performance is primarily a result of the necessary innate ability, it was also pointed out to the participants that their ability and skills can be developed and that effort will lead to increased performance (VandeWalle et al., 2001).

The third aim of the seminar revision was thus the inducement of a situational learning goal orientation for the seminar and specifically for the simulation sessions. The next specific aspect targeted by the redesign was the improvement of the students' readiness for the simulation sessions.

3.3.2.1.1.4 Fostering readiness

Especially in training settings in which resources are scarce and training itself is resource intensive, such as the present setting, it is important to realize that practice is a complex concept that implies more than simple task repetition (Schmidt & Bjork, 1992) and that can consequently be optimized by changing more than merely the amount of repetition. One of the variables that has been found to affect practice and learning is the trainees' readiness.

Based on the literature described earlier (Cannon-Bowers, Rhodenizer et al., 1998; Hilkey et al., 1982), the intervention created to improve trainees' readiness was included in segments 2-4 and segments 12-13. The first measure was the clarification of the learning objectives in segment 2 to create an awareness among the students as to which effects the seminar and the full-scale simulation sessions should have on their KSAs and how these elements related to the OSCE. The second measure was included in the learning contract and overview of the seminar content (segment 3) which clarified what was expected of the trainees in the training contexts (Cannon-Bowers, Rhodenizer et al., 1998). The third measure was the use of the

videotapes that modeled a short simulation session. The videos thus served several functions: In the sense of an anchor, they provided the background of the problem the students subsequently had to work on; in the sense of fostering the students' readiness, they provided relevant information on adequate behaviors to be displayed in the simulator to improve their readiness. In another context, the use of pre-training videos has already been shown to improve readiness: the used videos led to improved role-expectations and to the elicitation of more adequate behavior in the first therapy session (Hilkey et al., 1982).
The need to support and evaluate trainees' readiness has been emphasized by McIvor (2004) who remarked that "intuitively, simulation education will benefit from participants beginning the simulation with a thorough understanding of the cognitive goals of the course" (p. 66). Data to support this intuition are, however, still lacking.

The fourth aspect specifically included in the redesign of the seminar thus was the improvement of the students' readiness for the full-scale simulation sessions by explicating which effects the simulation sessions were supposed to have and what was in general expected of them. As the next aspect, the students' more specific training expectations were targeted.

3.3.2.1.1.5 Fostering training expectations

From a more detailed viewpoint the students' specific didactics-oriented training expectations were sought to be influenced as realistic training expectations have been shown to lead to greater motivation, commitment, and better reactions (Cannon-Bowers, Rhodenizer et al., 1998; Hicks & Klimoski, 1987).
As a more specific concept of readiness, the students' expectations concerning their role in the simulation setting and the didactic approach in the sessions were influenced by the presentation of the learning objectives, learning contract, and the video (segments 2-4, and segment 12). It was pointed out to the students that they needed to participate actively in the seminar and the simulation sessions, that they would have the opportunity to apply their theoretical knowledge in the simulation sessions, that they would work independently and in teams, and that they could improve their skills in anesthesiology.

As the fifth aspect it was thus decided to try to improve the students' specific training expectations in regard to the didactic approach and the resulting requirements for the students. The last aspect that was specifically targeted was transfer from the seminar to the simulation sessions.

3.3.2.1.1.6 Fostering transfer

In the needs assessment process it had become evident that the simulation session instructors felt that the students were not able to transfer the knowledge they had acquired in the seminar to the simulation setting. In the redesign process of the seminar it was thus sought to include elements that would foster the transfer of the knowledge acquired in the seminar to the simulation setting. Based on the research findings presented in section 2.2.1.5.1.5, the following elements were used to foster the students' transfer.

First of all, the focus on learner activity throughout the seminar was emphasized in order to support transfer (Campbell, 1988). Secondly, the use of multiple patient cases was designed to improve transfer in the sense of stimulus variability (Ellis, 1965) and also in the sense of providing authentic problem situations (Reinmann-Rothmeier & Mandl, 2001). Thirdly, the deduction of a standard anesthesia procedure (segment 9) as an underlying principle was employed to support transfer (Goldstein, 1993; McGehee & Thayer, 1961).

As the fourth and major element to support transfer, the video transfer anchor was developed (segments 4 and 12). It was designed along the guidelines of the anchored instruction approach (Bransford et al., 1990; Cognition and Technology Group at Vanderbilt, 1990; cf. section 2.1.2.2). Its use was based on findings that students who learned with anchors were more likely to transfer the acquired knowledge (Bransford et al., 1990; Cognition and Technology Group at Vanderbilt, 1990). The video was recorded in the HANS full-scale simulator with one of the instructors acting as the physician. It showed the physician examining the patient before starting the anesthetic (segment 4) and later conducting a standard anesthesia (segment 12). The video was thus designed in a narrative format that was supposed to increase the students' intrinsic motivation by presenting a realistic patient problem.

The first video-segment (segment 4) explicated the general goal of the seminar, that is, the acquisition of the knowledge to anesthetize patients. It could be reached by more specific subgoals, such as identifying the necessary drugs, the necessary steps, etc. Since the anchor was taking place in the simulator and the patient was being examined by the physician, it also allowed the identification of the main characteristics of the problem situation. This set-up furthermore made the anchor authentic and complex. The only design guidelines that were not followed during the design of the anchor were that the anchor was not followed by a parallel anchor and that it did not contain all of the information the students needed to solve the task. The first aspect was not deemed necessary as the students would subsequently act in the application context displayed in the anchor during the simulation sessions. The second aspect was not deemed necessary since the anchor was used as an introduction to the teacher-guided seminar.

The anchor further served as an advanced organizer for the seminar and the subsequent full-scale simulation sessions. Ausubel (1960) introduced the "advance organizer", a conceptual framework providing the trainee with an overview of the training that is to come.

Although research findings are controversial (see Patrick, 2003), it seems safe to conclude that it is beneficial for learners to provide an overview at the beginning of practice that incorporates different types and levels of information and serves as a context and synthesis. Other research in the field of anesthesiology, although in a slightly different context, has provided evidence that students perceived a video introduction to the simulator as helpful (Morgan et al., 2001).

The concept of transfer support was thus incorporated in the design of the seminar in different forms and with different methods. Care was taken to thereby prepare the students to use the simulation sessions as learning opportunities to further build on the knowledge they acquired in the seminar and not as application settings in which they were supposed to primarily display what they had already learned. The latter approach would have likely induced a performance orientation that was not considered desirable.

3.3.2.1.2 Conclusion

The seminar was revised in phase one of the study to improve a set of individual characteristics that have been identified as important input factors for the effectiveness of training programs (cf. Figure 1). It was also revised to incorporate principles of the learning design that have been specified in the training model.

The instructionally revised seminar was thus supposed to focus more on student learning than on presenting the greatest amount of information without enough consideration of learning processes. The chosen instructional strategy sought to combine constructivist and cognitivist learning principles and methods to create an integrated approach to learning. The focus was put on activating the learners in the seminar, improving their motivation, learning goal orientation, readiness, and specific expectations. They were furthermore prepared to transfer the theoretical knowledge acquired in the seminar to the subsequent simulation sessions to be able to build their practical anesthesia skills on the basis of this knowledge.

3.3.2.2 Design of the revised full-scale simulation sessions

The second intervention targeted the full-scale simulation sessions. The redesign was based on the analysis of the original simulation sessions (cf. 3.2.2.1), the needs assessment phase with the respective learning objectives, and the theoretical considerations.

The following sections first describe how the scenarios were developed and which content was covered in them (3.3.2.2.1). Then the underlying integrated didactic approach is characterized (3.3.2.2.2), followed by the conclusion of this redesign process (3.3.2.2.3).

3.3.2.2.1 Scenario development and scenario content

The simulation scenarios were developed through a standardized, multistage process during a workshop with the two senior anesthesia simulation SMEs. The workshop was facilitated by the author of this dissertation.

In the first stage, a list of clinical cases was developed in form of a brainstorming. The basis for this brainstorming were the identified learning objectives (see 3.3.1.2) and the capabilities of the simulator.

In the second stage, the following set of requirements the cases had to fulfill was developed (Table 8).

Table 8. *Requirements for the cases used as the basis for the scenarios*

Case requirements
• The used cases support the learning goals. • The cases take place in a simulated OR. • The cases can be realistically simulated with the patient simulator. • The cases allow for preoperative patient evaluation, induction, and maintenance of anesthesia. • The cases have different levels of difficulty. • The cases allow each student to manage one aspect of the task. • The cases can easily be standardized and thus reproduced with the simulator.

In the third stage, the list of cases was reviewed by the SMEs. They prioritized the cases based on their utility for the achievement of the learning goals and the fulfillment of the requirements. The top four cases were then selected.

In stage four, they were scripted into scenarios to fulfill all of the requirements. The scenario scripts served multiple purposes. They were used to standardize the simulation sessions by serving as a scenario outline for the instructor and the simulator operator. They contained a brief patient history (reason for admission, patient condition, examination information), instructions for the simulator's parameter settings, and the scenario's learning objectives. Based on this information, the simulator operator also formulated his, that is, the mannequin's, answers to the students' questions during the examination (e.g., the nature and severity of experienced pain, shortness of breath, habits, or previous therapeutic treatments). The operator was further instructed to assure that the simulator responded accurately to the students' interventions but to avoid the simulator's "death".

In the fifth stage, the scenario scripts were reviewed by other simulation faculty and staff to assure that they were suitable for attaining the listed learning goals as well as sufficiently realistic.

The following section describes the scenarios, as they were included in the curriculum.

3.3.2.2.2 An integrated approach to learning as the didactic basis of the full-scale simulation sessions' instructional strategy

The following sections first describe the simulation sessions' structure and content based on the previously presented training development steps (3.3.2.2.2.1), then they describe how constructivist design principles can be implemented with the help of simulators (3.3.2.2.2.2), and how cognitive apprenticeship elements can be used to provide instructional support to students and to thus achieve an integrated approach to learning (3.3.2.2.2.3).

3.3.2.2.2.1 Full-scale simulation sessions' structure and content

In accordance with the SMEs, the simulation sessions were redesigned based on the scenario scripts and the learning theory considerations. As in the preparatory seminar, an integrated approach to learning (Reinmann-Rothmeier & Mandl, 2001; cf. 2.1.3.1) was employed and Tannenbaum and Yukl's (1992) recommendations for the design of simulations were applied (2.2.1.5.2.2).
This approach led to the following structure and content of the two simulation sessions (Table 9, Table 10).

Table 9. *Structure of the revised full-scale simulation session a)*

Segment	Content	Method	Time
Introduction to the simulator			
1	HANS and the OR: Characteristics, equipment	Instructor presentation	10'
Scenario I: Standard anesthesia induction			
2	Scenario briefing	Instructor presentation	25'
3	Scenario	Group work supported by instructor	
4	Reflection	Instructor-facilitated group discussion	
Scenario II: Rapid sequence induction			
5	Scenario briefing	Instructor presentation	25'
6	Scenario	Group work supported by instructor	
7	Reflection	Instructor-facilitated group discussion	

The first full-scale simulation session (session a), Table 9) started with the students' introduction to the simulator (segment 1). Upon entering the simulator, the instructor briefly described the set-up of the simulation center with the OR and the control room. He then proceeded to explain the equipment in the OR (anesthesia machine, monitoring, drugs, etc.), the simulator with its functions and limitations, and how it can be manipulated. While demonstrating the mannequin's functions, he also modeled the basic tasks of auscultating and intubating the mannequin. After this presentation, the students had the opportunity to inspect the simulator and get acquainted with the equipment. Their questions concerning the simulator were answered by the instructor.

After the introductory phase, the students were divided into two groups of about four people each. One group proceeded to the control room where they watched the scenario through the one-way mirror. The other group worked through the first scenario.

The respective group in the simulator was asked to coordinate the scenario as a team. They consequently had to distribute the tasks among themselves. One student was, for example, responsible for taking the patient's history, one for managing the airway, one for monitoring the basic patient parameters, and the other for administering the necessary drugs. Each student thus had the opportunity to manage one aspect of the scenario.

At the beginning of the scenario the students were briefed about the general storyline of the scenario (segment 2). The instructor provided basic information on the patient and his reason for being in the hospital, such as "the patient was admitted to the hospital for elective surgery…".

After this briefing, the scenario began and the instructor retreated to the side of the room (segment 3). The students were then expected to start the scenario by examining the patient (i.e., auscultation, palpation, etc.), diagnose him, and initiate the appropriate anesthetic treatment for the described surgical procedure. After taking the patient's history, they had to induce and briefly maintain anesthesia. They thus acted as anesthesiologists treating the patient.

In scenario I, the students were called in to anesthetize a patient for an elective procedure. When they examined and questioned the patient, they found out that he was young, healthy, and reported no risk factors; he had been admitted to the hospital the day before to have metal extracted that had been used to stabilize a fracture. He was appropriately premedicated. The students could thus proceed to induce a standard anesthesia.

In scenario II, the students were asked to treat a patient who had been admitted to the hospital for an emergency appendectomy. In their patient examination, the students were supposed to find out that the patient had eaten an hour ago and felt nauseous. Other than that, he reported no risk factors. The students were thus expected to perform a rapid sequence induction due to the patient's aspiration risk.

The students worked through the scenario, at demand guided by the instructor. In the role of the facilitator, the instructor supported the students only where necessary, for example when having attempted a subtask unsuccessfully several times. That is, the instructor did not influence student management of the scenario nor did he demonstrate correct maneuvers.

When they had accomplished the task, the instructor marked the end of the scenario and started the reflection phase (segment 4) in which he first debriefed the students to allow them to step out of their roles and in which he then facilitated the student's reflection on the scenario.
The groups then traded places and the second scenario was conducted in the same fashion. The didactic basis of the different segments is described in more detail in section 3.3.2.2.2.3.

Table 10 displays the structure of the second simulation session the students attended (session b). The structure of the session was the same as in session a). The only element not repeated in session b) was the introduction to the simulator. The scenarios used for this session were more complex than in session a). The students now had to induce and maintain anesthesia for patients who had cardiac or pulmonary risks .
In scenario I, the students were introduced to a 66-year old patient who had been admitted to the hospital due to a fracture. Upon examination the patient reported heart problems and the use of respective heart medication, heavy smoking and several previous surgeries. The students were expected to diagnose the patient as a cardiac risk patient and thus use the appropriate drugs for the anesthesia induction. During the induction, the patient showed signs of hypotension. The students were expected to treat the hypotension with the appropriate drugs, e.g., Adrenalin.
In scenario II, the students met a patient who was admitted to the hospital for an acute cholecystectomy. During the examination, the patient felt nauseous and reported hay fever. The students were expected to identify the patient as a pulmonary risk patient and perform a rapid sequence induction due to the aspiration danger. After the induction, the patient had an allergic reaction that the students were supposed to treat.

Table 10. *Structure of the revised full-scale simulation session b)*

Seg-ment	Content	Method	Time
Scenario I: Anesthesia induction of a cardiac risk patient			
1	Scenario briefing	Instructor presentation	30’
2	Scenario	Group work supported by instructor	
3	Reflection	Instructor-facilitated group discussion	
Scenario II: Rapid sequence induction of a pulmonary risk patient			
4	Scenario briefing	Instructor presentation	30’
5	Scenario	Group work supported by instructor	
6	Reflection	Instructor-facilitated group discussion	

3.3.2.2.2.2 Realization of constructivist design principles in full-scale simulations

As has been illustrated in the theory part of this dissertation, full-scale simulations allow for an excellent realization of constructivist approaches to learning, such as the situated cognition approach. It has been noted (Lupien, 1998) that there are two main reasons for preferring simulation environments over real-world training for situated learning: *First*, the learning situations can be designed to include high degrees of interaction, provide for specific feedback, match the learner's level of expertise, require the learner to actively participate, and still provide a meaningful individual learning experience. *Second*, the sole objective of the simulation is the learners' education. In uncontrolled and uncontrollable real-world environments where students are trained while working with real patients, the patient's safety has to be the primary goal and extraneous circumstances may occur that require immediate attention at the cost of training.

Full-scale simulators are thus training tools that optimally allow for the realization of constructivist design principles (see Schmitz, 2002, for a more detailed discussion).

With the help of the simulator, scenarios can be created that provide *authentic* application situations for the learner. The level of complexity can be varied by the designer and corresponds to the complexity of the real application situation. It is to be noted, that authenticity is a relative concept. Simulators involving the replication of human elements will achieve only moderate degrees of realism. Nevertheless, they can still supply an authentic learning context if they ensure that the learner is required to perform the essential components of the real task (cf. Honebein et al., 1993). Simulators can furthermore achieve the *situatedness* of instruction because the scripted scenarios allow for embedding the problem in a broad context.

The scripted scenarios allow for the provision of *multiple contexts* that are at the same time meaningful. The scenarios enable the learner to use his knowledge in different problem areas, decontextualizing it, making it more flexible, and thus more applicable in real world settings. Learners can also be encouraged to view problems from different *perspectives* when they can play different roles in the scenarios and later reflect on these roles. This also supports the formation of flexible knowledge.

A *social context* for learning can also be easily created in the simulator. If a group of trainees participates in the simulation, the trainees are required to cooperate during the simulation to deal with the presented situation. In the reflection phase after the scenario, the discussion of the previous problem-solving attempts further supports generative learning through reciprocal explanation and thus deeper understanding.

On a more critical note it should of course not be neglected to see that the realization of these principles can also be problematic even with the help of full-scale simulators. *Authenticity* can be difficult to realize due to fidelity issues of the used simulator. This is especially problematic in the area of medicine, where a human being has to be simulated and the simulator can thus never achieve the same level of physical fidelity a simulator that simulates technical systems can. Authenticity is also limited by the fact that the scenarios are often much shorter than the real event or procedure. Although *multiple contexts* can be constructed, their generation is oftentimes difficult or resource intensive. In the area of medical simulation, multiple contexts can, however, rather conveniently be developed if different patient problems are considered as different contexts. Depending on the type of simulation the *social context* can also be difficult to realize if, for example, parts of the medical team are role-players instead of participants from different disciplines (see Schmitz, 2002).

Beyond these realization challenges, a more severe problem that directly stems from the above mentioned affordances, is, that simulation is often used as the sole constructivist training strategy and thus often also suffers from the disadvantages of purely constructivist learning environments: The lack of support can leave learners overwhelmed, especially those in the initial skill acquisition phase (Jonassen et al., 1993), and learning might require an extensive amount of time (Gräsel, Prenzel, & Mandl, 1993; Perkins & Cunningham, 1992).
Constructivist simulation settings should thus be enhanced by providing students with instructional support, especially if they are in the early stages of skill acquisition as in the current study context. Again, simulators also provide many opportunities for *instructional support*. In contrast to the real working environment, the simulation can, for example, be stopped to allow the instructor to actively provide input, or the course of the simulation can be changed, that is, its level of difficulty can be varied, to meet the learners' need for support. To provide an integrated approach to learning in the simulator, that combines the benefits of constructivist and cognitivist didactic elements, the incorporation of cognitive apprenticeship elements was chosen (cf. 2.1.3.2). They are detailed in the following section.

3.3.2.2.2.3 Cognitive apprenticeship elements providing instructional support for the learners

Cognitive apprenticeship elements (Brown et al., 1989) used in the full-scale anesthesia simulation sessions of this study were *modeling, coaching*, *scaffolding*, *fading*, *articulation*, *reflection*, as will be described below.

Modeling
Modeling as an instructional method was used at two points in time. It was first used in the video the students watched during the seminar. In the video, the instructor functioned as a positive *model*, performing the complete task of anesthetizing the patient.
When the students entered the simulator in the first full-scale simulation session, they were introduced to the simulator and its functions. In this segment, the instructor again served as a positive model. However, the instructor modeled only relevant subtasks this time, such as auscultating and intubating the simulator, not the complete task.

Coaching, scaffolding, fading

In the simulator, the instructor first introduced the students to the scenario, that is, the complex problem they would have to solve. He provided basic information about the patient and his history. Every further pieces of information were to be extracted from the patient by the students.

After this general introduction, the instructor assumed a reactive role, acting as a *coach* or facilitator (Leinhardt, 1993; cf. 2.2.1.5.1.1). He stepped back from the "patient" and followed the students' problem-solving activities. When the students were stuck in their process or pursued a wrong path of treating the patient, the instructor provided support in the form of hints, feedback, or recommendations.

This support was only given where needed and was thus adjusted to the performance levels of the different groups (*scaffolding*). It was furthermore *faded* in the course of the simulation session.

Articulation

To promote the development of the learners' metacognitive strategies and multiple perspectives as well as the group's collaboration, the instructor reminded students in the introductory phase to verbalize their thoughts related to the problem, diagnosis, and management. The students were encouraged to exchange their ideas, compare their solution strategies, and discuss arguments for the chosen solution strategies.

Reflection

Following the end of the scenario, a reflection and debriefing segment was added. The debriefing serves to assure that the students are able to leave their role and cope with any emotions that might have occurred during the simulation (Stafford, 2005). In the debriefing part, the instructor thus marked the end of the scenario, instructed the students to "step out of role", and asked them to voice their general impression of the scenario.

In the reflection part of the segment, the learning objectives and the learning process were addressed. Students were first asked to voice their experiences and recount the events as they had perceived them. The instructor then facilitated a group discussion concerning both, the strong points of the students' performance and the points that should be improved (Ellis & Davidi, 2005). Individual and group performance were thereby reflected. He emphasized that the students should try to provide feedback to each other, encouraging them to identify their own strengths, deficiencies, and areas for improvement. The feedback was aimed to be supportive, avoiding blame or humiliation, and focusing on the tasks and behaviors rather than on the individuals (Kluger & DeNisi, 1996). Again assuming the role of the coach or facilitator, the instructor only augmented or corrected the feedback where necessary. He thus ensured that any incorrect intervention was explored and discussed.

The students were asked to summarize the main learning points to be remembered and reviewed for the exam. The importance of this last step has again been highlighted by a recent article (Owen & Follows, 2006).

Exploration

The element of *exploration* was not used within the simulation session. *Exploration* implies seeking out further problem settings. This was not possible within the tight time frame of the simulation sessions. Students were instead encouraged at the end of the reflection phase to explore further paper patient cases in their preparation for the OSCE.

3.3.2.2.3 Conclusion

In conclusion, it can be stated that the described design process led to a standardized structure for the teaching sessions that was based on the incorporation of constructivist and supportive instructional elements as detailed in the presented model (Schaper, Schmitz et al., 2003). As the previous sections showed, simulators have unique and advantageous training-relevant characteristics, which allow training designers to create didactically sound learning environments.

The didactic revision of the seminar and the didactic revision of the simulation sessions were the main interventions designed for training group 1 in phase 1 of the redesign process. The only other intervention training group 1 received, was the change in the practice schedule that will be described below.

3.3.2.3 Design of the revised curricular sequence

The next set of interventions concerned the sequence of the curricular elements. The review of the original training in the needs and organizational assessment phase showed that the original curriculum contained a practice session on Tuesday with full-scale simulation session a) following directly after the three-hour preparatory seminar. The next full-scale simulation session (session b) then took place on Friday of the same week. In the following week, the students participated in the part-task part-scale emergency medicine training sessions (see Figure 3).

The following sections (3.3.2.3.1 and 3.3.2.3.2) describe the redesign of the curricular sequence that was undertaken in two phases.

3.3.2.3.1 Redesign of the curricular sequence: Phase 1: Spacing changes

The needs assessment had revealed that the Tuesday schedule was very dense for the students with the first simulation session (session a) following directly after the introductory seminar. Furthermore, the interviews with the SMEs revealed that the students did not transfer their knowledge from the preparatory seminar to the full-scale simulation session a) in a satisfactory fashion.

Together with the didactic redesign of the preparatory seminar, it was thus decided to change the position of the first simulation session to Wednesday afternoon (see Figure 3). This allowed the students to review the preparatory seminar contents as necessary and to deepen their understanding by working on further patient cases in their self-study period on Tuesday afternoon.

Following the suggestions of the cognitive apprenticeship approach (De Corte, 2001), the students first participated in both of the full-scale simulation sessions after the seminar, thus performing the global anesthesia skills before the local skills in the part-task emergency medicine training the following week (see Figure 3).

3.3.2.3.2 Redesign of the curricular sequence: Phase 2: Order and spacing changes

After the above described didactic and sequence changes had been implemented in the curriculum for training group 1 (phase 1), a new meeting was arranged with the SMEs. In this meeting, the affordances of the curriculum for additional changes were again investigated. It became evident that the curricular order could be further changed. For this second phase of the interventions, it was thus decided to now change the curricular order and thereby again the spacing. This change was based on considerations concerning the appropriate order of part-task and complete-task training. Interwoven with the question of part- versus complete task training was the question of the degree of simulator fidelity: The part-task training was conducted using SimMan™ part-scale simulators (see 3.2.2.2), whereas the complete task was simulated with the Human Patient™ full-scale simulator. Furthermore interwoven was the question of practice distribution or spacing.

Concerning the practice of part versus complete tasks, constructivist approaches, such as the cognitive apprenticeship approach, recommend practicing the global, that is, the complete or whole task first, before practicing the local (i.e., part) tasks to achieve authentic settings (Honebein et al., 1993). In the case of this curriculum, the complete task is the induction and maintenance of anesthesia, as practiced in the full-scale simulations. The complex task of inducing general anesthesia can, however, also be separated into a set of subroutines, such as ventilation, intubation, etc. Some of these sub-skills were practiced in the part-task emergency medicine training. Following the constructivist approach, training groups 0 and 1 practiced the global tasks first and then practiced the local tasks.

For training group 2 another order was sought based on several research findings concerning the order of part- and whole-task training as well as full- and part-scale simulation training. In a current paper, for example, Murray (2005) specifically recommended using a part-task approach to training complex medical skills and patient management problems. In addition, research showed that low fidelity simulators are particularly beneficial in early phases of the learning process (e.g., Dennis & Harris, 1998) and that trainees are able to transfer the skills acquired in low-fidelity simulators to full-scale simulators as well as to the job (Gopher et al., 1994; Jentsch & Bowers, 1998; Koonce & Bramble, 1998; Prince & Jentsch, 2001). Part-scale simulation of simplified tasks has proven especially useful for helping trainees practice and receive feedback on basic skill elements (Thornton & Cleveland, 1990). By training different parts of the complete process or task beforehand, the development of these basic skills is fostered. If students are forced to learn the complete task in a complex and authentic environment right from the start, the development of these basic skills is oftentimes overlooked. Once the trainees reached a certain level of competency, they are able to combine the basics skills in a complex environment. Thornton and Cleveland (1990) further recommended the sequencing of the simulations from simple to complex. This sequence enables trainees to focus on acquiring basic skills first and to then advance to practicing multiple skills in the complex full-scale simulation environment. Low-fidelity training has also been recommended as a supplement for high-fidelity training in the CRM training area (Prince & Jentsch, 2001). It should be used to offer participants the opportunity for practice and should thus bridge the gap between theoretical inputs and the high-fidelity simulations. Kozlowski and Salas (1997) further pointed out that transfer is enhanced by training individual-level skills to proficiency before training at the team level or training non-technical skills is offered. The part-task simulation sessions were more focused on the individual's skills than the full-scale and complete task simulation sessions in which the students had to perform as a team.

Prince and Jentsch (2001) caution training designers to carefully embed low-fidelity simulations within the greater training context, for example, by deriving its objectives from the training needs analysis. They conclude by stating that "low-fidelity systems offer valid training solutions that increase the usefulness of high-level, complex training environments" (Prince & Jentsch, 2001, p. 162).

The above mentioned research findings thus seem to support the use of part-task and consequently in this setting also part-scale training first and then using complete-task and full-scale simulations after the students acquired the basic skills.

Other research (e.g., Honebein et al., 1993) has however, pointed out, that tasks often become more intuitive in their regular context. The context contained in the full-scale simulator thus provides more cues for the students to initiate the required behaviors than the part-task simulators. Complete task contexts furthermore provide a form of an advance organizer (Ausubel, 1960) in which the students can experience the complete procedure they will be required to perform at the end of their training. Research furthermore showed that highly organized complex tasks render whole methods more efficient than part methods, unless their sub-tasks are psychologically meaningful (Fredericksen & White, 1989; Goldstein & Ford, 2002).

Together with the constructivist approaches, these findings would thus support the usage of full-scale complete task simulations first, followed by part-task part-scale simulations.

Based on these considerations and findings, the curricular sequence was again changed after phase two of the redesign process (see Figure 3). Training group 2 thus received the first full-scale simulation session (session a) on the day following the introductory seminar. This session served as an advance organizer and complete context to orient the students towards the task they would need to perform at the end of the curriculum. In the following four part-task emergency medicine training sessions the students then received the opportunity to practice the relevant sub-tasks. With the provision of the first full-scale simulation session before the part-task training it was assumed that it would be easier for the students to integrate the practiced subtasks in the second full-scale simulation session (Goldstein & Ford, 2002).

3.3.2.3.3 Conclusion

The redesign of the curricular sequence undertaken in two phases thus led to two different curricular sequences for training groups 1 and 2. The respective sequences are displayed in Figure 3. The different sequences reflect different research findings and their respective recommendations for ordering and spacing curricular elements. In the course of the training evaluation study, it was thus hoped to determine which of these sequences is more beneficial for the current training context.

3.3.3 Instructor training

Only slowly is the need for instructor training in the area of simulation being recognized (e.g., Seropian, Brown, Gavilanes, & Driggers, 2004). This is often due to the use of simulation as the sole training method; a method that is viewed as so powerful that further didactic considerations are not necessary. Consequently, it is often deemed sufficient for the instructors to possess the necessary medical skills to run the scenarios and the necessary technical skills to operate the simulator. As the organization analysis showed, the need for didactic training had also not been fully recognized among the HANS simulation faculty. Within the redesign process, it became evident, however, that the simulation faculty would also need to be trained in the revised didactic set-up.

Together with the two SMEs who were involved in the redesign process, it was decided to conduct a one-hour training session for the simulation center faculty who were involved in teaching the preparatory seminar as well as the full-scale simulation sessions. The training session served to familiarize the instructors with the new learning goals and to improve their teaching skills. Most of the instructors had not previously attended any training in didactics.

The lecture-based training session lasted approximately one hour and was conducted after semester 0 in which training group 0 received the original curriculum. In the training session, the faculty members were first given information regarding the nature of the study and the newly developed learning goals for the students. They then received the new scenario scripts that were reviewed in the group. Discussion was encouraged to promote understanding and commitment to the revised teaching content.

Furthermore, the instructional strategy of the revised preparatory seminar was presented to the instructors. They also received a detailed instructor's guide explaining the different segments (see appendix B).

The second part of the training session consisted of the introduction and discussion of the constructivist didactic guidelines and how they can be incorporated in the seminar and simulation sessions.

The instructor training concluded with a detailed discussion of the instructors' role during the simulation sessions, as described above.

After the instructors had implemented the new teaching approach with training group 1, a one-hour refresher session was conducted for them. The session started by reviewing the semester and discussing any difficulties that had arisen during the teaching sessions. After this discussion, the didactic principles were again covered.

3.3.4 Conclusion

The development of the revised anesthesiology curriculum followed the guidelines explicated in the training model. It comprised instructional systems design tasks and training design tasks as well as the training of the instructors who would teach the training program. The revision that was conducted in two phases resulted in two kinds of a revised curriculum (see Figure 3). As the model also shows, the evaluation of the training program is an inherent part of a systems approach to training development. The two different kinds of curricula were therefore to be evaluated in a study. The method of the conducted study will be described in the following chapter.

3.4 Method of the evaluation study

After the study setting (3.1), the original curriculum (3.2), and the development of the revised curriculum (3.3) have been described above, the following sections explicate the method of the evaluation study that was undertaken to investigate the effectiveness of the revised curricula. The description starts out by presenting the hypotheses (3.4.1). Section 3.4.2 then delineates the study's design. The employed evaluation instruments are presented in section 3.4.3, followed by the description of the study's participants (3.4.4) and the resulting procedure (3.4.5). The chapter concludes with elaborations on the analytic strategy (3.4.6).

3.4.1 Hypotheses

Based on the theoretical background of the training model and the resulting redesign of the simulation curriculum, the following hypotheses were formulated. It is to be noted that the presented training model would have allowed for a plethora of hypotheses which would, however, have been far beyond the scope of this study. Consequently, the training model was primarily used as a guideline for the design of the training interventions and as a basis to deduct the following set of hypotheses that could be investigated within the limitations of this evaluation study.

The hypotheses are structured along the levels of Kirkpatrick's model (1994) and the influence variables specified in the training model (Schaper, Schmitz et al., 2003). Section 3.4.1.1 explicates the hypotheses formulated for the reaction level, followed by 3.4.1.2 describing the hypotheses for the learning level and section 3.4.1.3 describing the hypotheses for the transfer level. Section 3.4.1.4 deals with the hypotheses concerning the students' individual characteristics and section 3.4.1.5 with exploratory hypotheses about the relations within the model.

3.4.1.1 Reaction level hypotheses

Concerning the first evaluation level, training reactions, the following hypotheses were formulated (Table 11).

Table 11. *Hypotheses concerning the students' reactions to the training*

Hypothesis number	Hypotheses: Training reactions
1.	Training groups 1 and 2 rate the full-scale simulation training as more valuable than training group 0.
2.	The three training groups do not differ in their ratings concerning the effects of the simulator's realism.
3.	The three training groups do not differ in their ratings concerning the simulator's suitability as a learning tool.
4.	Training groups 1 and 2 can adapt more easily to the simulations than training group 0.
5.	Training group 2 considers the level of difficulty most adequate, followed by training group 1 and 0, respectively.
6.	Training group 2 considers the used scenarios most motivating, followed by training group 1 and 0, respectively.
7.	The three groups do not differ in their ratings concerning the scenarios' relevancy.
8.	The three groups differ significantly in the degree that the full-scale simulation training fulfilled their expectations, with training group 2 being most satisfied, followed by training groups 1 and 0, respectively.
9.	The three groups differ significantly in their reactions to the instructors, with training group 2 rating them highest, followed by training groups 1 and 0, respectively.

A look at Table 11 shows that based on the above described redesign process, different kinds of hypotheses resulted concerning the students' reactions to the training program.

In hypotheses 2, 3, and 7 no differences between the training groups were postulated since the respective aspects were not changed in the course of the redesign process: the redesign process did not affect the simulator's realism, its suitability as a learning tool, or the scenarios' clinical relevancy.

Hypotheses 1 and 4 postulated differences between training group 0 against training groups 1 and 2, because only the first redesign phase of the training program was supposed to affect the respective aspects.
In *Hypothesis 1,* it was assumed that the redesign of the preparatory seminar would lead to increased perceptions of the training's value because the instructor explicated why the training was important in the contract part of the seminar (cf. 3.3.2.1).
Also based on the redesign of the preparatory seminar is *Hypothesis 4*. Due to the more detailed information on the simulation sessions that the participants received during the preparatory seminar, for example by the used transfer anchors, it postulates that groups 1 and 2 can adapt more easily to the simulation situation (Cannon-Bowers, Rhodenizer et al., 1998; Cognition and Technology Group at Vanderbilt, 1992; cf. 3.3.2.1).

Hypotheses 5, 6, 8, and 9 postulate differences between each of the groups because the respective aspects were affected by both of the redesign phases.

In *Hypothesis 5,* it was assumed that training group 2 considered the scenarios' difficulty level as most appropriate because they had participated in the revised full-scale simulation sessions that provided instructional support and had also participated in the interspersed part-task emergency training practicing their basic subskills, such as mask ventilation or intubation (Collins et al., 1989; Murray, 2005). Training group 1 was assumed to give the second best ratings based on their participation in the revised simulation sessions but still before they had the opportunity to practice the basic skills. Training group 0 was supposed to find the difficulty level least adequate because they participated in the original simulation sessions that did not provide additional instructional support for the learners.

Hypothesis 6 concerns the degree of motivation the students experienced through the scenarios. It was assumed that training group 2 was most motivated because of the authentic task augmented by the instructional support they received to prevent their being overwhelmed and because they could integrate the previously practiced skills in the full-scale simulation scenarios (Collins et al., 1989; Reinmann-Rothmeier & Mandl, 2001). Training group 1 was supposed to find the scenarios more motivating than training group 0 based on the additional support provided by the instructor (Collins et al., 1989).

Hypothesis 8 concerned the groups' overall expectation fulfillment concerning the curriculum. Based on the two sets of interventions, it was assumed that training group 2 showed the highest satisfaction followed by training groups 1 and 0, respectively. For training groups 1 and 2 the expectations were influenced by the information received during the revised seminar and the perception of the resulting didactic strategies was influenced by the instructor training. Since the instructor training was repeated after training group 1, it is thus assumed that they were best able to realize the didactic strategy for training group 2, thus leading to the highest expectation fulfillment for the latter group.

Hypothesis 9 concerned the students' reactions to the instructors' teaching skills. Based on the instructor training and the according refresher (cf. 3.3.3), it was assumed that the students in training group 2 would rate the teaching skills most positively because at that point the instructors had received the training and the refresher, followed by training group 1, at which point the instructors had only received the first training sessions. Training group 0 was assumed to provide the lowest ratings of the teaching skills because at that point most of the instructors had never received any didactic training.

3.4.1.2 Learning level hypotheses

For the second evaluation level, the following hypotheses were formulated (Table 12). Again, they were derived from the training literature reported in the description of the model and the resulting training design interventions.

Table 12. *Hypotheses concerning the students' learning*

Hypo-thesis number	**Hypotheses: Learning**
10.	Participants in all training groups will rate their performance as worse, i.e., will assign higher ratings, in the *retrospective pretest 2* than in *pretest 2*, indicating a response-shift.
11.	The instructionally revised training leads to greater improvements in participants' self-perceived competencies than the original training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively.
12.	The instructionally revised seminar leads to greater improvements in participants' self-perceived competencies than the original seminar, with participants in training groups 1 and 2 showing greater improvements than participants in training group 0.
13.	The instructionally revised simulation training leads to greater improvements in participants' self-perceived competencies than the original simulation training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively.
14.	Training group 2 receives the instructors' best performance ratings, followed by training groups 1 and 0, respectively.

The first hypothesis on the learning level (*Hypothesis 10*) concerns the occurrence of a response-shift. It is based on findings in the training literature (Howard & Dailey, 1979; Sprangers & Hoogstraten, 1989) that found that self-report measures, especially if trainees lack adequate information to rate their level of functioning before the intervention, are subject to a change of their internal standard of measurement.

Hypothesis 11 is the general training intervention hypothesis on the learning level, stating that the trainees participating in the revised curriculum learn more than the trainees in the original curriculum. Learning in this study is thus operationalized as the increase from the respective pre- to the posttest-scores. Hypothesis 11 was further broken down to the two separate interventions of the seminar and the simulation sessions, leading to hypotheses 12 and 13. *Hypothesis 12* states that the two groups (training groups 1 and 2) that received the revised form of the preparatory seminar, learn more than the group that received the original version (training group 0). *Hypothesis 13* states that the revised simulation training curriculum leads to greater improvements than the original one. This hypothesis reflects the two redesign phases in stating that training group 2 also learns more than training group 1 (cf. section 3.3.2.3.2).

Based on the same considerations, *Hypothesis 14* states that the two curricular revisions also lead to more other-reported learning than the original version. This hypothesis was specifically formulated to determine if the possible self-reported learning effects could also be found using more objective ratings by the instructors as observers.

3.4.1.3 Transfer level hypothesis

On the transfer level, the following hypothesis was developed (Table 13).

Table 13. *Hypotheses concerning the students' transfer*

Hypothesis number	Hypotheses: Training transfer
15.	Members of training group 2 receive the best OSCE scores, followed by members of training group 1 and training group 0, respectively.

Based on the two phases of the curricular redesign process, *Hypothesis 15* proposes that training group 2 shows the highest transfer, training group 1 shows less transfer, and training group 0 shows the least amount of transfer. The hypothesis is based on the training model and the resulting two phases of the training redesign, as detailed in sections 3.3.2.3.1 and 3.3.2.3.2. It was specifically assumed that training groups 1 and 2 would show more transfer because their simulation sessions were designed based on an integrated approach to learning, following principles of the cognitive apprenticeship approach (Collins et al., 1989) that have been shown to improve transfer (Gräsel & Mandl, 1993). Training group 2 was furthermore supposed to show more transfer than training group 1 because they were assumed to additionally benefit from the revised curricular sequence of providing an overview of the complete task first, then practicing relevant subskills, and then integrating these skills again in the full-scale simulation session when practicing the complete task again (Honebein et al., 1993; Murray et al., 2005).

3.4.1.4 Individual characteristics hypotheses

The next set of hypotheses was developed to investigate the individual characteristics specified in the training model as input variables (Table 14). This set of hypotheses deals only with those characteristics that were to be influenced by the redesigned training interventions.

Table 14. *Hypotheses concerning the students' individual characteristics*

Hypo-thesis number	**Hypotheses: Individual characteristics**
16.	Training groups 1 and 2 have higher expectations concerning their active participation in the simulation sessions and the application of their theoretical knowledge in the simulation sessions than group 0.
17.	Training groups 1 and 2 perceived a greater opportunity to actively participate and apply their theoretical knowledge in the simulation sessions than group 0.
18.	Training groups 1 and 2 show higher expectation fulfillment than group 0.
19.	Training groups 1 and 2 have higher instrumentality expectancies concerning the simulation sessions than group 0.
20.	Training groups 1 and 2 are less performance goal oriented and more learning goal oriented than group 0.
21.	Training groups 1 and 2 were readier for the full-scale simulation training than group 0.

Hypotheses 16-18 deal with the students' specific didactic expectations, perceptions, and the resulting expectation fulfillment (cf. section 3.3.2.1.1.5). It was assumed that due to the redesign of the preparatory seminar, groups 1 and 2 would develop higher and thus more realistic expectations concerning their active participation and knowledge application in the simulation sessions than training group 0 whose expectations were not guided in the original version of the seminar. Group 1 and 2's expectations were influenced in the revised seminar through the explication of the learning contract and the transfer anchor (Cognition and Technology Group at Vanderbilt, 1990). Due to the standardization and redesign of the revised simulation sessions, it was also assumed that the students perceived more opportunities for the their active participation and knowledge application (*Hypothesis 17*). The resulting better match between the students' expectations and perceptions should consequently lead to better expectation fulfillment scores for training groups 1 and 2 than for training group 0 (*Hypothesis 18*).

The revised version of the seminar also aimed at influencing the students' instrumentality expectancies (Lawler, 1973; Vroom, 1964; cf. section 3.3.2.1.1.2) concerning the simulation sessions, again through the explication of the learning contract. *Hypothesis 19* thus states that training groups 1 and 2 rate the simulation sessions as more instrumental than training group 0.

As explicated in *Hypothesis 20*, the revised version of the seminar also aimed at influencing the students' situational goal orientation (Dweck, 1986; cf. section 3.3.2.1.1.3). It was therefore assumed that training groups 1 and 2 are less performance goal oriented and more learning goal oriented than training group 0.

Hypothesis 21 refers to the students' readiness for the simulation training (Holton, 2005, cf. section 3.3.2.1.1.4). Based on the clarification of the learning goals, the learning contract, and the transfer anchors, it was assumed that training groups 1 and 2 would be readier for the simulation sessions than training group 0.

3.4.1.5 Hypotheses about relations within the model

The following hypotheses concern a set of possible relations within the model (Table 15). They were exploratory in nature as the focus of the study lay on the test of the designed interventions and not on the validation of assumed relationships between the model's elements.

The hypotheses were derived from the presented training model (Schaper, Schmitz et al., 2003), from Holton's evaluation model (2005), as well as from Cannon-Bowers et al.'s model (1995). Care was taken not to overdetermine the resulting model by choosing only a small set of relations to be tested.

Table 15. *Hypotheses concerning the relations within the model*

Hypothesis number	**Hypotheses: Training reactions**
22.	Individual characteristics variables (readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy) predict training reaction scores.
23.	Individual characteristics variables predict learning post-training scores.
24.	Individual characteristics variables predict learning difference scores.
25.	Learning post-training scores mediate the relationship between individual characteristics variables and behavior.
26.	The students' self-reported anesthesia skills at the end of the training ($\text{learning}_{\text{post}}$) predict their transfer performance (behavior).

In *Hypothesis 22* it was assumed that the individual characteristics variables of student readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy predict the students' reaction scores. This assumption was based on Holton's (2005) model in which, for example, students' motivation is supposed to influence their reactions to the training, as well as on Cannon-Bowers et al.'s model (1995) in which, for example, expectation fulfillment was shown to influence training reactions.

The next hypotheses concern student learning. Learning is further differentiated in learning post-training scores and learning differences scores. In the training literature, learning is often only operationalized as posttest performance, for example in the meta-analysis conducted by Alliger et al. (1997). In a training context it seems, however, also important to operationalize learning as the difference between the pretest and the posttest scores as a high or low posttest score does not contain information on possible improvements the trainee made. Consequently, this differentiation was taken into consideration in the developed hypotheses.

Hypothesis 23 assumes that a set of individual characteristics influences the learning post-training scores, an assumption that is also found in Holton's (2005), Cannon-Bowers et al.'s (1995), and Baldwin and Ford's (1988) model. In addition, exploratory *Hypothesis 24* states that the individual characteristics variables also influence the learning differences scores.

Hypothesis 25 states that the individual difference variables also influence behavior, that is, the students' transfer performance in the OSCE, but that they do so through the students' posttest scores. That is, the students' posttest scores are assumed to mediate the relationship between the individual difference variables and the transfer measure (behavior), a relationship that was also in parts assumed in Holton's (2005), Cannon-Bowers et al.'s (1995), as well as Baldwin and Ford's (1988) models.

Hypothesis 26 postulates that the students' posttest ratings predict their transfer scores, as was again assumed in Holton's (2005) and Cannon-Bowers et al.'s (1995) models.

3.4.2 Study design

The study had to be conducted in the context of the mandatory curriculum described above (3.1, 3.2). The general design considerations (3.3.1.4.1) indicated that a non-equivalent control group quasi-experimental design with pre-, post-, and transfer-test design could be employed to test the aforementioned hypotheses. The response-shift considerations (3.3.1.4.2) led to the inclusion of a retrospective pretest. Based on the research question, learning objectives, designed interventions, and the restrictions described above, the following design was developed (Figure 3).

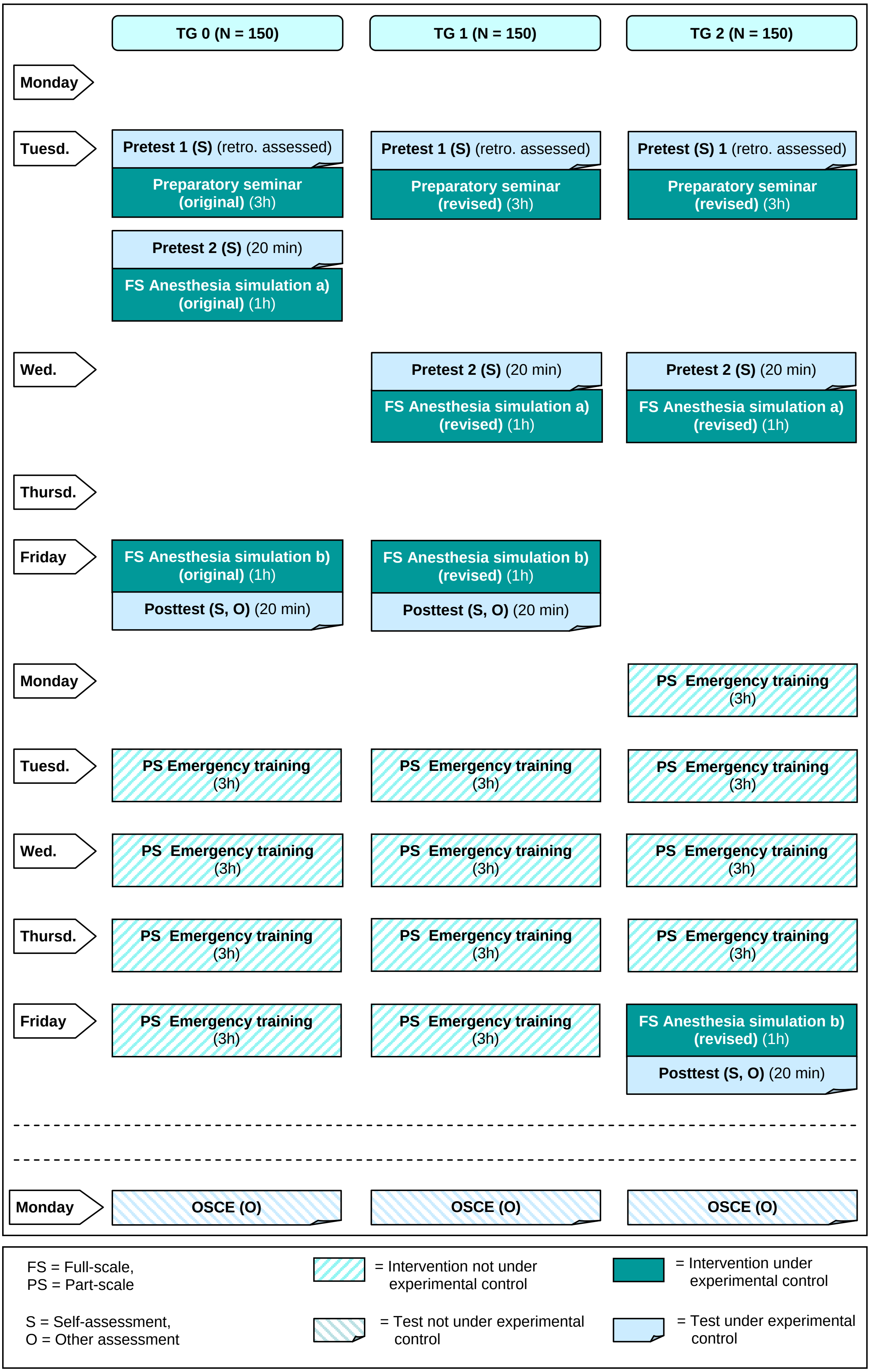

Figure 3. Design of the evaluation study

The above figure shows the design of the study with the relevant interventions and tests. Solidly colored elements were directly controlled within the study, striped elements were part of the study but could not be experimentally controlled. The different elements also contain their respective durations.

Duration of the curriculum

As the figure shows, the training interventions took place within two back-to-back weeks of the 14-week surgery module. Since students rotated through the different blocks, such as the two-week anesthesiology block, in sub-groups, the block was conducted six times within the surgery module.

The OSCE is conducted over all of the topics covered in the module and takes place in the last week of the module (week 14).

The rotation systems also entails that the time the students have from the last simulation input to the OSCE varies from two days to 12 weeks.

Interventions and tests

Overall, the study included three different types of *interventions* for the students that have been described above: The preparatory introductory anesthesiology seminar (1 x 3 hours), the full-scale anesthesia simulation sessions (2 x 1 hour), and the part-scale emergency medicine trainings (4 x 3 hours).

The study furthermore included several different *tests* to evaluate the effectiveness of the curriculum. Students were tested at three points of measurement. They filled out the *pretest questionnaire* before entering the first full-scale simulation session (FS anesthesia simulation a). This pretest questionnaire assessed their competencies at two different points in time: They were first asked to rate their competencies as they were before they had participated in the preparatory seminar (pretest 1, retrospectively assessed), then they were asked to rate their competencies at the current point in time, that is, after their participation in the seminar (pretest 2).

The *posttest questionnaire* was filled out immediately after the students' participation in the second full-scale simulation session (FS anesthesia simulation b). The posttest questionnaire furthermore contained a retrospective rating of pretest 2 to check for a response-shift. Like the pretest, the posttest questionnaires were self-report ratings on the individual level. At the posttest point of measurement, a further rating was completed by the instructor teaching the respective session. He rated the students' performance in the simulation session on a group level. The pre- and posttest self-report questionnaires were especially designed for the purpose of this study, as were the instructor questionnaires.

At the completion of the surgery block (week 14), all of the students participated in the OSCE which served as the *transfer-test*. The OSCE was conducted as part of the mandatory curriculum. Since this exam counted for the students' medical school report, it could not be altered as part of the study.

The evaluation instruments are described in more detail in section 3.4.2. The following paragraphs detail the resulting designs for the three different training groups.

Design for the three training groups

Training group 0 received the anesthesiology curriculum in its original version. Original version means that the different interventions were developed without any specific instructional design considerations (see 3.2). On Tuesday morning, training group 0 participated in the original three-hour preparatory introductory anesthesiology seminar (cf. 3.2.1). After their lunch-break, they completed pretest 2, which also contained the retrospectively assessed pretest 1, and then participated in full-scale simulation session a). On Friday of the same week, they participated in full-scale simulation session b) and filled out the posttest immediately thereafter. The instructor also rated the group's performance in the simulation session.

In the following week, training group 0 participated in the part-scale emergency training (Tuesday to Friday).

In week fourteen, they participated in the OSCE exam as the transfer test.

Training group 1 participated in a revised version of the anesthesiology curriculum (see 3.3). On Tuesday morning, they attended the instructionally revised version of the preparatory seminar. On Wednesday afternoon, they first filled out the pretest and then participated in the instructionally revised simulation session a). As for training group 0, the pretest questionnaire contained assessments of their competencies before the seminar (pretest 1) and now, after the seminar and before the simulation session (pretest 2).

Following simulation session b), they filled out the posttest, and the instructor also rated the group's performance in the simulation session.

Like group 0, they participated in the part-scale emergency training in the following week.

Again, like group 0, they participated in the OSCE exam to assess their transfer in week fourteen.

Training group 2 received the same instructionally revised curricular elements as training group 1 did. The design for group 2 was developed after training group 1 had completed the curriculum (see 3.3.2.3). Training group 2 participated in the same preparatory seminar as group 1 and also completed the pretest before simulation session a) on Wednesday afternoon.

Then they participated in the part-scale emergency training sessions from Monday to Thursday of the following week.

On Friday of the second week, they participated in full-scale anesthesia simulation b), after which they completed the posttest. The instructor also rated the group's performance in the simulation session.
Like the other groups, they participated in the OSCE transfer test in week fourteen of the curriculum.

Design limitations
The inspection of the above illustrated study design reveals several limitations. These limitations arose from the applied setting of the study that contained several restrictions. The first restriction was, that the students could only be assessed at the described three points of measurement due to other evaluation efforts that were taking place at the same time, initiated by the medical school. This meant that pretest 1 could only be assessed retrospectively. Together with the restriction concerning the study plan due to the working schedules of the teaching faculty described above, it further meant that another posttest was missing. The second posttest should have been included after the part-scale emergency training sessions.
Due to time and resource restrictions, it was furthermore not possible to rate the students' individual performance by the instructors or other observers. These ratings could thus only be collected on a group level.

3.4.3 Evaluation instruments

The following sections describe the instruments that were used to assess the evaluation criteria (cf. 3.3.1.3). Section 3.4.3.1 describes the developed questionnaires, followed by the description of the OSCE (3.4.3.2), and concluding with the resulting instruments used at the different times of measurement (3.4.3.3).

3.4.3.1 Evaluation questionnaires

The following sections first describe the development of the student questionnaires (3.4.3.1.1), followed by a description of the developed instructor questionnaires (3.4.3.1.2).
The questionnaires had to be developed because no instruments were specified in the respective literature that could be used to assess the above mentioned research question. An exception was the goal orientation measure (see 3.4.3.1.1.4).
The underlying multi-step development process was the same for the student and instructor questionnaires. The scales were constructed rationally: Based on the theoretical considerations described in the model, on similar instruments reported in the literature, on the conducted needs assessment, the derived learning objectives, and evaluation criteria, a list of items was developed. This list was then reviewed with the SMEs for the items relevance and adequacy. The best-suited items were selected and then pretested with a student worker at the clinic of anesthesiology. The student worker rated them for their comprehensibility and clarity. The

items that had been accepted at all of the stages of the review process were then included in the resulting questionnaires.
The items were to be answered on a 6-point Likert scale. Following the suggestions of the SMEs, the poling of the Likert scale reflected the German school grading system. 1 thus indicated the most agreement or best rating (“totally agree”, “optimal”); 6 indicated the least agreement (“do not agree at all”, “mostly not”). This poling was chosen because the participants were used to it based on their school experience and based on the other evaluation instruments they needed to fill out in the medical school.

3.4.3.1.1 Student questionnaires

Based on the theoretical considerations explicated in the training model, the needs assessment, the derived learning objectives, and the evaluation criteria, the following scales were included in the student questionnaires: anesthesia competencies (3.4.3.1.1.1), training expectations, perceptions, and expectation fulfillment (3.4.3.1.1.2), instrumentality expectancy (3.4.3.1.1.3), situational goal orientation (3.4.3.1.1.4), students’ readiness (3.4.3.1.1.5), self-efficacy (3.4.3.1.1.6), training reactions (3.4.3.1.1.7), and demographics (3.4.3.1.1.8).

3.4.3.1.1.1 Self-perceived anesthesia competencies

The main criteria to be assessed in the evaluation study were student learning and transfer in the area of anesthesia competencies. Based on the above described design and its inherent limitations, it was decided to assess learning with the help of self-report questionnaires.

Since no instrument that assesses skill-based elementary anesthesia competencies could be identified in the conducted literature search, a new scale had to be developed based on the simulation sessions’ learning objectives. The scale development followed the steps described above (3.4.3.1).

After the rational item reduction, 14 items remained to be included in the questionnaire (Table 16). Based on the learning objectives, the items cover cognitive, interpersonal, and psychomotor elements. Taken together, these elements represent the basic competency to conduct anesthesia.

Table 16. *Self-perceived anesthesia competency scale items*

Item number	**Questionnaire item** How do you rate your competency…
1	to gather relevant information through communicating with the patient.
2	to administer a general anesthesia.
3	to administer the necessary drugs for a standard anesthesia.
4	to ventilate a patient with a mask.
5	to determine the exact time for an endotracheal intubation.
6	to conduct an endotracheal intubation.
7	to use interdisciplinary knowledge to form preliminary diagnoses.
8	to determine the following preliminary diagnoses (using data stemming from anamnesis, clinical exam, and monitoring): Myocardial infarction, asthma attack, anaphylaxis, pulmonary embolism.
9	to determine the above listed differential diagnoses.
10	to develop therapy plans for the differential diagnoses listed above.
11	to administer the appropriate drugs for the differential diagnoses listed above.
12	to determine the appropriate drug dosage for the differential diagnoses listed above.
13	to actively gather information about a possible drug contraindication.
14	to efficiently distribute tasks among team members.

Note. For the original German questionnaire items refer to appendix C.

The self-perceived competency items were to be rated on the six-point Likert-type scale following the German school grading systems as displayed in the following table (Table 17).

Table 17. *Self-perceived competency items rating scale*

Rating scale					
① Optimally existent	② Mostly existent	③ Satisfactorily existent	④ Partly existent	⑤ Hardly existent	⑥ Mostly not existent

The self-perceived competency scale was included in the pretest- and the posttest-questionnaire (part A).

3.4.3.1.1.2 Students' training expectations, perceptions, and expectation fulfillment

Based on Cannon-Bowers et al.'s (1995) model, students' training expectations were examined as one of the motivation variables. Items were generated to tap the didactic aspects of the training method (cf. Cannon-Bowers et al., 1995), namely active participation and knowledge application. In addition to the respective literature findings, these aspects reflect the SMEs' teaching experiences. Previous semesters had shown that the students did not seem to expect that they had to participate actively and apply what they had learned theoretically.

After the item generation and selection process, 8 items remained to gather students' expectations concerning simulation sessions (Table 18).

Table 18. *Expectation scale items*

Item number	Questionnaire item I expect from the HANS simulation sessions to be able to…
1	actively participate in the lesson.
2	apply my theoretical knowledge.
3	work independently.
4	learn from my mistakes.
5	develop diagnoses in a team.
6	carry out tasks of an anesthesiologist.
7	realistically monitor the patient.
8	try out my acquired competencies in anesthesiology.

Note. For the German items used in the questionnaire refer to appendix C.

The students were asked to indicate their agreement to the above items using the following Likert-type 6-point rating scale (Table 19).

Table 19. *Expectation/perception items rating scale*

Rating scale					
① Strongly agree	② Mostly agree	③ Rather agree	④ Rather disagree	⑤ Mostly disagree	⑥ Strongly disagree

The students' expectations regarding the simulation sessions were assessed in part B of the pretest questionnaire.

As described in Cannon-Bowers et al.'s (1995) model, parallel items to those in the pretest were used in the posttest (part B) to assess trainees' perceptions regarding the didactic aspects of the training. The item stem was accordingly changed to "In the HANS simulation sessions, I was able to…".

With the students' expectations collected in the pretest and their perceptions of the training collected in the posttest, their training fulfillment can be investigated (Tannenbaum et al., 1991). For this purpose, the obtained scores will be used to generate an expectation fulfillment score as a function of expectations and perceptions[10](see Cannon-Bowers et al., 1995; Tannenbaum et al., 1991):

$$Ef = \sum_{i=1}^{j} = (E_i - P_i)$$

where EF = total expectation fulfillment score, i = item, j = number of expectation – perception item pairs, E = expectations, and P = perceptions, rendering an expectation score

[10] Data for students' desires, reflecting the importance they assign to the assessed concepts were not collected due to time restraints.

that could vary between –5 and +5. If expectations exceed perceptions on a German school grading scale, expectations will have numerically lower ratings than perceptions, thus resulting in a negative fulfillment score. If perceptions exceed expectations, perceptions will have numerically lower ratings, thus yielding a positive fulfillment score. In addition to the total expectation fulfillment score, expectation fulfillment scores will be calculated for each item or factor.

3.4.3.1.1.3 Instrumentality expectancy

Training motivation was primarily assessed using a valence-instrumentality-expectancy approach (Lawler & Suttle, 1973; Vroom, 1964). Following the suggestion of Mathieu and colleagues (1992) only the performance-outcome expectancies were operationalized in this questionnaire section. This resulted in an instrumentality approach investigating trainees' perceptions that doing well in the training would lead to better outcomes (performance-outcome association; expectancy II). The effort-performance expectations were not separately operationalized as they are related to self-efficacy expectations which were assessed in a different scale (cf. Noe, 1986).

When using measures of performance-outcome expectancies, it is important to use items which assess trainees' beliefs concerning training performance levels and the attainment of both intrinsic and extrinsic outcomes (Noe, 1986).

Following these considerations, a list of items was developed based on items presented by Mathieu and colleagues (see also Holton, 2005; Mathieu et al., 1992). They used items in the format of "Successful completion of this training course will help me produce higher quality work" (p. 835). Items used in the present study were developed to measure the relationship between doing well in the training and future performance (extrinsic outcomes) as well as optimal learning (intrinsic outcome). Again, the development process followed the steps described above (3.4.3.1).

Table 20. *Instrumentality expectancy scale items*

Item number	Questionnaire item
1	If I learn a lot during the simulation sessions, I will have a better chance to succeed in the final exam.
2	Through the simulation sessions I will acquire skills that will be useful to me in my applied clinical work.
3	In respect to the final exam, it is better to participate in the simulation sessions than reading a book on the subject would be.[1]
4	I consider the simulation sessions to contain a learning opportunity that I will not have in any other form of instruction.

Note. For the German items used in the questionnaire refer to appendix C.

The students were asked to indicate their agreement with the instrumentality expectancy items on a 6-point Likert-type scale (see Table 19).

The instrumentality expectancy scale was included in the pretest questionnaire (part C1).

3.4.3.1.1.4 Situational goal orientation

The current study also aimed at investigating if the students' goal orientation could be influenced by an instructional intervention. In the goal orientation area, several instruments exist to collect the respective data. Since it was assumed that the students' goal orientation could be influenced by an intervention, the concept of situation goal orientation was assessed. Button (1996) has developed a situational goal orientation scale. It was decided to use this measure because it is targeted at learning experiences in general and conceptualized to assess situational goal orientation. The items were received by a personal correspondence from Scott Button. They were translated and the following minor adjustments were made to assure that the items matched the curriculum parameters: Item 8 was changed from "Who is going to see my results?" into "Will my performance today influence my course grade?" since the results were already visible during the simulation sessions. Item 3 ("I hope this isn't something I already know.") was further specified by rewording it into "I hope we won't cover material I already know". The instruction for rating the questions (see appendix C) was also based on the instruction used by Button. The scale was changed from Button's 9-point scale to the six-point scale that was also used for the other items of the questionnaire. The item sequence was furthermore changed so that performance and learning goal items were not presented in a block but alternated. The used items are displayed in Table 21.

Table 21. *Situational goal orientation scale items*

Item number	Questionnaire item
1	Am I going to learn something?
2	I hope I don't make any mistakes.
3	I hope we won't cover material I already know.
4	I wonder how my scores will compare to others?
5	I hope there are some tough parts that I really get to work at.
6	I hope my performance is up to par.
7	Will my performance today influence my course grade?
8	I'd like to get a chance to discuss my mistakes with others.
9	Will I look competent?
10	I'm eager to get started and try to figure things out.

Note. For the German items used in the questionnaire refer to appendix C.

Again, the students were asked to indicate their agreement with the situational goal orientation items on a 6-point Likert-type scale (see Table 19).

The items were included in the pretest questionnaire (part D).

3.4.3.1.1.5 Students' readiness

Based on the training model, the trainees' readiness for the training was assessed as another individual characteristic. The scale was constructed based on Holton's (1996; 2005) Learning Transfer System Inventory (LTSI) and adapted where necessary. Holton constructed the scale as a "training specific scale", highlighting the importance of adapting the items to the training. Again, the development followed the steps described above (3.4.3.1). The process resulted in a set of five items that were included in the questionnaire (Table 22). The items covered the students' general knowledge of the goals of the full-scale simulation sessions and their readiness to engage in practice-oriented interventions. Following Holton (2005), the readiness items were included in the posttest questionnaire and thus assessed the students' readiness retrospectively. This procedure was deemed adequate as the students' are better able to assess their readiness after they have participated in the training and thus have, for example, experienced the effects the training had on their skills.

Table 22. *Readiness scale items*

Item number	Questionnaire item
1	I knew before the HANS-sessions which effects the sessions should have on my skills.
2	I knew up front what I could expect from the HANS sessions.
3	I was aware of the HANS-sessions' expected results from the beginning.
4	As far as I know, the skills conveyed in the HANS sessions reflect the examination requirements.
5	Usually, practice helps me to increase my performance.

Note. For the German items used in the questionnaire refer to appendix D.

Again, the same 6-point Likert-type scale was used for the indication of the students' agreement (see Table 19).

The items were retrospectively assessed in the posttest (part C1).

3.4.3.1.1.6 Self-efficacy

Another motivational variable that was assessed, was the students' self-efficacy. When investigating self-efficacy in the realm of training interventions, the employed measures should aim at effective responses to learning and change (Noe, 1986). Schwarzer (1994) further points out that the selection of the degree of generality with which self-efficacy is operationalized depends on the study's goals. For this study, a task-specific operationalization was chosen because the competencies of interest were very specific anesthesia competencies. Based on the above mentioned learning goals and the conducted task analysis, key anesthesia tasks were identified. These tasks were then used in the construction of the task-specific self-efficacy scale (cf. Chen, Thomas, & Wallace, 2005). Following Bandura's conceptualizations, the self-efficacy items were designed to include cognitive expectancies concerning future behavior, as well as emotional and motivational aspects (Bandura, 1997). Again, the development was conducted in the described fashion (3.4.3.1). The process yielded the following self-efficacy scale (Table 23).

Table 23. *Self-efficacy scale items*

Item number	Questionnaire item
1	I'm convinced that I will achieve the learning goals of the anesthesia simulation sessions.
2	I know exactly that I am able to accomplish the tasks during the simulation sessions.
3	I remain calm anticipating possible difficulties during the simulation because I can trust my own abilities.
4	I will be able to apply the content of the simulation sessions well.
5	I'm convinced that I will succeed in using my interdisciplinary medical knowledge to form a diagnosis.
6	On average, I'm probably more capable than others to complete anesthetic tasks.
7	I know that I will succeed in independently conducting a standard anesthesia.
8	I will not need a long time to deduct and justify diagnoses, such as myocardial infarction, asthma, pulmonary embolism, ...

Note. For the German items used in the questionnaire refer to appendix C.

The same rating scale was used as shown in Table 19.

The self-efficacy scale was included in the pre-test questionnaire (part C2).

3.4.3.1.1.7 Reaction items

The questionnaire was to furthermore contain a set of reaction items to assess this first level of training effectiveness. Again, these items had to be specifically constructed for the present training, following the process outlined above (3.4.3.1). The simulation and training literature was reviewed to generate a list of reaction items (e.g., Alliger et al., 1997; Cannon-Bowers et al., 1995; Devitt et al., 2001; Helmreich & Merritt, 1998).

Based on the research questions and the design, three different reaction aspects emerged: the students' general reactions to the curriculum and its elements (3.4.3.1.1.7.1), the students' reactions to the curricular sequence (3.4.3.1.1.7.2), and their reactions to the instructors' teaching skills (3.4.3.1.1.7.3).

3.4.3.1.1.7.1 General student reactions

After the revision process (3.4.3.1), 12 items remained in the scale. They investigated the students' reactions to the simulations and the simulator, to the used scenarios, and their overall satisfaction with the program (Table 24).

Table 24. *General reaction scale items*

Item number	Questionnaire item
1	The HANS-simulation was an important component of the Heicumed curriculum.
2	I can recommend the use of the HANS-simulation in its current form for the future.
3	The simulation allowed me to work on relevant problems in a realistic situation.
4	The scenarios' level of difficulty was adequate for my level of knowledge.
5	The used scenarios motivated me to further work on the content of the anesthesiology curriculum.
6	The simulator's realism was motivating.
7	I could easily adjust to the simulation.
8	I consider the simulator a suitable learning tool for me.
9	I did not feel restrained in the simulation situation.
10	I could acquire skills using the simulator that I could not have acquired using different training tools.
11	The simulator's realism was helpful for learning the content.
12	Overall, the HANS training fulfilled my expectations.

Note. For the German items used in the questionnaire refer to appendix D.

Again, the students were asked to indicate their agreement with the reaction items on the 6-point Likert-type scale used for the other items (see Table 19).

The general reaction items (Table 24) were included in the posttest questionnaire as part C2.

3.4.3.1.1.7.2 Student reactions to curricular sequence

Due to the changes in the curricular sequence for training group 2, four additional items were included as a result of the second redesign phase. They were used to assess the students' reactions to the course structure, that is, the new sequence and the coherence of the contents covered in the different curricular elements (Table 25). Reactions to the course structure have previously been identified as one factor underlying reaction measures (Morgan & Casper, 2000).

Table 25. *Reaction to seminar sequence scale items*

Item number	**Questionnaire item**
1	The sequence "Seminar" – "HANS I" – "Emergency Training" – "HANS II" was appropriate.
2	The contents of the above elements were well coordinated.
3	The contents of the "Emergency Training" prepared me well for "HANS II".
4	Due to the experience gained in the "Emergency Training" I profited more from "HANS II" than from "HANS I".

Note. For the German items used in the questionnaire refer to appendix D.

The same scale was used to rate the answers to these items (see Table 19).

The seminar sequence items were contained in group 2's posttest questionnaire only (part E). This exclusive inclusion was based on the intervention's development process that had to be conducted in two phases (cf. 3.3.1.4.1, 3.3.2.3.2).

3.4.3.1.1.7.3 Student reactions to instructor's teaching skills

Overall, the assessment of teaching behaviors can be accomplished by using three sources of data: Students (either as observers or indirectly through the assessment of their learning), observers, and self-reports (Hartman & Nelson, 1992). For this study, the assessment of the instructors' teaching skills was operationalized in form of a reaction measure, thus using the students as observers. The items were developed based on existing scales and conceptualizations (Bartlett, 1982; Borich, 1989; Borich, 1992; Mayo et al., 1993), as well as on considerations of the instructional strategy and the according instructor role in the present training.

The 14 items resulting from the development and review process tapped the instructors' structuring skills, questioning skills, scaffolding skills, adaption skills, motivating skills, as well as their personal competence and motivation (Table 26).

Table 26. *Reaction to instructors' teaching skills scale items*

Item number	Questionnaire item
1	The instructor's questions were understandable.
2	The instructor provided for a structured lesson.
3	Through the instructor's feedback I received specific cues about how to improve my performance.
4	The instructor clarified the lesson's objectives.
5	It was important to the instructor that the group comprehended the content.
6	The instructor adapted the lesson to the group's needs.
7	The instructor motivated me to further contemplate the cases.
8	The instructor presented the content clearly and understandably.
9	The instructor created a pleasant learning climate.
10	The instructor considered the group's knowledge level when providing explanations.
11	The instructor questioned the group's statements.
12	The instructor was competent.
13	The instructor's practical guidance was appropriate.
14	The instructor was motivated.

Note. For the German items used in the questionnaire refer to appendix D.

Again, the same answering scale was used as before (see Table 19).

The above items were included in the posttest questionnaire of every group as part D.

3.4.3.1.1.8 Demographics

The questionnaire was also to include items assessing the demographic characteristics of the participants. These items were going to serve to determine the comparability of the three training groups.

The participants' age and gender were assessed, as well as their clinical experience. The participants first had to indicate whether or not they possessed prior clinical experience (item 3). If they had previous experience, they were asked to subsequently indicate which type, in an open-ended question format (see Table 27).

Table 27. *Demographic items*

Item number	Questionnaire item
1	Age (in years)
2	Gender (male / female)
3	Previous clinical experience (yes/no)
	Type of clinical experience (open answer)

Note. For the German items used in the questionnaire refer to appendix C.

The items were to be answered in the format indicated in the parentheses (Table 27). The type of clinical experience was to be specified in an open-answer format if the students indicated that they did possess previous clinical experience.

The demographic items were contained in the pretest questionnaire (Part E).

3.4.3.1.2 Instructor questionnaires

In addition to the student questionnaires used to assess students' self-reported learning, instructor questionnaires were used. The instructor questionnaires were designed to provide an external rating of the students' performance in the simulator.
The instructors filled out the questionnaire directly following simulation session b) (cf. Figure 3.), that is, at the same time, the students filled out the posttest questionnaire.
The number of items that could be included in the instructor questionnaire was unfortunately very limited due to time constraints the instructors faced during the teaching sessions. The instructors were only available for three minutes in between different student groups. The items contained in the instructor questionnaires were also specifically developed for the present study. Their development followed the process described in 3.4.3.1.

The resulting instructor questionnaire contained six items tapping the student groups' performance in the simulator. These items were selected by the SMEs from an item pool of group performance items. Selection criteria were the items' relevancy and importance as well as ease of observation. The items assessed the students' preparation, active participation, motivation, homogeneity, patient skills, and cooperation (Table 28).

Table 28. *Instructor ratings of student performance items*

Item number	**Questionnaire item** The group was…
1	was well prepared.
2	participated actively.
3	was motivated to learn.
4	was homogenous concerning their level of proficiency.
5	could approach the patient.
6	cooperated among themselves.

Note. For the German items used in the questionnaire refer to appendix E.

The instructors rated the above items on the same 6-point Likert-type rating scale used in the student questionnaire (see Table 19).

3.4.3.2 Objective standardized clinical examinations (OSCE)

To assess transfer of learning on level three, *behavior*, exam scores were taken from the regular OSCE (Objective Standardized Clinical Examination) every student had to participate in as part of the degree program. These examinations were not under the control of the study designer as they were part of the mandatory curriculum and the grades counted towards the students' report. This also implied that the questions could not be matched to the questionnaire items.

Studies investigating the OSCE found it to be a reliable and valid method for assessing clinical knowledge and skills in anesthesia and other medical areas (e.g., Cohen, Shay, Norden, & Tome, 1997; Donnelly, Sloan, Plymale, & Schwartz, 2000; Sloan, Donnelly, Schwartz, & Strodel, 1995). OSCE performance is usually summarized in an overall score, representing a combination of history, physical examination, interpersonal skills, and technical skills (Donnelly et al., 2000).

In this study, medical school faculty assessed the students' knowledge and skills in 6 different stations. The content of each curriculum block was thus tested in a separate station, with the anesthesiology content tested in one of the stations. The scores used as the transfer measure were taken from the anesthesia station of the OSCE.

The anesthesia station consisted of different questions the students had to answer as well as different tasks they had to perform. Questions covered, for example, issues of appropriate drugs, differential diagnoses, and therapeutic plans applied to new patient problems. Tasks the students had to perform included, for example, specific aspects of patient examination, patient ventilation (e.g., mask ventilation and intubation), and drug administration. These tasks had to be performed on a part-scale simulator. The type of transfer that was assessed by the questions and tasks the students' had to answer / fulfill can thus be described as *performance after training* (Cannon-Bowers et al., 1995). It not only contains aspects of learning and retention but also assesses if the learners are able to apply their knowledge to different tasks and in different settings after a period of time has elapsed. In the OSCE, the students first of all had to apply the theoretical knowledge they acquired in the seminar and the simulation sessions to new patient problems as presented in the questions the instructors asked. Secondly, they had to apply the anesthesia competencies they had acquired in the simulation sessions to a new type of simulator and to new types of patient problems.

Due to examination regulations of the medical school, the exact questions or tasks of the OSCE cannot be published.

The students were graded on a 0-25 point scale. All examiners used checklists to provide standardized evaluations. Each answer or performed task was scored using a 0 to 5 grading scale, with 0 indicating unacceptable answer/performance and 5 excellent answer/performance. Each student had to complete five questions or tasks and could thus attain a maximum of 25 points.

Each semester, the questions on the exam change, but are constructed to be parallel tests.

The OSCE took place in week fourteen of the module.

The type of transfer that was assessed was thus performance after training in a new setting and applied to new tasks (Cannon-Bowers et al., 1995). The students were thus required to retain the learned skills over a period of time and apply them to a different context.

3.4.3.3 Resulting pre-/post- and transfer-test instruments

The above described questionnaire scales and exams were combined for the study as displayed in the following tables (Table 29-30).

Table 29. *Resulting instrument structure: Students*

Time of measurement	**T_1**	**T_2**	**T_3**
Type of instrument	**Pretest Questionnaire**	**Posttest Questionnaire**	**OSCE**
Part A	Self-reported anesthesia competency scale (assessing present competencies (pretest 2) and competencies before the seminar (pretest 1)	Self-reported anesthesia competency scale (assessing present competencies and retrospectively competencies of pretest 2)	Anesthesiology questions and tasks
Part B	Training expectations	Training perceptions	
Part C1	Instrumentality expectancy scale	Student readiness scale	
Part C2	Self-efficacy scale	General reaction scale	
Part D	Situational goal orientation scale	Reaction to teaching skills scale	
Part E	Demographics scale	Reaction to seminar sequence scale[a]	

Note. [a] only included in the training group 2's questionnaire.

Note that the pretest questionnaire at t_1 was used to collect data for two points of measurement for the self-perceived anesthesia competencies: Pretest 1 before the seminar and pretest 2 before the simulation sessions and after the seminar.

Table 30 shows the questionnaire the instructors had to fill out at T_2.

Table 30. *Resulting questionnaire structure: Instructors*

Time of measurement	**T_2**
Type of instrument	**Posttest Questionnaire**
	Assessment of student group performance

3.4.4 Participants

As has previously been mentioned, students and instructors of the University of Heidelberg's medical school participated in the study. The following sections first describe the student participants (3.4.4.1), followed by a description of the instructor participants (3.4.4.2).

3.4.4.1 Student participants

The student participants completed the simulation curriculum as part of their required course work. Overall, three cohorts of students participated in the study. The first cohort of students participated in the summer semester of 2003 (training group 0), the second cohort in the winter semester of 2003/2004 (training group 1), and the third cohort (training group 2) in the summer semester of 2004. The cohorts had an anticipated size of about 150 students per semester.
The students were in their third year of medical school and had not received prior instruction in anesthesiology. Whereas participation in the curriculum was mandatory, participation in the study was not.

3.4.4.2 Instructor participants

Eight anesthesiologists participated in the study as instructors, two of which were board certified. All eight were clinical faculty who spent the majority of their time in patient care activities. In addition to their clinical work, they were also involved in teaching and had previous experience as simulation instructors.

3.4.5 Procedure

The following sections describe the data collection process. They are structured along the conducted tests. Section 3.4.5.1 illustrates the pretest, followed by the posttest (3.4.5.2), and the transfer test (3.4.5.3) descriptions.

3.4.5.1 Pretest

Immediately before attending simulation session one, the students were informed of the study and asked for their participation in the pre-, post-, and transfer-test. It was pointed out to them that participation was voluntary and not part of their course requirements, that the individual results would not be revealed to the instructors and would be treated confidentially.
They were told that the pre- and posttest data would be matched with the help of the code on the questionnaires' first page. They were furthermore informed that if they additionally revealed their names on the questionnaires, their self-evaluations would be compared with their OSCE grades for the transfer-test. The self-evaluations and grades would only be viewed by the primary researcher of the study and used only for research purposes. By disclosing their names, participants agreed to the recording of their grades.
The students then received the pretest questionnaires and were asked to fill them out in the following 15-20 minutes. It was pointed out to the students that they were asked to rate their anesthesia competencies in regard to two different points in time. They were furthermore encouraged to ask any questions they might have while filling out the questionnaire.

As has been mentioned in section 3.4.2, the training groups thus received the pretest after having attended the preparatory seminar. With the pretest questionnaire, data for two points of measurement were collected. Retrospective data were collected for the students' self-perceived competencies prior to the preparatory seminar (pretest 1) and for their current competencies (pretest 2) prior to the simulation sessions.

Upon completion of the pretest questionnaires, the students entered the simulator to participate in simulation session a).

3.4.5.2 Posttest

Upon completion of the second simulation session (simulation session b), the students were asked to fill out the posttest questionnaire before leaving the HANS simulation center. Again, the posttest took about 15-20 minutes to complete. The procedure was equivalent to the pretest. The instructions were slightly changed. This time it was pointed out to the students that the posttest questionnaire also asked them to come up with two different ratings of their competencies (posttest, and retrospective pretest 2). The general purpose of the retrospective pretest was briefly explained to them to improve acceptance.

At the same time, the instructor was asked to fill out the instructor questionnaire to rate student group performance.

Training groups 0 and 1 filled out the posttest questionnaires after the completion of the two full-scale simulation sessions and before their participation in the part-scale emergency training. Training group 2 participated in the posttest after having completed the two full-scale simulation sessions and the part-scale emergency training.

3.4.5.3 Transfer-test

As part of their regular course requirements, students participated in the OSCE at the end of the respective semester (week 14 of the surgery block). The examinations were conducted by medical school faculty according to the school's examination guidelines. The students had to answer questions and complete tasks using a part-scale simulator (cf. 3.4.3.2).

From the different stations contained in the OSCE only the attained anesthesia grades were used as the transfer-test data. The anesthesia station lasted 5 minutes.

Upon completion of the OSCE, the students' grades were transferred to the author by the curriculum coordinator to be included in the study.

3.4.6 Analytic strategy

The following sections briefly describe the statistical procedures used to present and analyze the data. Section 3.4.6.1 contains information on how the quality management of the data was conducted, section 3.4.6.2 describes how the questionnaire was analyzed, namely the employed factor analytic procedures, and section 3.4.6.3 illustrates the employed inferential methods.

All analyses were run using SPSS, version 11.5.1.

3.4.6.1 Data quality management

To ensure the validity of the results, the data quality is very important. The following steps were taken to secure the quality of the collected and entered data.

In the first step, the data set was checked for data entry errors and corrected where necessary.

In the second step, the scores were examined for obviously biased answer tendencies of individual participants (e.g., the same rating on every item of a scale). Participants with these answering strategies were excluded from the analysis of the particular scale and their ratings were substituted with missing values.

In the third step, the data set was checked for outliers. Although the data set contained outlying values, it was decided not to exclude subjects on this basis. As Bortz and Döring (1995) point out, it is important to collect data from a representative sample and the exclusion of outliers would have restricted the data set too much.

In the fourth step, the missing data problem was considered. Missing data can be due to experimental attrition, meaning subjects who participated in the first measurement, do not participate anymore in one or more of the following measurements. It can furthermore be due to subjects not answering all of the items in the questionnaire. Both causes can lead to a substantial decrease in sample size and to a reduction of the study's internal validity if the attrition is systematic. The threat of systematic attrition can unfortunately not be ruled out in this study. Due to the size of the present data set and the design-related unequal group sizes, the contained missing values were not substituted by the item mean in the analyses, as this procedure would not have led to equal cell sizes either.

3.4.6.2 Scale analysis

To investigate the psychometric properties of the questionnaire developed for this study, factor analytic procedures were employed.

Factor analyses are multivariate procedures that can be used to identify the latent construct structure of a set of variables and to thus represent a set of observed variables in terms of a smaller number of underlying variables or factors (Bortz, 1993). These extracted factors form

coherent, that is, highly correlated, subsets of variables and are usually orthogonal of each other (Bortz & Döring, 1995).
Factor analytic procedures often serve exploratory purposes to determine the scale properties of newly developed instruments. In this study, the factor analytic procedures were also used for exploratory purposes.

A factor analysis is generally structured following four steps (cf. Bortz, 1993; Janssen & Laatz, 1999; Kim & Mueller, 1994).

1. The basis of the factor analysis is the examination of the correlation matrix of the original scores. Based on these intercorrelations it can be decided which variables will be used for the further analysis.
2. In the second step, a number of factors are extracted, based on Pearson's product-moment-correlation coefficients. Bortz (1993) describes several methods for the extraction of the appropriate number of factors, among them are the following: the Kaiser-Guttmann criterion, after which factors with an eigenvalue greater than 1 are extracted; the visual inspection of the scree plot in which factoring should be stopped where the eigenvalues start to level off (Bortz, 1993); addition of factors if their inclusion explains substantially more variance, which might, however, lead to the extraction of many more factors than the previous methods; the conceptual investigation of the extracted factors, in which the solution offering the best interpretability should be favored (Rost & Schermer, 1989).
3. In the third step, the extracted factors are rotated to find a more meaningful or interpretable factor structure. Factors can be more easily interpreted, if their loadings approach simple structure, that is, if a number of variables correlate highly with the respective factor but do not correlate with other factors (Tabachnick & Fidell, 1996). Again, the interpretability of the resulting factor solution should be examined as an important test of the analysis (Tabachnick & Fidell, 1996).
4. As an optional fourth step, factor scores can be calculated and used for further analyses of the data or in future studies. They are estimates of "the score subjects would have received on each of the factors had they been measured directly" (Tabachnick & Fidell, 1996, p. 678).

For the current study, principal components analysis (PCA) with orthogonal rotation (Varimax) was used to obtain uncorrelated factors, which generally leads to more interpretable results. In principal components analyses the original set of observed variables is linearly transformed into a considerably smaller set of uncorrelated latent variables or factors that still represents most of the information contained in the original data set (Duntemann, 1994). The resulting factors thus still explain the maximum amount of variance (Bortz, 1993).
It is to be noted, however, that factor analytic procedures do not lead to one optimal factor solution, thus leaving room for various interpretations. As there are no objective criteria that

determine the selection of the appropriate solution, heuristics have to be used as guidelines (Bortz & Döring, 1995, p. 355). Following Guadagnoli and Velicer (1988), Bortz (1993, p. 484, 509f) recommends interpreting a Varimax-rotated factor structure only if

a) every meaningful factor contains at least 4 variables with loadings greater than .60 or
b) every factor contains at least 10 variables with loadings greater than .40 or
c) $n \geq 300$ if a factor contains less than ten variables with loadings greater than .40. If the sample is smaller than 300 the replication of the factor structure is to be awaited.

The interpretation of the respective factors is furthermore more valid if variables with high loadings on one factor have low loadings on the other factors (simple structure) (Revenstorf, 1976).

The following assumptions need to be met by the data to be factor analyzed: normal distribution of the variables, variables that are measured at least at the interval level and intercorrelated variables. Large samples are furthermore strongly recommended for factor analytic procedures (Bortz, 1993). If, as in the present study, factor analyses are only used for explorative purposes, the restrictions on factor analytic procedures are often relaxed in order to fully explore the data (Tabachnick & Fidell, 1996). Tabachnick and Fidell remark that "as long as PCA and FA are used descriptively as convenient ways to summarize the relationships in a large set of observed variables, assumptions regarding the distribution of variables are not in force. If variables are normally distributed, the solution is enhanced" (1996, p. 640).

3.4.6.3 Hypothesis tests

The following paragraphs briefly describe the statistical procedures used in this study to test the hypotheses. Section 3.4.6.3.1 deals with the methods used for the detection of a response shift, section 3.4.6.3.2 with the methods used to investigate the effects of the training, and section 3.4.6.3.3 with the methods used to exploratively assess the relations within the model.

3.4.6.3.1 Analyses of the response-shift

As has been described above, the posttest questionnaire also included retrospective ratings for pretest 2 to assess the possible occurrence of a response shift. Analytically, the occurrence of a response shift will be tested by comparing the pretest 2 scores collected in the pretest questionnaire with the retrospective pretest 2 scores collected in the posttest questionnaire. If a response-shift does not occur, the conventional pretest-posttest approach, or the retrospective-posttest approach can be used as a measure of change since they provide comparable estimates of the effects (Sprangers & Hoogstraten, 1989). If a response-shift does occur, the retrospective-posttest comparison results in a less conservative estimate of the training effect and is considered to provide a more valid estimate of the treatment effect (Howard & Dailey, 1979).

A repeated measurement MANOVA will be used to investigate the occurrence of a response shift (Howard & Dailey, 1979). Depending on the results of these analyses it will be decided if the pretest 2 or the retrospective pretest 2 scores will be used in the further analyses to test the effects of the intervention.

3.4.6.3.2 Analyses of the training effects

The effects of the training will be tested within the framework of the General Linear Model (GLM). Specifically, Multivariate Analyses of Variance (MANOVAs) will be used to analyze the questionnaire data. When using MANOVAs, the following assumptions have to be considered (cf. Bortz, 1993; Tabachnick & Fidell, 1996; Werner, 1997):

The *first* assumption concerns the normal distribution of the error terms (Werner, 1997). Since the MANOVA can be considered robust against violations of this assumption (see Werner, 1997), it will not further be investigated.

The *second* assumption concerns the homogeneity of covariance matrices. It means that the variances and covariances of the repeated measures must be the same for each group and can be tested by calculating Box's *M*. If sample sizes are unequal and Box's *M* test reaches significance at $p < .001$, the robustness of the MANOVA is questionable (Tabachnick & Fidell, 1996). One should, however, also consider the number of dependent variables and the discrepancy between cell sizes. Tabachnick and Fidell (1996) further suggest to use Pillai's criterion if the variance-covariance matrices are not homogenous (cf. Olson, 1979). Pillai's criterion is the most powerful and robust test (Bühl & Zöfel, 2005) and will thus be used in every analysis. Based on this criterion and the large sample size, the calculated MANOVAs will be considered robust against the violation of this assumption.

The *third* assumption concerns the homogeneity of covariances between repeated measurements (sphericity). This assumption can be tested using Mauchly's test of sphericity. The assumption of sphericity is relevant when there are more than two points of measurement (J. Werner, 1997). Whenever this assumption is violated, the second SPSS output table containing the adjusted tests will be reported.

3.4.6.3.3 Analyses of the relations within the model

To test the relations within the model in an exploratory fashion, multiple regression analyses will be used. Multiple regression allows for the examination of patterns of relationships between multiple independent variables (predictors) and a single dependent variable (criterion). Following Cannon-Bowers et al.'s model (1995), multiple regressions will thus be computed to assess the assumed relations within the model (Bortz, 1993; Tabachnick & Fidell, 1996).

Based on this goal, standard regression, i.e., simultaneous entry, will be used to evaluate each independent variable in regard to what it individually adds to the prediction of the criterion

that is different from the predictability provided by the other independent variables. Using this strategy, it is consequently important to consider both, the full correlation of the independent variable with the dependent variable, as well as its unique contribution in the regression equation.
Since the conducted tests of the relations between the model's variables are exploratory in nature, the scales' overall scores will be used. It seems furthermore appropriate to use the overall scale scores to limit the resulting number of independent variables to be entered into the regression analyses since increasing numbers of independent variables also call for an increase in sample size (cf. Tabachnick & Fidell, 1996). An exception is the goal orientation scale. Since its two subscales have previously been established in the respective literature (Button et al., 1996), the subscales themselves will be entered into the regression equations.
To avoid overdetermination, only several of the possible relationships in the model will be tested. The selection of relationships to be tested was based on the respective literature (Cannon-Bowers et al., 1995; Holton, 2005).

To test for mediation effects of the respective variables, the procedure suggested by Baron and Kenny (1986) will be used. Baron and Kenny (1986, p. 1177) describe the set of conditions that must be met to infer support for mediated models. These can be tested with the help of three regression equations:

a) The mediator must be regressed on the independent variable (predictor). In this equation, the independent variable must be shown to affect the mediator.
b) The dependent variable (criterion) must be regressed on the independent variable. Again, the independent variable must be shown to affect the dependent variable.
c) The dependent variable must be regressed on both, the independent variable and the mediator. In this equation, the mediator must be shown to affect the dependent variable. Furthermore, the effect of the independent variable on the dependent variable must be less in this third equation than in the second equation. Complete mediation occurs if the independent variable has no effect when the mediator is controlled. Due to the correlation between the independent variable and the mediator, multicollinearity occurs when the dependent variable is regressed on both, the independent variable and the mediator. This effect leads to a loss of power in the third equation. It is therefore recommended to not only examine the significance of the resulting coefficients, but also their absolute size. Sobel (1982) provided a special significance test for this indirect effect of the independent variable on the criterion via the mediator.

Multiple regression analyses require the following assumptions to be met by the data: Multivariate normal distribution, multivariate homoscedasticity, multivariate linearity, and independence of residual scores. They are, however, largely robust against violations of these assumptions, especially in the light of large sample sizes and the exploratory nature of the analyses (Tabachnick & Fidell, 1996). The analyses will consequently be run even if not all of the assumptions are met by the data.

Singularity and multicollinearity are of further concern in regression analyses since they indicate redundancy in the data and weaken the analysis (Tabachnick & Fidell, 1996). The data will be checked for singularity and multicollinearity and the results reported where indicated.

4 Results of the present study

The following chapters present the results of the evaluation study. They contain the description of the sample characteristics (4.1), the training outcomes (4.2), the individual characteristics variables (4.3), and the relations within the model (4.4).

4.1 Sample characteristics

The following sections describe the sample of this study in more detail with the information contained in the demographics part of the questionnaire.

Using this data, the comparability of the training and control groups was investigated. As has been mentioned above, it was not possible to randomly assign participants to one of the three groups. In such quasi-experimental settings, selection has to be considered as a threat to a study's internal validity (Bortz & Döring, 1995; Cook & Campbell, 1979) because the subjects may differ in influential variables other than the independent variable. To reduce this threat, the comparability of the two groups was investigated along the variables that have been collected in the demographics section of the questionnaire. The comparability was tested with non-parametric tests or, where appropriate, with analyses of variance.

4.1.1 Demographic information

The following sections contain information about the training groups' age and gender (4.1.1.1), and their previous clinical experience (4.1.1.2). For the original questionnaire items (section E) refer to appendix C.

4.1.1.1 Number of participants, age, and gender

Training group 0 consisted of 187 participants (41.3 % males, 58.7 % females) with a mean age of 23.63 years (*SD* = 2.22) and a range of 13 years.

Training group 1 was comprised of 110 participants (40.6% males, 59.4% females). Their age ranged from 22 to 32 years, with a mean age of 24.34 years (*SD* = 2.10).

Training group 2 included 165 participants (44.3% males, 55.7% females) who were on average 23.76 years old (*SD* = 1.91), with a range of 13 years (see Table 31).

Table 31. *Age of training group participants*

Age	*M*	*SD*	min	max
Training group 0 (n = 187)	23.63	2.22	21	34
Training group 1 (n = 110)	24.34	2.10	22	32
Training group 2 (n = 165)	23.76	1.91	21	34

To compare the three groups, an ANOVA was used. The results of the ANOVA revealed a significant effect ($F_{(2,\ 433)} = 3.81$, $p = .023$). Post-hoc tests showed that training group 0 differed significantly from training group 1 ($p = .021$).

The comparability of the three groups concerning the gender distribution (see Table 32) was tested with a chi-square-test. The test revealed no differences between the three groups concerning the gender distribution ($\chi^2(2, N = 438) = .44, p = .801$).

Table 32. *Gender distribution*

	Training group 0		**Training group 1**		**Training group 2**	
Gender	***N***	**%**	***n***	**%**	***n***	**%**
Male	76	41.3	39	40.6	70	44.3
Female	108	58.7	57	59.4	88	55.7
Overall	184	100.0	96	100.0	158	100.0

4.1.1.2 Prior clinical experience

The next item asked the participants to indicate if they possessed prior clinical experience, such as internships or paramedic work (see Table 33).

Table 33. *Prior clinical experience*

	Training group 0		**Training group 1**		**Training group 2**	
Clinical Experience	***n***	**%**	***n***	**%**	***n***	**%**
Yes	59	32.2	31	32.3	72	45.6
No	124	67.8	65	67.7	86	54.4

The groups' comparability concerning their clinical experience was investigated with a chi-square-test. Since all of the categories had an expected frequency larger than five, the chi-square-test's assumption was met. The test indicated a significant difference between the three groups' clinical experience ($\chi^2(2, N = 437) = 7.66, p = .022$).

Table 34. *Prior clinical experience frequencies*

			Training group			Overall
			0	1	2	
Clinical Experience	Yes	Frequency	59	31	72	162
		Expected frequency	67.8	35.6	58.6	162.0
		% of training group	32.2	32.3	45.6	37.1
		Standardized residuals	-1.1	- .8	1.8	
	No	Frequency	124	65	86	275
		Expected frequency	115.2	60.4	99.4	275.0
		% of training group	67.8	67.7	54.4	62.9
		Standardized residuals	.8	.6	-1.3	
Overall		Frequency	183	96	158	437
		Expected frequency	183.0	96.0	158.0	437.0
		% of training group	100.0	100.0	100.0	100.0

The inspection of the expected frequencies and the standardized residuals (see Table 34) showed that training group 0 and training group 1 had fewer participants with experience than would have been expected, whereas training group 2 had more participants with experience than expected and fewer participants without experience than expected. The standardized residuals indicated that both cells of training group 2 contributed to the significant outcome of the test.

This difference in the groups' clinical experience was considered in more detail in the following analyses by investigating the relevancy of the experience.

The following table (Table 35) contains the students' clinical experiences as further specified in an open-ended question, asking the students to indicate the type of clinical experience they have had if they had answered "yes, I have prior clinical experience" in the previous item. In the first step, the participants' entries were counted. Only those experiences were put in the same categories that had been referred to by the same terms, resulting in 27 categories. In the next step, these categories were examined for overlapping types of experiences which were then combined with the help of subject matter experts. This procedure yielded 18 categories (see appendix F). These 18 categories were then classified as HANS-training-relevant experiences and non-relevant experiences by subject-matter experts. It was furthermore differentiated between experiences made as medical students and thus as part of the curriculum, and as experiences gained in a non-medical-school setting. This resulted in the following categorization (Table 35).

Table 35. *Categorization of students' prior clinical experience*

Type of experience	**Frequency**			
	Overall*	**Training group 0**	**Training group 1**	**Training group 2**
Training-relevant clinical experience	**77**	**28**	**16**	**23**
as student-in-practical-training	26	7	3	10
as paramedic	51	21	13	13
Training-non-relevant clinical experience	**95**	**31**	**15**	**49**
as student-in-practical-training	87	24	14	40
as nurse	8	5	0	4
Others	14	2	1	5

Note. For the original German entries refer to appendix F.
*Overall frequency of entries does not necessarily equal number of participants or sum of the different groups' frequencies since participants could indicate different experiences, which were all counted for the overall count, but only the most relevant experiences were counted for the different groups.

Table 35 contains the overall count of entered experiences. For the overall count, all of the entries were counted. For the different groups' counts, only one type of experience, namely the most relevant, was entered and thus counted per participant. If clinical experience was not further specified, it was coded as non-relevant student experience. The differing frequencies are thus attributable to entries for the item "do you have prior clinical experience?" which were in some cases not further specified.

The conducted chi-square test for the distribution of relevant experiences yielded a non-significant result ($\chi^2(2, N = 163) = 5.18, p = .075$), thus indicating, that the type of prior experience did not differ among the training groups.

4.1.2 Conclusion

In conclusion it can be stated that the three groups were unequal in size due to procedural factors in the study. Group sizes will furthermore vary from analysis to analysis because of missing values in the questionnaire data. The consequence will have to be discussed in the respective sections.

Concerning the groups' mean age, the three groups can be considered comparable although the conducted ANOVA yielded a significant effect for age, because the age difference is not even one year and can thus be neglected.

Concerning the gender distribution, the three groups can also be considered comparable.

Concerning the groups' prior clinical experience, they cannot be considered equal as training group 2 contained too many participants with experience and too few participants without experience.

4.2 Training outcomes

The following chapters present the outcomes of the training on the three levels: Training reactions (4.2.1), learning (4.2.2), and behavior (4.2.3).

4.2.1 Level 1: Training reactions

The last part of the posttest questionnaire operationalized the first level of evaluation (Kirkpatrick, 1994) assessing trainees' reactions to the training on a 6-point Likert-scale, with the value 1 indicating strong agreement and the value 6 indicating strong disagreement.
This part of the scale was not factor analyzed since it contained too many conceptually different aspects. Reaction items were instead separately analyzed using ANOVAs[11] following the rationally constructed scale partitions. To conduct multiple comparisons among pairs of means, Scheffé's procedure was used.
The following sections contain the participants' reactions to the simulation and the simulator (4.2.1.1), to the scenarios (4.2.1.2), to the revised curricular sequence (4.2.1.3), their overall satisfaction (4.2.1.4), and their reactions to the instructors (4.2.1.5).

4.2.1.1 Reactions to the full-scale simulation and the simulator

The reactions to the simulation and the simulator were differentiated in the perceived training value (4.2.1.1.1), the appraisal of the simulator's realism (4.2.1.1.2), the simulator's suitability (4.2.1.1.3), and the adjustment to the simulation (4.2.1.1.4). The results are discussed in section 4.2.1.1.5.

4.2.1.1.1 Full-scale simulation training's value for the curriculum

The first aspect to be rated by the trainees was the full-scale simulation training's value for the Heicumed curriculum.

Hypothesis 1
Training groups 1 and 2 rate the full-scale simulation training as more valuable than training group 0.

The group means (see Table 36) show that all of the groups agreed strongly to moderately to both of the items, considering the simulation an important component of the Heicumed curriculum and recommending the use of the HANS simulation in its current form for future semesters. The ratings concerning the participants' recommendation of the simulator training

[11] Since the different groups were not of the same size and the data did not meet the normality and/or variance homogeneity assumptions, non-parametric tests (Kruskal-Wallis) were also conducted to cross-check the ANOVA results. Non-parametric tests yielded equivalent results. The alpha level was not adjusted for the conducted ANOVAs due to their exploratory character.

were overall slightly better than the ratings concerning the importance of the HANS training for the curriculum.

Table 36. *Assessment of the full-scale simulation training's value*

Item	Training group 0 (n = 177)		Training group 1 (n = 98)		Training group 2 (n = 137)	
	M	*SD*	*M*	*SD*	*M*	*SD*
1. The HANS-simulation was an important component of the Heicumed curriculum.	1.55_a	.80	1.76_a	.86	1.77_a	.95
2. I can recommend the use of the HANS-simulation in its current form for the future.	1.31_a	.67	1.54_b	.79	$1.49_{a,b}$	.78

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row not sharing subscripts ($_{a,b}$) are significantly different ($p < .05$) from each other.

The conducted one-way ANOVA revealed for item 1 a significant main effect for group ($F_{(2, 409)} = 3.04$, $p = .049$). The according effect size was small ($eta^2 = .02$). Post-hoc tests for the pair-wise comparison of means showed no significant differences.

The analysis of item 2 showed a significant main effect for group as well ($F_{(2, 409)} = 4.07$, $p = .018$), with the means of training group 0 and training group 1 differing significantly ($p = .034$). Again, the effect size was only small ($eta^2 = .02$).

Hypothesis 1 could thus not be supported by the data since no significant difference could be found in the pair-wise comparisons for item 1 and the direction of the significant pair-wise comparisons was contrary to the hypothesized direction for item 2.

4.2.1.1.2 Effects of the simulator's realism

Two items were used to determine if the simulator's realism was considered motivating as well as helpful for learning the lesson content.

> *Hypothesis 2*
> The three training groups do not differ in their ratings concerning the effects of the simulator's realism.

Again, the means of all groups show that the participants agreed strongly to moderately strong with both of the items (see Table 37).

Table 37. *Assessment of the effects of the simulator's realism*

Item	Training group 0 (n = 177)		Training group 1 (n = 98)		Training group 2 (n = 137/139)	
	M	*SD*	*M*	*SD*	*M*	*SD*
6. The simulator's realism was motivating.	1.69_a	.82	1.99_b	.99	$1.93_{a,b}$	.98
11. The simulator's realism was helpful for learning the content.	1.63_a	.81	1.84_a	.97	1.77_a	.87

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row not sharing subscripts are significantly different ($p < .05$) from each other.

The ANOVA for item 6 resulted in a significant but small group effect ($F_{(2,\ 409)} = 4.25$, $p = .015$, $\text{eta}^2 = .02$). Multiple comparisons among means showed that training group 0 found the simulator's realism to be significantly more motivating than training group 1.

The analysis of item 11 indicated no significant difference between the groups ($F_{(2,\ 409)} = 1.96$, $p = .142$, $\text{eta}^2 = .01$).

Hypothesis 2 could thus be supported for item 11 but not for item 6.

4.2.1.1.3 Simulator's suitability as a training tool

The simulator's suitability as a training tool was assessed with two items. Item 8 referred explicitly to the suitability, whereas item 10 inquired if the participants thought that they could acquire skills with the simulator's help that they could not have acquired using different training tools.

Hypothesis 3
The three training groups do not differ in their ratings concerning the simulator's suitability as a learning tool.

The item means indicate strong to moderately strong agreement with the items for all groups, with training group 1 showing the least agreement, followed by training group 2 and then training group 0 (see Table 38).

Table 38. *Assessment of the simulator's suitability as a training tool*

Item	**Training group 0 (n = 176/177)**		**Training group 1 (n = 98)**		**Training group 2 (n = 137)**	
	M	***SD***	***M***	***SD***	***M***	***SD***
8. I consider the simulator a suitable learning tool for me.	1.65	.85	1.91	1.05	1.85	.87
10. I could acquire skills using the simulator that I could not have acquired using different training tools.	1.67	.83	1.88	1.00	1.82	1.01

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.

The ANOVA for item 8 resulted in a significant difference between groups ($F_{(2,\ 408)} = 3.08$, $p = .047$), without, however, yielding significant differences between pairs of means. The corresponding effect size was small ($eta^2 = .02$).

The analysis of item 10 showed no significant differences between the groups ($F_{(2,\ 409)} = 1.86$, $p = .157$, $eta^2 = .01$).

Hypothesis 3 could thus be supported by the data for item 10. On item 8, the groups did differ in their ratings but the calculated contrasts between the groups did not yield significant effects.

4.2.1.1.4 Adjustment to the simulation

The next set of items referred to the ease with which the participants could adjust to the simulation situation and were thus able to engage in it.

> *Hypothesis 4*
> Training groups 1 and 2 can adapt more easily to the simulations than training group 0.

A look at the means (see Table 39) indicates that participants showed less agreement to this set of items than to the previous ones, although they still agreed moderately strong. The highest agreement for item 7 is portrayed by training group 2, followed by groups 0 and 1, respectively. For item 9, groups 1 and 2 rate the item slightly better than group 0.

Table 39. *Assessment to the adjustment to the simulation*

Item	Training group 0 (n = 177)		Training group 1 (n = 98/97)		Training group 2 (n = 137)	
	M	*SD*	*M*	*SD*	*M*	*SD*
7. I could easily adjust to the simulation.	2.14	1.00	2.21	1.13	2.08	.92
9. I did not feel restrained in the simulation situation.	2.32	1.05	2.26	1.06	2.26	1.05

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.

The conducted ANOVAs neither revealed a significant effect for item 7 ($F_{(2,\ 409)}$ = .50, p = 605, eta^2 = .00) nor for item 9 ($F_{(2,\ 408)}$ = .142, p = .868, eta^2 = .00).

Hypothesis 4 could thus not be supported by the data.

4.2.1.1.5 Conclusion

Overall, the results of the reaction data in regard to the simulation and the simulator were mixed.

Concerning the participants' reactions to the simulation and the simulator the results showed that the interventions did not lead to the expected better ratings of the HANS training's value for the curriculum. All groups considered the simulation an important component of the curriculum. Training group 0 indicated the most agreement with recommending the training for the use in future semesters. Training group 1 agreed significantly less to this item. *Hypothesis 1* could thus not be supported. A possible explanation for this finding is that the first set of changes in the simulation were not as successful as planned. The implementation of the new training methods might not yet have been optimal. This explanation is supported by a slight increase in the agreement of group 2. It is furthermore possible that the hypothesis could not be supported due to a ceiling effect since the ratings of training group 0 were already very high.

No differences between the groups were expected concerning their perceptions of the simulator's realism (*Hypothesis 2*). This hypothesis was partially supported by the data, in that the groups did not differ in their ratings concerning the helpfulness of the simulator's realism for learning the content. Concerning the motivational effect of the simulator's realism, training group 0 again perceived it as more motivating than training group 1.

Concerning the simulator's suitability as a learning tool, the three groups did as expected not differ in considering the simulator a training tool that offered learning opportunities no other tool could. *Hypothesis 3* could thus be supported by this item. Not supported was this hypothesis by the direct ratings of the simulator's suitability as a learning tool. A significant effect for group was found, although the subsequent pairwise comparisons did not yield

significant effects. On the descriptive level, training group 0 again showed the most agreement with the item, thus giving the best ratings for the simulator's suitability as a learning tool. Overall, *Hypothesis 3* could thus only partially be supported.
Due to the video anchors, training groups 1 and 2 were expected to be able to better adjust to the simulation sessions. This hypothesis (*Hypothesis 4*) could not be supported. The effect of the video anchors might not have been strong enough to lead to changes in the participants' ability to adjust to the following simulation situation.

4.2.1.2 Reactions to the scenarios

The following items focused on the participants' reactions to the used simulation scenarios or cases. They investigated the students' appraisal of the scenarios' suitability (4.2.1.2.1) and relevancy (4.2.1.2.2) and are discussed in section 4.2.1.2.3.

4.2.1.2.1 Scenarios' suitability

The first item (item 4) assessed the participants' opinion of the used scenarios' level of difficulty, the second item (item 5) asked participants to rate the scenarios' ability to motivate them to further work on the content of the anesthesiology curriculum.

Hypothesis 5
Training group 2 considers the level of difficulty most adequate, followed by training groups 1 and 0, respectively.

Hypothesis 6
Training group 2 considers the used scenarios most motivating, followed by training groups 1 and 0, respectively.

The descriptive analysis of the means (cf. Table 40) reveals that training group 0 showed the least agreement with item 4, whereas training groups 1 and 2 indicated a slightly higher agreement concerning the adequacy of the scenarios' level of difficulty.

The inspection of item 5's means (cf. Table 40) shows that training group 0 showed the most agreement to the item. Training groups 1 and 2 attained the same means reflecting a slightly lower agreement with the item.

Table 40. *Assessment of the scenarios' suitability*

Item	Training group 0 (n = 176/177)		Training group 1 (n = 98)		Training group 2 (n = 137)	
	M	*SD*	*M*	*SD*	*M*	*SD*
4. The scenarios' level of difficulty was adequate for my level of knowledge.	2.47_{a}	1.12	$2.20_{a,b}$	.93	1.96_{b}	.83
5. The used scenarios motivated me to further work on the content of the anesthesiology curriculum.	1.79_{a}	.83	2.09_{b}	1.01	2.09_{b}	1.04

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row not sharing subscripts are significantly different (p < .05) from each other.

The conducted ANOVAs yielded a significant group effect for item 4 ($F_{(2,\ 408)}$ = 10.12, p = .000). The corresponding effect size was small (eta^2 = .05). Post-hoc comparisons of the group means showed that the rating of training group 0 differed significantly from that of training group 2 (p = .000).

The analysis of item 5 also revealed a significant overall effect for group ($F_{(2,\ 409)}$ = 5.20, p = .006, eta^2 = .03). Post-hoc comparisons resulted in significant differences between training group 0 and training group 1 (p = .038), as well as between training group 0 and training group 2 (p = .020).

Hypothesis 5 was thus partially supported by the data.
Hypothesis 6 was not supported since the found effect was contrary to the anticipated direction.

4.2.1.2.2 Simulation relevancy

One item was used to assess the simulations' perceived relevancy for the students.

> *Hypothesis 7*
> The three groups do not differ in their ratings concerning the scenarios' relevancy.

The means show that the students agreed strongly to moderately strong to the statement that they were able to work on relevant problems in a realistic situation. Group 0 perceived the scenarios most relevant, followed by groups 2 and 1, respectively (Table 41).

Table 41. *Assessment of the simulation's relevancy*

Item	Training group 0 (n = 176)		Training group 1 (n = 98)		Training group 2 (n = 137)	
	M	*SD*	*M*	*SD*	*M*	*SD*
3. The simulation allowed me to work on relevant problems in a realistic situation.	1.60_a	.81	1.89_b	.93	$1.68_{a,b}$	.87

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row not sharing subscripts are significantly different ($p < .05$) from each other.

The conducted ANOVA yielded a significant small effect for group ($F_{(2, 408)} = 3.49$, $p = .032$, $eta^2 = .02$). Post-hoc analyses showed that training group 0 differed significantly from training group 1 ($p = .033$).

Hypothesis 7 could thus not be supported by the data.

4.2.1.2.3 Conclusion

Concerning the students' reactions to the used scenarios the following results were found.
As assumed in *Hypothesis 5*, participants of training group 2 rated the used difficulty level of the scenarios as most adequate for their level of knowledge and significantly more so than training group 0. The found effect size was small (Cohen, 1988). This provides first evidence that the intervention of the preparatory seminar and the change in the curriculum sequence was successful: the students felt better prepared for the simulation sessions. Not supported was the assumption that based on their previous learning and the received support in the scenarios training group 2 would also consider the scenarios most motivating (*Hypothesis 6*). With easier scenarios, the students do not feel as motivated anymore to further work on the anesthesiology content. In regard to the motivational effect of the scenarios, a more adequate difficulty level might have detrimental motivational effects as group 0 indicated the highest motivation due to the scenarios, when at the same time they found them least adequate for their level of knowledge. It thus seems very likely that the experience of scenarios that were too difficult for the participants' level of knowledge resulted in more motivation to further work on the subject.
Not in line with the expectations were the results concerning the perceived relevancy of the scenarios (*Hypothesis 7*). Training group 0 perceived the scenarios more relevant than training group 1. One possible explanation for this finding is that in the unstandardized simulation sessions received by group 0, the instructors put more emphasis on the scenarios' relevancy for the students' exams or clinical practice.

4.2.1.3 Reactions to the revised curricular sequence

Due to the changes in the curricular sequence for training group 2, four additional items were used to assess the students' reactions to the new sequence and the coherence of the contents covered in the different curricular elements.

4.2.1.3.1 Curricular sequence and coherence

The results indicate an overall strong agreement with the curricular sequence (Table 42).

Table 42. *Reactions to the curricular sequence and coherence of the curriculum elements*

Item	Training group 2 (n = 137)		
	M	*SD*	*α = .77*
1. The sequence "Seminar" – "HANS I" – "Emergency Training" – "HANS II" was appropriate.	1.94	1.07	
2. The contents of the above elements were well coordinated.	2.07	1.02	
3. The contents of the "Emergency Training" prepared me well for "HANS II".	2.07	1.20	
4. Due to the experience gained in the "Emergency Training" I profited more from "HANS II" than from "HANS I".	2.02	1.22	

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.

The students considered the sequence of the different curricular elements as appropriate and their contents as well coordinated. The position of the "Emergency Training" before the second simulation session also seemed to be well received by the students, as they indicated that they felt well prepared for the second simulation session and profited more from it than from the first due to their experience gained in the "Emergency Training".

4.2.1.3.2 Conclusion

The specific items used in the questionnaire of training group 2 showed that the new sequence was well received by the participants and prepared them well for the simulation sessions. Although there is room for improvement concerning the coordination of the different elements' content, the ratings were high considering that is was the first implementation of the curriculum changes. The data thus provide preliminary support for the assumption that the students should be prepared before participating in full-scale and complex simulation sessions. The interpretability of the results, however, is limited by the fact that data for the more specific sequencing items were only available for semester 2 and not for the previous semesters which were conducted in a different sequence.

4.2.1.4 Overall satisfaction with the full-scale simulation training

One item was used to assess the overall satisfaction with the simulation training.

Based on the training interventions, the following hypothesis was formulated.

Hypothesis 8
The three groups differ significantly in the degree that the full-scale training fulfilled their expectations, with training group 2 being most satisfied, followed by training groups 1 and 0, respectively.

The different groups' means (see Table 43) showed that they agreed strongly to moderately strong to the statement that the simulator training met their expectations. Training group 2 felt most satisfied followed by training groups 0 and 1, respectively.

Table 43. *Assessment of the overall expectation fulfillment by the full-scale simulation training*

Item	**Training group 0 (n = 177)**		**Training group 1 (n = 97)**		**Training group 2 (n = 137)**	
	M	***SD***	***M***	***SD***	***M***	***SD***
12. Overall, the HANS training fulfilled my expectations.	1.78	.93	1.78	.99	1.74	.79

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.

The conducted ANOVA of item 10 showed no significant differences between the groups ($F_{(2, 408)} = .11$, $p = .897$, $eta^2 = .00$).

Hypothesis 8 could thus not be supported by the data.

It is possible that the implemented changes were not strong enough to lead to the anticipated changes. It is furthermore possible that the operationalization of the satisfaction construct was not adequate with an item inquiring a form of expectation fulfillment.

4.2.1.5 Reactions to the instructors

The following sections first present the results of the students' ratings of the instructors' teaching skills (4.2.1.5.1) and then discuss them (4.2.1.5.2).

4.2.1.5.1 Instructors' teaching skills

As a reaction measure in a broader sense, fourteen items were used to assess students' reactions to the eight instructors, providing information on the instructors' teaching skills. The results are reported according to the rationally constructed scale dimensions. Factor analyses were not conducted due to the small number of items in relation to the number of dimensions. Based on the two training sessions the instructors received, the following hypothesis was formulated.

Hypothesis 9
The three groups differ significantly in their reactions to the instructors, with training group 2 rating them highest, followed by training groups 1 and 0, respectively.

Table 44. *Descriptive statistics: Assessment of the instructors*

Item	Training group 0 (n = 176)		Training group 1 (n = 97)		Training group 2 (n = 137)		
	M	*SD*	*M*	*SD*	*M*	*SD*	*α*
Instructor's structuring skills							**.83**
2. The instructor provided for a structured lesson.	1.83_a	.88	1.73_a	.65	1.42_b	.58	
4. The instructor clarified the lesson's objectives.	2.18_a	1.02	$2.02_{a,b}$	.92	1.80_b	.92	
8. The instructor presented the content clearly and understandably.	1.91_a	1.00	$1.76_{a,b}$	.76	1.61_b	.83	
Instructor's questioning skills							**.64**
1. The instructor's questions were understandable.	1.62_a	.71	1.74_a	.68	1.53_a	.70	
11. The instructor questioned the group's statements.	1.95_a	.90	$1.89_{a,b}$	.90	1.69_b	.81	
Instructor's scaffolding skills							**.68**
3. Through the instructor's feedback I received specific cues about how to improve my performance.	2.17_a	1.20	$2.11_{a,b}$	1.07	1.81_b	.91	
13. The instructor's practical guidance was appropriate.	1.70_a	.87	$1.67_{a,b}$	.69	1.46_b	.66	
Instructor's adaption skills							**.87**
5. It was important to the instructor that the group comprehended the content.	1.94_a	.94	$1.72_{a,b}$	.76	1.53_b	.68	
6. The instructor adapted the lesson to the group's needs.	2.14_a	1.01	$1.96_{a,b}$	.93	1.84_b	.98	
10. The instructor considered the group's knowledge level when providing explanations.	2.01_a	.99	$1.93_{a,b}$	.79	1.72_b	.87	
Instructor's motivating skills							**.75**
7. The instructor motivated me to further contemplate the cases.	2.10_a	.96	2.25_a	1.12	2.10_a	1.12	
9. The instructor created a pleasant learning climate.	$1.65_{a,b}$	.85	1.79_a	.91	1.47_b	.91	
Instructor's personal competence and motivation							**.77**
12. The instructor was competent.	$1.34_{a,b}$	.60	1.45_a	.58	1.26_b	.49	
14. The instructor was motivated.	1.46_a	.71	1.61_a	.76	1.39_a	.66	

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row not sharing subscripts are significantly different ($p < .05$) from each other.

The inspection of the item means (Table 44) reveals that the means differ in the expected direction for the *structuring skills*, *scaffolding skills*, and *adaption skills* scales: the instructors received the best ratings from training group 2 in each item, followed by the ratings of training 1 and 0, respectively. The *questioning skills* scale shows the same pattern for item 11, concerning the instructor's questioning of the group's statements. Item 1 of the *questioning skills* scale as well as the *motivating skills* and *instructor's personal competence and motivation* scales' items show a different pattern, however. On these items, training group 2 still gave the best ratings, but training group 1 gave worse ratings than training group 0.

In spite of the limited number of items per scale, the scales have satisfying to good (Bortz & Döring, 1995) internal consistencies (Cronbach's α).

Table 45. *Test of group differences among assessment of the instructors*

Source	Pillai-Spur/ Sum of squares	*F*	Hypo-thesis *df*	Error *df*	*p*	Eta^2
Instructor's structuring skills	.06	4.45	6	810	*.000*	.03
2. The instructor provided for a structured lesson.	13.65	12.53	2	406	*.000*	.06
4. The instructor clarified the lesson's objectives.	10.74	5.75	2	406	*.003*	.03
8. The instructor presented the content clearly and understandably.	7.38	4.65	2	406	*.010*	.02
Instructor's questioning skills	.03	2.83	4	814	*.024*	.01
1. The instructor's questions were understandable.	2.49	2.56	2	407	.078	.01
11. The instructor questioned the group's statements.	5.73	3.78	2	407	*.024*	.02
Instructor's scaffolding skills	.03	2.81	4	812	*.025*	.01
3. Through the instructor's feedback I received specific cues about how to improve my performance.	10.51	4.51	2	406	*.012*	.02
13. The instructor's practical guidance was appropriate.	4.77	4.06	2	406	*.018*	.02
Instructor's adaption skills	.05	3.34	6	806	*.003*	.02
5. It was important to the instructor that the group comprehended the content.	12.41	9.30	2	404	*.000*	.04
6. The instructor adapted the lesson to the group's needs.	7.22	3.75	2	404	*.024*	.02
10. The instructor considered the group's knowledge level when providing explanations.	6.50	3.88	2	404	*.021*	.02
Instructor's motivating skills	.02	2.31	4	812	.057	.01
7. The instructor motivated me to further contemplate the cases.	1.62	.73	2	406	.483	.00
9. The instructor created a pleasant learning climate.	5.94	3.79	2	406	*.023*	.02
Instructor's personal competence and motivation	.02	1.95	4	810	.10	.01
12. The instructor was competent.	2.16	3.44	2	405	*.033*	.02
14. The instructor was motivated.	2.82	2.84	2	405	.059	.01

The descriptive pattern was inferentially supported by the conducted MANOVAs (see Table 45). The group differences on the *structuring skills*, *questioning skills*, *scaffolding skills* and *adaption skills* scales all reached significance on the multivariate level as well as the univariate level (except for item 1, cf. Table 44).
Post-hoc comparisons (cf. Table 44) showed significant differences ($p < .05$) between all the means of training group 0 and training group 2 for the above mentioned scales (except for item 1). Item 2 yielded an additional significant difference between training group 1 and training group 2.
The effect for group on the *instructor's motivating skills* scale as well as the *instructor's personal competence and motivation* scale did not reach significance on the multivariate level. The univariate analyses revealed that two items on each scale (items 9 and 13) reached significance nevertheless. The respective effect sizes are all small (Cohen, 1988). A look at the post-hoc comparisons shows that the difference between training group 1 and training group 2 reached significance for these items ($p < .05$). This leaves the motivating item (item 7) and the instructor's personal motivation item (item 14) as the only two items not reaching significance on these scales.

Hypothesis 9 could thus be partially supported by the data.

4.2.1.5.2 Conclusion

Overall, the results show that training group 2 gave the highest ratings and *Hypothesis 9* was thus supported by the data from 4 out of 6 aspects. This suggests that the short training sessions for the instructors were successful in changing the instructors' teaching behavior towards better structured lessons, better questioning techniques, more appropriate scaffolding, better adapting the lesson to the group's needs. Although better ratings for the teaching behaviors were found on the descriptive level from training group 0 to training group 1, they did not reach significance, indicating that the instructors did take actions to change their teaching behaviors after receiving the first didactic instruction. Since the majority of significant differences were found between training group 0 and training group 2, it can be assumed, however, that the successful behavior change was only implemented after the repeated intervention, emphasizing the need for repeated exposure to new teaching concepts.
Further support for this explanation is provided by item 12 assessing the instructors' competence. Their competency ratings dropped somewhat for training group 1 but increased again for training group 2, indicating that trying out a new approach towards teaching is in the beginning accompanied by a reduction in competence perceived by the students.
The motivation items' (7 and 14) results concerning the instructors' ability to motivate their students as well as the portrayal of their personal motivation revealed no significant change. This finding implies that the teaching guidelines might not have been specific enough in exemplifying how to better motivate students or that this teaching skill is not as easily

implemented as the others. Concerning the portrayal of personal motivation, it is again possible that it was not made clear enough how to portray personal motivation or that the instructors' motivation did not change due to the new teaching approach. Future interventions might have to further emphasize motivational topics and tap the concept of teacher motivation more directly by also using teachers' self-report data.

The results have to be interpreted with caution, though. First of all, the data did not consistently meet the normality assumption (Bortz, 1993) and the found effect sizes were only small. Secondly, it has to be noted that two sources of variance are confounded in the data set, namely the different students as raters and the different instructors' behavior. The line of argumentation is based on the assumption that the different groups of students' ratings are reliable and valid indicators for the instructors' teaching skills. Further research is thus warranted using trained observers to rate the instructors' skills employing, for example, behavioral markers. Future studies should also invest further resources in questionnaire development, trying to employ more items as well as better tested items.

4.2.2 Level 2: Learning

On the second evaluation level, *learning*, students' self-perceived anesthesia competencies were assessed. Self-ratings were obtained for three points of measurement: Before the seminar, after the seminar, and after the second full-scale simulation session.

The following sections present the preliminary analyses of the scale's psychometric qualities (4.2.2.1), the findings of the self-reported learning scales (4.2.2.2), and the findings of the other-reported learning (4.2.2.3).

4.2.2.1 Preliminary analyses of the anesthesia competency scale's psychometric qualities

The following sections first investigate the anesthesia competency scale's factor structure (4.2.2.1.1) and then examine a possible change in the participants' internal frame of reference (4.2.2.1.2).

4.2.2.1.1 Examination of the anesthesia competency scale's structure

Principal components analyses (PCA) were calculated for the competency items (section A) in order to create a more parsimonious representation of the data.

To investigate the suitability of the questionnaire items and determine the underlying structure of the instrument, several criteria will be examined in a stepwise fashion. In the first step, the number of factors to be extracted will be inspected (4.2.2.1.1.1). The item adequacy will be determined in steps two (4.2.2.1.1.2) and three (4.2.2.1.1.3) by checking item communalities and factor loadings, respectively. After these statistical criteria, the criteria of factor interpretability will be used to determine if the analysis is conceptually useful (4.2.2.1.1.4).

The factors' reliability will be examined in the last step (4.2.2.1.1.5). All of the criteria will be taken into consideration in the last section presenting the final factor solution (4.2.2.1.1.6).

For the analyses the pretest data were used since it can be assumed that the underlying structure of the instrument can be detected more easily before the trainees participate in the simulation intervention.

4.2.2.1.1.1 Investigation of the anesthesia competency scale's factor extraction

An examination of the assumptions showed that the normality assumption was not met by the data. Following Tabachnik and Fidell (1996) this is not considered critical in exploratory settings. The investigation of the correlation matrix revealed correlations ranging from .172 to .791, all reaching significance ($p < .000$). Bartlett's test of sphericity also reached significance ($p < .000$). The correlation matrix can thus be considered factorable (Tabachnick & Fidell, 1996), as was further supported by the Kaiser-Meyer-Olkin measure of sample adequacy (*KMO* = .894) (Janssen & Laatz, 1999).
A principal components analysis with orthogonal varimax rotation with Kaiser normalization was thus conducted investigating the 14 competency items. Due to missing values, only 399 data sets were available for the analysis.
The following table (Table 46) contains the components' eigenvalues before and after the rotation.

Table 46. *Anesthesia competency factors' eigenvalues and explained variance*

	Initial factor matrix			Rotated factor matrix		
Component	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**
1	6.58	46.97	46.97	3.16	22.54	22.54
2	1.40	10.00	56.96	2.99	21.39	43.93
3	1.15	8.21	65.17	2.12	15.16	59.09
4	1.01	7.21	72.39	1.86	13.30	72.39

Note: Principal components method of extraction.

As Table 46 shows, 4 factors should be extracted using the Kaiser-Guttman criterion. A second criterion used to determine the number of factors to be extracted is the screeplot.

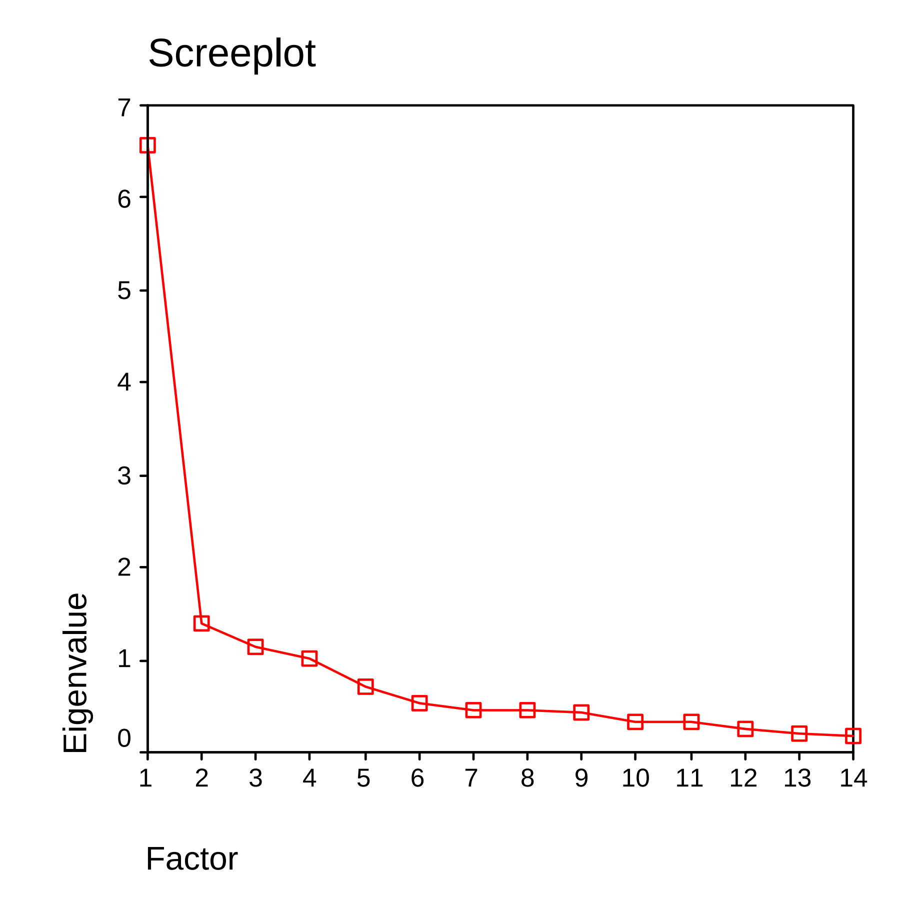

Figure 4. Screeplot of the anesthesia competency factors' eigenvalues

The analysis of the screeplot (see Figure 4) reflects the drop in the initial eigenvalues. The first drop can be observed after factor 1, the second drop after factor 4.

Taken together, the Kaiser-Guttman criterion as well as the screeplot suggest the extraction of four factors. As can be seen in Table 46, these four factors explain 72.39% of the variance. They explain between 22.54% and 13.30% of the variance each.

4.2.2.1.1.2 Investigation of the anesthesia competency scale's item adequacy: communalities

The following table (Table 47) presents the rotated factor solution for the competency items, including the factor loadings a_{ij} of the variables and their respective communalities h^2.

The investigation of the communalities h^2 reveals that the amount of variance of the observed variables accounted for by the factors is rather high, ranging from 63% (item 7) to 83% (item 9) indicating homogeneity among the variables (cf. Tabachnick & Fidell, 1996).

All items can be included in the final factor solution on account of their communality.

Table 47. *Rotated factor loading matrix of the anesthesia competency items*

Item number	Communality h^2	Factor loading a_{ij}			
		1	2	3	4
9	.83	*.86*	.20	.12	.20
8	.79	*.85*	.18		.18
7	.63	*.70*	.19	.21	.25
10	.74	*.69*	.31	.41	
5	.75	.33	*.78*	.15	.11
6	.66	.14	*.77*	.11	.19
2	.74	.18	*.72*	.43	
4	.67	.19	*.67*		.44
3	.69	.24	*.64*	.48	
12	.77	.23	.25	*.80*	
13	.72	.16		*.59*	
11	.77	.55	.35	*.58*	
14	.72	.14	.11	.32	*.77*
1	.65	.30	.19		*.72*

Note. Method of extraction: Principal components analysis. Method of rotation: Varimax with Kaiser-normalization.
Factor loadings with absolute values smaller than .10 are not included in the table (empty cells).

4.2.2.1.1.3 Investigation of the anesthesia competency scale's item adequacy: factor loadings

The items' factor loadings range from .58 (item 11) to .86 (item 9) on the respective factor (see Table 47). Following Comrey and Lee (as cited in Tabachnick & Fidell, 1996) factor loadings above .55 can be considered good (30% of overlapping variance), above .63 (40% over overlapping variance) as very good, and above .71 (50% of overlapping variance) as excellent. In this data set six variables have excellent factor loadings, six have very good factor loadings and two items have good loadings. All of the loadings can thus be considered acceptable.

A different recommendation is offered by Guadagnoli and Velicer (1988) taking into account not only the actual factor loading, but also the number of variables loading on each factor. They suggest to only interpret factors if more than four variables load on them with .60 or higher. Factors 3 and 4 would thus have to viewed as critical. Factor 3 since it has only one variable loading higher than .60 and only contains three variables overall. Factor 4 since only two variables load on it. A factor with only two variables has in general to be viewed as critical. However, taking into account the overall number of items (only 14) and the high correlation of the variables with each other (factor 4) (cf. Tabachnick & Fidell, 1996) further criteria should be considered before a valid decision concerning their inclusion or exclusion can be reached.

The examination of the loadings' structure shows that most variables load highly on only one factor, approaching simple structure. Exceptions, and thus critical, are items 2 (factor 2), 3 (factor 2), 4 (factor 2), 10 (factor 3), 11 (factor 3). They load highly or moderately on more than one factor. As further criteria, the interpretability of these items and internal consistency of the factors will have to be considered to decide whether these items should be excluded.

4.2.2.1.1.4 Investigation of the extracted anesthesia competency factors' interpretability

After the examination of the above statistical considerations, the next important test of the analysis is the interpretability of the resulting factors from a conceptual perspective (Tabachnick & Fidell, 1996). For this purpose, the rotated factor solution with the according questionnaire items is presented in Table 48.

Table 48. *Rotated anesthesia competency factor solution with items of the factors and their respective loadings*

Item number	Questionnaire item How do you rate your competency…	Factor loading
Factor 1		
9	to determine the above listed differential diagnoses.	.86
8	to determine the following preliminary diagnoses (using data stemming from anamnesis, clinical exam, and monitoring): Myocardial infarction, asthma attack, anaphylaxis, pulmonary embolism.	.85
7	to use interdisciplinary knowledge to form preliminary diagnoses.	.70
10	to develop therapy plans for the differential diagnoses listed above.	.69
Factor 2		
5	to determine the exact time for an endotracheal intubation.	.78
6	to conduct an endotracheal intubation.	.77
2	to administer a general anesthesia.	.72
4	to ventilate a patient with a mask.	.67
3	to administer the necessary drugs for a standard anesthesia.	.64
Factor 3		
12	to determine the appropriate drug dosage for the differential diagnoses listed above.	.80
13	to actively gather information about a possible drug contraindication.	.59
11	to administer the appropriate drugs for the differential diagnoses listed above.	.58
Factor 4		
14	to efficiently distribute tasks among team members.	.77
1	to gather relevant information through communicating with the patient.	.72

Note. Method of extraction: Principal components analysis. Method of rotation: Varimax.
Table contains the English translation of the German questionnaire items. For the original German items refer to the questionnaire included in appendix C.

Factor 1: Diagnostic competencies

Factor 1 contains four items concerning the students' competencies in forming diagnoses, thus presenting an interpretable factor.

Factor 2: Anesthetic competencies and airway management

Factor 2 is comprised of items tapping competencies related to conducting general anesthesia and within this process ventilating the patient. The only item that contains an additional aspect is item 3, as it deals with the administration of drugs.

Factor 3: Drug knowledge and administration competencies

Three items dealing with the students' knowledge of drugs and their competencies of administering these drugs load on factor 3. Item 13 additionally taps the concept of communication, investigating the students' ability to actively gather information about possible contraindications.

Factor 4: Communication competencies

Factor 4 is characterized by two items referring to communication skills and can thus be easily interpreted.

4.2.2.1.1.5 Investigation of the anesthesia competency factors' internal consistency

As a test of the factors' reliability, their internal consistency using Cronbach's α was examined (see Table 49).

Table 49. *Cronbach's alphas of the anesthesia competency factors*

	Original item set	
	Number of items	Standardized alpha coefficient
Factor 1	4	.87
Factor 2	5	.85
Factor 3	3	.78
Factor 4	2	.61

Good tests should have alpha coefficients higher than .80 (Bortz & Döring, 1995). This means that factor 4 and possibly also factor 3 have to be considered as critical due to their lower alpha coefficients. It is to be noted, however, that the alpha coefficients will increase with the length of the scale implying that the reliabilities can still be considered acceptable in the light of the number of items.

4.2.2.1.1.6 Conclusion: Final factor solution of the anesthesia competency scale

Taking into consideration all of the aforementioned criteria, the following factor solution can be derived (see Table 50).

Factor 1: Diagnostic competencies

Four items load on factor 1 tapping the competencies of forming diagnoses. Their communalities and loadings are acceptable, their loadings approach simple structure, except for one item (item 10), they can easily be interpreted and are internally consistent. All of the mentioned criteria thus support the retention of the four items.

Factor 2: Anesthetic and airway management competencies

Factor 2 is comprised of five items related to conducting general anesthesia and ventilation. The investigation of the factor structure showed that item 2 also loads on factor 3, and item 4 also loads on factor 4. These items were nevertheless retained in this factor based on conceptual considerations. Item 3 also does not only load on factor 2 (.64) but also moderately high on factor 3 (.48). Since the conceptual investigation also revealed that item 3 contains aspects related to factor 3 (drug administration) it was reassigned to factor 3. The following reliability analysis showed that this reallocation was acceptable as the scale α without item 3 still reaches .82.

Factor 3: Drug knowledge and administration competencies

Factor 3 was initially composed of only three items dealing with the students' knowledge of drugs and their competencies of administering these drugs. Due to the above mentioned changes factor 3 now contains four interpretable items with acceptable communalities and factor loadings. The loading structure, however, is critical. For items 13 and 11 it cannot be considered simple, with item 11 also loading on factor 1 and item 13 also loading on factor 4 (cf. Table 48). Since item 13 additionally taps into the communication concept of factor 4, it will be assigned to the latter factor. Due to conceptual considerations item 11 will be retained in factor 3.

Subsequent reliability analyses showed that the factor's reliability did not considerably change (α = .80).

Factor 4: Communication competencies

Factor 4 was initially characterized by two interpretable items referring to communication skills. Communalities and loadings, as well as loading structure, supported the retention of these items. Due to the considerations described above, item 13 was reassigned to factor 4.

The addition of item 13 resulted in an increase of the scale's internal consistency to α = .69 which can still not be considered "good" (Bortz & Döring, 1995), but is acceptable in the light of the scale length.

Table 50. *Final factor solution of the anesthesia competency items after respective changes*

Item number	Questionnaire item How do you rate your competency…	Factor loading	Cronbach's alpha
Factor 1: Diagnostic competencies			**.87**
9	to determine the above listed differential diagnoses.	.86	
8	to determine the following preliminary diagnoses (using data stemming from anamnesis, clinical exam, and monitoring): Myocardial infarction, asthma attack, anaphylaxis, pulmonary embolism.	.85	
7	to use interdisciplinary knowledge to form preliminary diagnoses.	.70	
10	to develop therapy plans for the differential diagnoses listed above.	.69	
Factor 2: Anesthetic and airway management competencies			**.82**
5	to determine the exact time for an endotracheal intubation.	.78	
6	to conduct an endotracheal intubation.	.77	
2	to administer a general anesthesia.	.72	
4	to ventilate a patient with a mask.	.67	
Factor 3: Drug knowledge and administration competencies			**.80**
12	to determine the appropriate drug dosage for the differential diagnoses listed above.	.80	
11	to administer the appropriate drugs for the differential diagnoses listed above.	.58	
3	to administer the necessary drugs for a standard anesthesia.	.48	
Factor 4: Communication competencies			**.69**
14	to efficiently distribute tasks among team members.	.77	
1	to gather relevant information through communicating with the patient.	.72	
13	to actively gather information about a possible drug contraindication.	.58	

Note. Table contains the English translation of the German questionnaire items. For the original German items refer to the questionnaire included in appendix C.

4.2.2.1.1.7 Intercorrelations among the anesthesia competency factors

The intercorrelations of the resulting questionnaire scales were investigated. Since the data is not normally distributed, Kendall's Tau was calculated in addition to Pearson's correlation coefficient. Kendall's Tau is especially suited for data sets which still contain outliers. The results are presented below (Table 51).

Table 51. *Correlation matrix of the competency factors*

		Factor 1	Factor 2	Factor 3	Factor 4
Factor 1: Diagnostic competencies	Pearson Kendall				
Factor 2: Anesthetic and airway management competencies	Pearson Kendall	*.57**[a]* *.42**[a]*			
Factor 3: Drug knowledge and administration competencies	Pearson Kendall	*.66**[a]* *.51**[a]*	*.69**[b]* *.54**[b]*		
Factor 4: Communication Competencies	Pearson Kendall	*.53**[a]* *.39**[a]*	*.44**[b]* *.32**[b]*	*.51**[b]* *.38**[b]*	

Note: [a] $N = 412$, [b] $N = 422$
** correlations are significant ($p \leq .01$, two-sided).

The examination of the correlation matrix shows that the four competency factors have medium to strong intercorrelations as indicated by Pearson's coefficient (Brosius, 2002). Kendall's coefficients are lower than the Pearson's coefficients, ranging in the weak to medium area (Brosius, 2002). Since the data did not meet the assumptions, Kendall's Tau is considered the more valid result. All of the correlations reach significance.
Overall, the intercorrelations suggest that the four subscales are not fully independent of each other but nevertheless assess different competency aspects. The scores will thus only be combined for each subscale, instead of the complete scale.

4.2.2.1.2 Changes regarding the students' internal frame of reference

The following sections investigate the occurrence of a response-shift as evidence for changes in the students' internal frame of reference (4.2.2.1.2.1) and subsequently discuss the findings (4.2.2.1.2.2).

4.2.2.1.2.1 Investigation of the occurrence of a response-shift

Students' self-perceived anesthesia competency ratings were obtained for three points of measurement. In the pretest questionnaire, data on two points of measurement were collected: The first point of measurement (*pretest 1*(t_1)) was retrospectively assessed and reflects the students' ratings of their competencies before they received any anesthesiology-related input. The second point of measurement (*pretest 2*(t_2)) asked for the self-perceived competencies before attending the simulator training and thus after the seminar. Two different sets of competency ratings were also assessed in the posttest questionnaire: Ratings for pretest 2 were again assessed retrospectively (*retrospective pretest 2*(t_2)) and posttest ratings (*posttest*(t_3)) were assessed reflecting the students' self-perceived competencies after the full-scale simulation training's second session (session b).

The following paragraphs investigate the differences between *pretest 2* and *retrospective pretest 2* of the different training groups to assess the occurrence of a response-shift. A repeated measurements MANOVA was used to test the effects of time of measurement (cf. Howard & Dailey, 1979).

> *Hypothesis 10*
> Participants in all training groups will rate their performance as worse, i.e., will assign higher ratings, in the *retrospective pretest 2* than in *pretest 2*, indicating a response-shift.

Table 52 contains the competency items' descriptive values pertaining to the response-shift.

Table 52. *Descriptive statistics for the competency subscales pretest 2 and retrospective pretest 2*

			Pretest 2		Retrospective Pretest 2	
Factor	**Group**	***n***	***M***	***SD***	***M***	***SD***
Factor 1	Training 0	138	3.84	.83	3.96	.91
	Training 1	83	3.43	.72	3.55	.84
	Training 2	104	3.60	.82	3.27	.89
Factor 2	Training 0	138	4.29	1.00	4.25	.93
	Training 1	83	3.78	.83	3.94	.97
	Training 2	104	3.37	.88	3.15	1.25
Factor 3	Training 0	138	4.79	.97	4.66	.89
	Training 1	83	3.99	.84	4.16	1.02
	Training 2	104	4.00	.85	3.61	1.20
Factor 4	Training 0	138	3.44	.86	3.50	.82
	Training 1	83	3.43	.81	3.40	.83
	Training 2	104	3.27	.81	3.00	.86

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix C.

As the inspection of the four subscales' means in Table 52 and the figures below (Figures 5-8) show, there is no uniform shift in the different groups' means from pretest 2 to the retrospective pretest 2. Training group 0 shows a slight increase in means for factors 1 and 4, but not for factors 2 and 3. Training group 1 shows the expected increase of means for factors 1, 2, and 3, but not for factor 4. The mean ratings of training group 2 decreased from the pretest to the retrospective pretest for all of the four subscales.

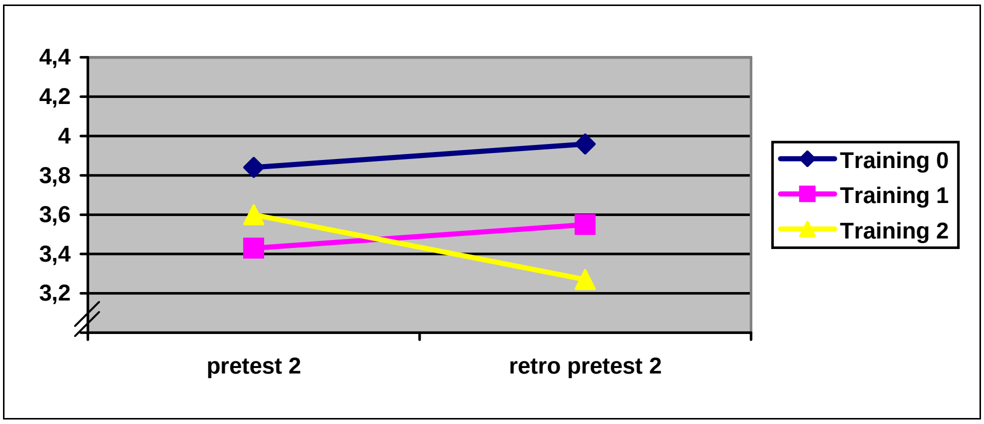

Figure 5. Response-shift anesthesia competency factor 1

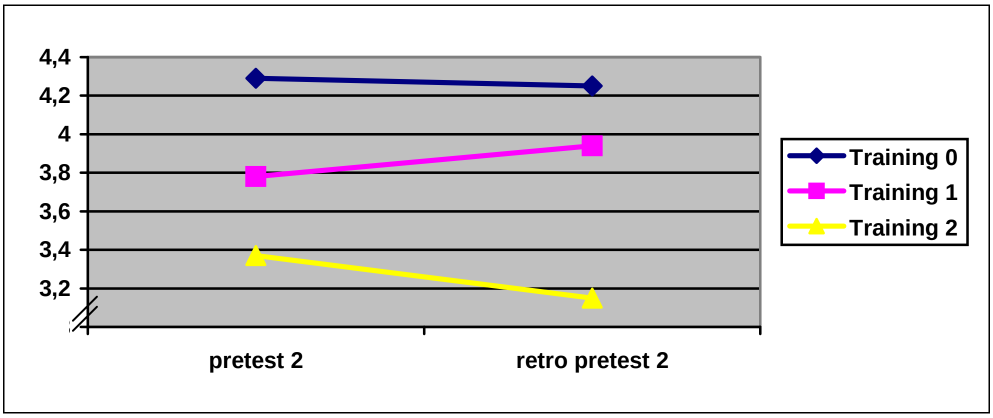

Figure 6. Response-shift competency factor 2

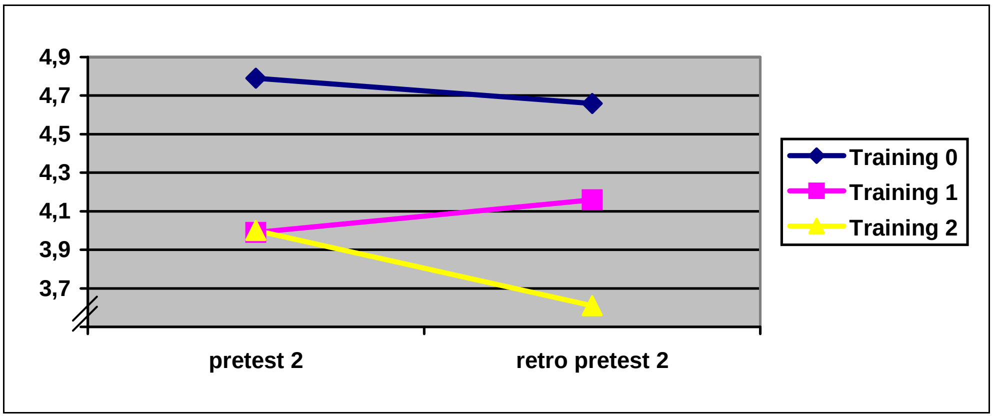

Figure 7. Response-shift competency factor 3

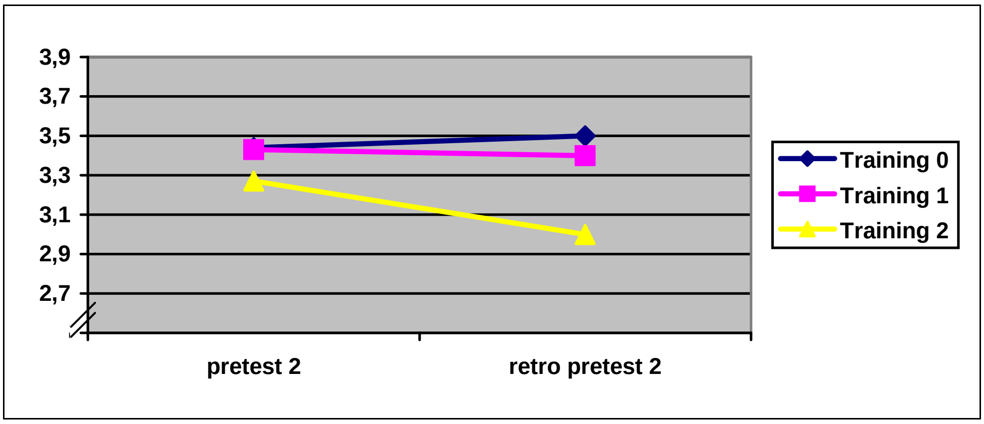

Figure 8. Response-shift competency factor 4

The conducted repeated measurements MANOVA (see Table 53) did not yield a significant effect for time (*p* = .195), although, both, the main effect for group and the interaction effect between time and group, reached significance (*p* < .01). The respective effect sizes indicated a medium effect for group and small effects for time and the interaction effect.

Hypothesis 10 could thus not be supported.

Table 53. *Multivariate analysis of the response-shift*

Source		Pillai-Spur	*F*	Hypothesis *df*	Error *df*	*p*	Eta^2
Between subjects	Group	.31	14.77	8	640	*.000***	.16
Within subjects	Time	.02	1.52	4	319	.195	.02
	Time * Group	.10	4.25	8	640	*.000***	.05

Note. ** *p* < .01; significant effects in italics.

4.2.2.1.2.2 Discussion of changes regarding the students' internal frame of reference: Response-shift

Based on both, the descriptive and inferential results, a response-shift could not be ascertained. It cannot ultimately be determined, however, why this shift could not be detected. The participants might have been able to adequately determine their level of functioning on the given dimensions before the training due to previous experiences in their studies and thus not have shifted their internal standards of measurement. This interpretation fits in the line of research on pretest information (e.g., Howard et al., 1979). It can be assumed that a response-shift did not occur due the participants' level of information about the subject area. Although they had not had explicit practical experience in the rated competencies, they had already received a three-hour anesthesia input, namely, the preparatory seminar. Howard, Dailey, and Gulanick (1979) specifically suggested providing participants with sufficient information to prevent the occurrence of a response-shift and to thus attain results in the retrospective test and the pretest that do not significantly differ from each other. The preparatory seminar could have served this function and thus enabled the participants to adequately judge their initial competencies. The overall rather low grades the participants' gave themselves can be seen as further supporting this interpretation.

The finding of the significant interaction effect and the only consistent finding on the descriptive level of training group 2's improved ratings (better grades) in the retrospective pretest might at first seem to indicate a shift in standards. It seems unlikely however, that these differences truly reflect a shift in the group's internal frame of reference. Due to the changed sequence of the curricular elements, it seems more likely that the participants might not have rated their competencies before the first simulation block (session a), but instead might have rated their performance before the second simulation block (session b), which in their sequence meant after the emergency training. The difference between pretest 2 and retrospective pretest 2 might thus rather reflect the training effect of the emergency training and not a change in the participants' frame of reference. Based on the training literature, a response shift would also be expected in the other direction. In addition, the extended time interval between pre- and retrospective pretest might have been too long for training group 2. Although adequate time intervals for the retrospective tests have not yet been determined (Schwartz & Sprangers, 2000), the response-shift concept assumes that participants can correctly remember their previous level of functioning, an assumption which has also been challenged (Norman, 2003). Unfortunately, the influence of the recalibration response-shift and recall bias could not be teased apart in this design.

In summary, not enough evidence could be found to support the response-shift hypothesis (*Hypothesis 10*). Based on these results, it is assumed that the students used the same internal frame of reference at the two different times of measurement. The original pretest data will consequently be used in the subsequent analyses.

4.2.2.2 Self-reported learning: Training variation effect on students' self-perceived anesthesia competency

The following hypothesis states the general assumption for the effect of the training program.

> *Hypothesis 11*
> The instructionally revised training leads to greater improvements in participants' self-perceived competencies than the original training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively.

Broken down to the two separate interventions, the following hypotheses emerge.

Hypothesis 12
The instructionally revised seminar leads to greater improvements in participants' self-perceived competencies than the original seminar, with participants in training groups 1 and 2 showing greater improvements than participants in training group 0.

Hypothesis 13
The instructionally revised simulation training leads to greater improvements in participants' self-perceived competencies than the original simulation training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively.

4.2.2.2.1 Investigation of the overall training variation effect

4.2.2.2.1.1 Descriptive and inferential analyses

The following table (Table 54) contains the descriptive statistics for the different training groups' self-perceived competency ratings.

Table 54. *Descriptive statistics for the competency factors*

		T_1 Pretest 1 (retrospective) Before intervention		T_2 Pretest 2 After seminar		T_3 Posttest After simulations	
Factor	**Group**	***M***	***SD***	***M***	***SD***	***M***	***SD***
Factor 1	Training 0	3.30	.59	3.08	.69	2.78	.79
	Training 1	3.73	.75	3.43	.72	3.07	.73
	Training 2	3.92	.82	3.61	.81	2.55	.66
Factor 2	Training 0	4.17	1.02	3.21	.82	2.38	.77
	Training 1	4.61	.86	3.78	.83	2.89	.84
	Training 2	4.45	1.03	3.40	.87	2.05	.62
Factor 3	Training 0	4.55	.63	3.66	.82	2.99	.84
	Training 1	4.87	.77	3.98	.84	3.19	.85
	Training 2	4.90	.82	3.99	.86	2.52	.79
Factor 4	Training 0	3.44	.68	3.05	.58	2.45	.69
	Training 1	3.81	.82	3.42	.80	2.73	.69
	Training 2	3.68	.84	3.27	.79	2.23	.56

Note. n_{TG0} = 26, $n_{TG\,1}$ = 84, n_{TG2} = 119.
6-point scale (1 = optimally existent to 6 = mostly not existent).

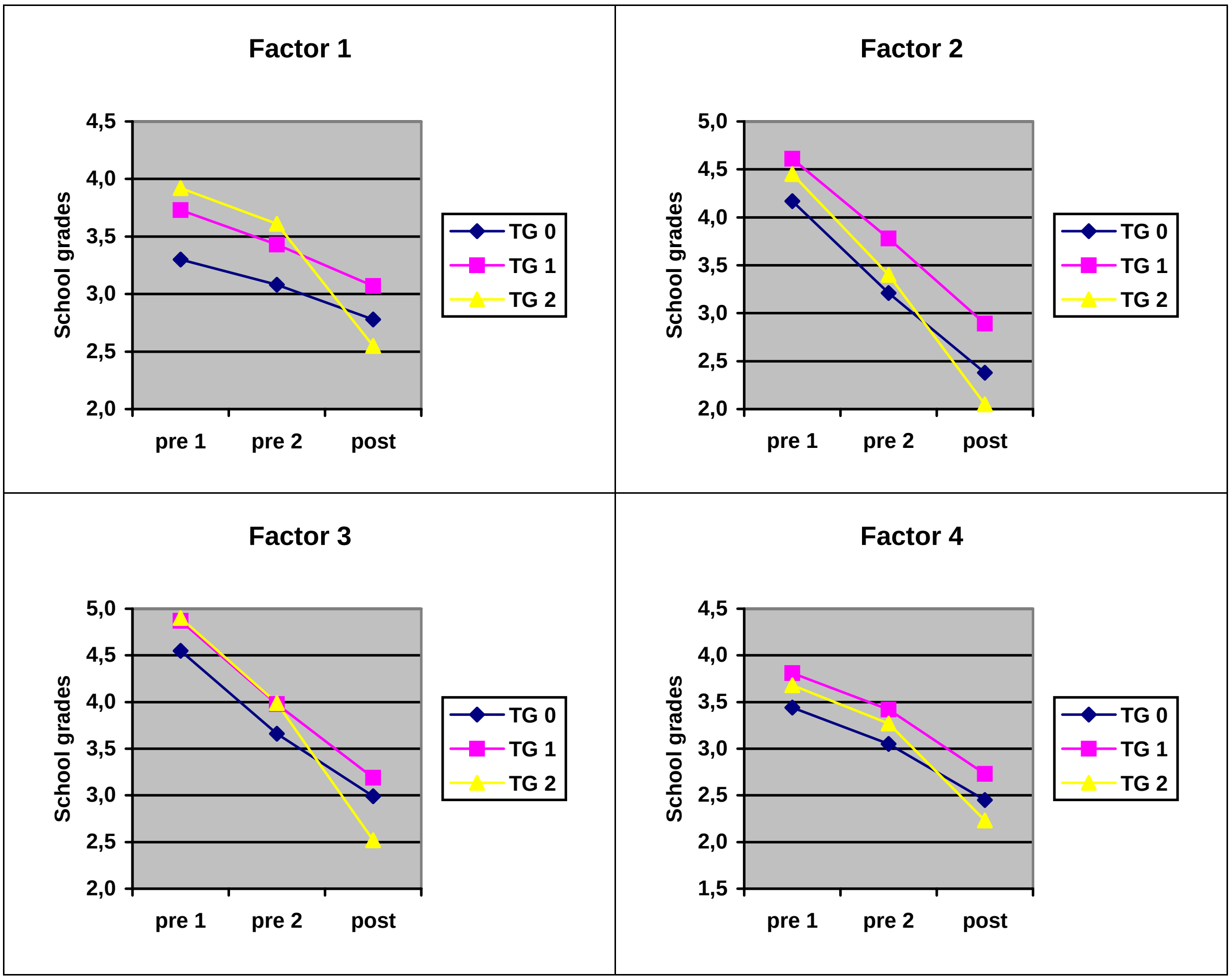

Figure 9. Means of the self-reported anesthesia competencies

The inspection of Table 54 and Figure 9 reveals that each group's ratings on each factor decrease from *pretest 1* to *pretest 2* to the *posttest* thus indicating an increase in the participants' self-perceived competencies. Overall, the self-ascribed competencies are rather low. They vary from 4.90 as the lowest grade in the pretest, to 3.30 as the highest grade in the pretest. In the posttest, they vary from 3.19 to 2.05.

It can further be seen that the training groups' *pretest 1* ratings of their competencies already differ from each other. Concerning the differences between the training groups, it is remarkable that for factors 1 and 3, training group 0 rated their competencies as best in pretest 1 and 2, followed by training groups 1 and 2, respectively. The order changes in the posttest, in which training group 2 rated their competencies as highest followed by groups 0 and 1, respectively. For factors 2 and 4, a slightly different pattern emerges: training group 0 again rated their competencies as best in pretest 1 and 2 but followed by training group 2 and 1, respectively. Again, a different order is visible in the posttest on all four factors in which training group 2 rated their competencies as best followed by training group 0 and 1, respectively.

A repeated measurements multivariate analysis of variance was conducted to analyze the effect of the training intervention. As predicted in the general hypothesis (*Hypothesis 11*), the multivariate analysis (see Table 55) yielded a significant interaction effect between time and group ($p \leq .000$). The size of the respective eta^2 indicates a small effect (cf. Cohen, 1988). The MANOVA furthermore yielded significant main effects ($p \leq .000$) for time and for group, with a large effect size for time and a small effect size for group.

Table 55. *Multivariate analysis of the self-perceived anesthesia competencies*

Source	**Pillai-Spur**	***F***	**Hypothesis *df***	**Error *df***	***p***	**Eta²**
Time	.77	70.81	8	900	*.000***	.39
Group	.15	4.50	8	448	*.000***	.07
Time * Group	.21	6.27	16	1808	*.000***	.05

Note. ** $p < .01$; significant effect in italics.

Subsequently conducted univariate analyses (Table 56) revealed that the significant multivariate effects also reached significance in every univariate analysis ($p \leq .05$), except for the group effect on factor 3 that was only marginally significant. The respective effect sizes were large for the main effect time, small for the main effect group, and small to medium for the interaction effect (cf. Cohen, 1988).

Table 56. *Univariate tests of the self-perceived anesthesia competencies*

Source	Subscale	*F*	Hypo-thesis *df*	Error *df*	*p*	Eta^2
Time	1. Diagnostic competencies	137.70	2	452	*.000***	.38
	2. Anesthetic competencies and airway management	453.80	2	452	*.000***	.67
	3. Drug knowledge and administration competencies	435.99	2	452	*.000***	.66
	4. Communication competencies	223.76	2	452	*.000***	.50
Group	1. Diagnostic competencies	3.12	2	226	*.046**	.03
	2. Anesthetic competencies and airway management	11.40	2	226	*.000***	.09
	3. Drug knowledge and administration competencies	3.02	2	226	.051	.03
	4. Communication competencies	5.38	2	226	*.005***	.05
Time * Group	1. Diagnostic competencies	24.72	4	452	*.000***	.18
	2. Anesthetic competencies and airway management	10.12	4	452	*.000***	.08
	3. Drug knowledge and administration competencies	16.24	4	452	*.000***	.13
	4. Communication competencies	5.79	4	452	*.001***	.05

Note. F ratios were generated from Greenhouse-Geisser's statistic for time and time*group.
** $p \leq .01$; * $p \leq .05$; significant effects in italics.

Scheffé post-hoc analyses for group showed significant differences ($p \leq .05$) between training group 0 and training group 1 for factors 1, 2, and 4, as well as significant differences between training group 1 and 2 for factors 2 and 4. On factor 3, the groups did not differ from each other. Calculated contrasts for time of measurement showed significant differences between each point of measurement on each factor.

4.2.2.2.1.2 Conclusion

The results of this general training effect analysis so far show that the training groups differed in their self-reported competencies, that the self-reported competencies differed over time, and that the effect of time of measurement varies depending on the participants' group. Overall, the self-reported ratings of the students are rather low which seems to adequately reflect their status as beginning students with little or no prior knowledge in anesthesiology.

Due to the main effect for group and the existing differences between the training groups at pretest 1, difference scores per factor will be calculated for the different groups to further investigate the hypothesized gains in learning due to the seminar and due to the simulation sessions.

4.2.2.2.2 Investigation of the training groups' seminar and simulation learning gains

4.2.2.2.2.1 Descriptive and inferential analyses

The first difference score d_1 was calculated between the first point of measurement t_1 and the second point of measurement t_2 ($d_1 = M(t_1) - M(t_2)$), thus indicating the learning gains due to the preparatory seminar. The second difference score d_2 was calculated between the second point of measurement t_2 and the third point of measurement t_3 ($d_2 = M(t_2) - M(t_3)$), thus describing the learning gains due to the simulation sessions. Positive gain scores thus reflect gains in learning based on decreasing raw scores.

The calculated difference scores where then entered as the dependent variable in the new 3 x 2 mixed design. The following table (Table 57) shows the descriptive statistics for the difference scores.

Table 57. *Descriptive statistics for the learning difference scores*

		d_1 ($M(t_1) - M(t_2)$)		d_2 ($M(t_2) - M(t_3)$)	
Factor	**Group**	***M***	***SD***	***M***	***SD***
Factor 1 Diagnostic competencies	Training 0	$.22_a$	.28	$.30_a$	.57
	Training 1	$.30_a$	.39	$.36_a$	.61
	Training 2	$.31_a$	.39	1.07_b	.81
Factor 2 Anesthetic competencies and airway management	Training 0	$.96_a$	.58	$.84_a$	.60
	Training 1	$.83_a$	.60	$.89_a$	.81
	Training 2	1.05_a	.72	1.39_b	.79
Factor 3 Drug knowledge and administration competencies	Training 0	$.89_a$	.47	$.67_a$	.78
	Training 1	$.90_a$	.60	$.79_a$	.70
	Training 2	$.91_a$	.61	1.48_b	.93
Factor 4 Communication competencies	Training 0	$.39_a$	.39	$.60_a$	.52
	Training 1	$.39_a$	.45	$.69_a$	.70
	Training 2	$.40_a$	.43	1.05_b	.83

Note. $n_{TG0} = 26$, $n_{TG1} = 84$, $n_{TG2} = 119$.
For all difference scores higher means indicate more learning.
Cells in the same column within each factor that do not share subscripts are significantly different ($p < .05$) from each other.

The inspection of Table 57 shows that the differences between the groups in their learning gains from pretest 1 to pretest 2 are only marginal, except for factor 2. The differences between the groups are overall bigger for the gain scores between pretest 2 and the posttest. Noticeable are the gain scores for training group 2 that are a lot bigger on every factor than the other groups' scores. On the second gain score, the descriptive pattern is the same on all factors: Training group 0 shows the least improvement, followed by training group 1 and then by training group 2, with the latter showing the highest improvements.

A MANOVA was conducted to inferentially support the results. It yielded significant effects ($p < .01$) for time, group, and the interaction between time and group (see Table 58). The respective effect sizes ranged in the small to medium realm (cf. Cohen, 1988).

Table 58. *Multivariate analysis of the learning difference scores*

Source	**Pillai-Spur**	***F***	**Hypothesis *df***	**Error *df***	***p***	**Eta²**
Time	.18	12.35	4	223	*.000***	.11
Group	.23	7.19	8	448	*.000***	.18
Time * Group	.16	4.75	8	448	*.000***	.08

Note. ** $p < .01$; significant effects in italics.

A look at the univariate results reveals that the significant effect for time is due to factors 1 and 4 (Table 59). The significant multivariate main effect for group is reflected by every factor, except for the only marginally significant effect for factor 3. The significant multivariate interaction effect is due to factors 1 and 3. Effect sizes were again small to medium (cf. Cohen, 1988).

Table 59. *Univariate tests of the learning difference scores*

Source	Subscale	*F*	Hypo-thesis *df*	Error *df*	*p*	Eta²
Time	1. Diagnostic competencies	19.30	1	226	*.000***	.08
	2. Anesthetic competencies and airway management	.83	1	226	.365	.00
	3. Drug knowledge and administration competencies	.79	1	226	.375	.00
	4. Communication competencies	11.56	1	226	*.000***	.11
Group	1. Diagnostic competencies	27.01	1	226	*.000***	.19
	2. Anesthetic competencies and airway management	13.46	1	226	*.000***	.11
	3. Drug knowledge and administration competencies	18.95	1	226	.051	.14
	4. Communication competencies	4.41	1	226	*.002***	.05
Time * Group	1. Diagnostic competencies	18.94	2	226	*.000***	.14
	2. Anesthetic competencies and airway management	2.31	2	226	.102	.02
	3. Drug knowledge and administration competencies	11.05	2	226	*.000***	.09
	4. Communication competencies	4.64	2	226	.011	.04

Note. F ratios were generated from Greenhouse-Geisser's statistic for time and time*group.
** $p \leq .01$; significant effects in italics.

Calculated Scheffé post-hoc tests (see Table 57) for the main effect group showed that the groups did not differ in the seminar difference scores (d_1). For the d_2 difference scores, the post-hoc tests showed that training group 2's gains differed significantly from group 0's and group 1's gains on all factors ($p < .05$). Group 0's and group 1's gains did not differ from each other.

4.2.2.2.2.2 Conclusion

The analysis so far revealed that the groups differ in their learning gains, that the learning gains due to the seminar and the simulation differ from each other, and that the learning gains vary depending on the group and thus the type of intervention.

To further elucidate these results, the seminar's and the simulations' effects will be investigated separately.

4.2.2.2.3 Investigation of the seminar's variation effect

In *Hypothesis 12* it was assumed that the instructionally revised seminar, received by training groups 1 and 2, leads to greater improvements in students' self-reported learning than the seminar in its original version, received by training group 0.

4.2.2.2.3.1 Descriptive and inferential analyses

On a descriptive level (see Table 57), an according trend can be detected in the gain scores for factor 1 *diagnostic competencies* and factor 3 *drug knowledge and administration competencies*: training group 0 shows the least improvements, followed by training groups 1 and 2, respectively. For factor 2 *anesthetic competencies and airway management*, training group 2 shows the predicted largest increase but training group 0's increase is larger than training group 1's. On factor 4 *communication competencies*, training group 0 and 1 reported the same increase in learning and as predicted, training group 2 reported the highest increase. It has to be noted, however, that the differences between the gain scores are overall small.

To inferentially test the hypothesis that the instructionally revised seminar leads to higher self-reported learning, a MANOVA was run with the four learning gain scores d_1 ($M(t_1)$ - $M(t_2)$) as dependent measures and training group as the between-subjects variable.

Table 60. *Multivariate analysis of the seminar difference scores*

Source	**Pillai-Spur**	***F***	**Hypothesis *df***	**Error *df***	***p***	**Eta²**
Group	.04	.99	8	448	.442	.02

As the above table (Table 60) shows, no significant effect for group was detected. This means that the three groups did not differ in their learning gains due to the seminar they received.
The multivariate finding also became evident in the univariate analyses in which none of the tests reached significance (Table 61).

Table 61. *Univariate analyses of the seminar difference scores*

Source	**Sum of Squares**	***F***	**Hypothesis *df***	**Error *df***	***p***	**Eta²**
Factor 1	.17	.58	2	226	.559	.01
Factor 2	2.27	2.57	2	226	.078	.02
Factor 3	.01	.01	2	226	.987	.00
Factor 4	.02	.04	2	226	.960	.00

Hypothesis 12 could thus not be supported.

4.2.2.2.3.2 Conclusion

The comparison of the learning gains due to the seminar indicates that the seminar in its revised version did not lead to a significant increase in self-reported learning, although the predicted differences were, as a tendency, found on the descriptive level regarding the students' *diagnostic competencies* (factor 1) as well as their *drug knowledge and administration competencies* (factor 3). *Hypothesis 12* could thus not be supported.

A look (Table 50) at the items of factor 2 *anesthetic competencies and airway management* and factor 4 *communication competencies* elucidates why the predicted results were not obtained for these factors: the items on factor 2 assessed *anesthetic competencies and airway management* competencies that could not specifically be practiced in a seminar since they involve psychomotor skills (e.g., mask ventilation) as well as cognitive components. In regard to the *communication competencies* (factor 4) it can be reasoned that the cases on paper used in the seminar did not effectively enough address the issue of communication with the patient.

It can thus be concluded that there are only first preliminary indications that the revised problem-oriented seminar based on an integrated approach to learning led to more learning than the traditional seminar in those competencies that can also be practiced on a cognitive level, namely *diagnostic competencies* and *drug knowledge and administration competencies*.

4.2.2.2.4 Investigation of the simulations' variation effect

It was hypothesized that the instructionally revised simulation training leads to greater improvements in participants' self-perceived competencies than the original simulation training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively (*Hypothesis 13*).

4.2.2.2.4.1 Descriptive and inferential analyses

As the inspection of Table 57 shows, the predicted results are found on the descriptive level. On every factor, the learning gains increase from training group 0 to training group 1 to training group 2. The differences between training groups 0 and 1 are comparatively small, whereas the difference to training group 2 is on every factor rather large.

To inferentially test the hypothesis that the instructionally revised simulation sessions lead to higher self-reported learning, a MANOVA was run with the four learning difference scores d_2 ($M(t_2) - M(t_3)$) as dependent measures and training group as the between-subjects variable (Table 62).

Table 62. *Multivariate analysis of the simulation difference scores*

Source	Pillai-Spur	*F*	Hypothesis *df*	Error *df*	*p*	Eta²
Group	.23	7.14	8	448	*.000***	.11

Note. ** $p < .01$; significant effect in italics.

As predicted, a significant multivariate effect was found for group. Its effect size was small. Subsequently calculated univariate analyses showed that the multivariate effect is based on significant effects on each factor (see Table 63). The respective univariate effect sizes vary from small to medium (cf. Cohen, 1988).

Table 63. *Univariate analyses of the simulation difference scores*

Source	Sum of Squares	*F*	Hypothesis *df*	Error *df*	*p*	Eta²
Factor 1	29.62	28.76.	2	226	*.000***	.20
Factor 2	12.84	10.55	2	226	*.000***	.09
Factor 3	29.58	21.29	2	226	*.000***	.16
Factor 4	8.11	7.16	2	226	*.001***	.06

Note. ** $p < .01$; significant effects in italics.

To further investigate which groups differed significantly from each other, pairwise comparisons were calculated (cf. Table 57). On every factor training group 2 differed significantly from training group 0 and significantly from training group 1 ($p < .05$). Training group 1 did not differ significantly from training group 0.

Hypothesis 13 was thus partially supported.

4.2.2.2.4.2 Conclusion

On a descriptive level, the results support *Hypothesis 13* stating that the instructionally revised training of training group 1 led to more learning than the training in its original version (training group 0). This improvement did, however, not reach significance when tested inferentially. The improvement that did reach significance compared to training groups 0 and 1, was the improvement of training group 2 who received the instructionally revised and longer training.

Due to the described design limitations and the resulting confoundation of type of training and amount of training, it can only be tentatively concluded that the instructionally improved training and curriculum changes in regard to the part- and full-scale training together lead to more learning than the original version of the training and the instructionally revised but short version.

4.2.2.2.5 Exploratory investigation of the seminars' and simulation sessions' learning effects

Since the effectiveness of simulation training in comparison to another training method could not be investigated separately, the following comparisons between the learning gains due to the seminar and the learning gains due to the simulation sessions were run.

4.2.2.2.5.1 Descriptive and inferential analyses

On a descriptive level (cf. Table 57) it is noteworthy that the reported learning gains were larger due to the simulation sessions than due to the seminar, except for training group 0 on factor 2 a*nesthetic competencies and airway management* and for training groups 0 and 1 on factor 3 *drug knowledge and administration competencies.*

To determine if there are significant differences between the learning gains d_1 and d_2, a repeated measurements MANOVA was calculated with the difference scores as dependent variables.

As the following table (Table 64) shows, the learning gain d_1 does not differ significantly from learning gain d_2 for training group 0. Training group 1 shows a significant multivariate effect that is due to the significant univariate effect for factor 4 (*communication competencies*), with a medium effect size. Training group 2 also shows a significant multivariate effect. This effect is due to significant univariate effects on all four factors. The respective effect sizes vary from small to large (Cohen, 1988).

Table 64. *Multivariate and univariate results for differences in difference scores d_1 vs. d_2*

Source		**Pillai-Spur/ Sum of squares**	***F***	**Hypo-thesis *df***	**Error *df***	***p***	**Eta²**
Training group 0							
Multivariate		.23	1.63	4	22	.203	.23
Univariate	Factor 1	.08	.36	1	25	.555	.01
	Factor 2	.20	.48	1	25	.495	.02
	Factor 3	.62	1.13	1	25	.298	.04
	Factor 4	.61	3.45	1	25	.075	.12
Training group 1							
Multivariate		.19	4.74	4	80	*.002***	.19
Univariate	Factor 1	.19	.70	1	83	.405	.01
	Factor 2	.15	.25	1	83	.620	.00
	Factor 3	.48	.94	1	83	.335	.01
	Factor 4	3.97	11.11	1	83	*.001***	.12
Training group 2							
Multivariate		.447	23.27	4	115	*.000***	.48
Univariate	Factor 1	34.10	75.30	1	118	*.000***	.39
	Factor 2	5.52	8.92	1	118	*.003***	.07
	Factor 3	19.35	25.43	1	118	*.000***	.18
	Factor 4	24.54	51.35	1	118	*.000***	.30

Note. ** $p < .01$; significant effects in italics.

4.2.2.2.5.2 Conclusion

The results indicate that the simulation sessions led to a greater learning gain for training groups 1 and 2 than the seminar did. The significant difference between the gains for training group 1 was due to the communication factor (factor 4).

Training group 0 that received both interventions in their original, instructionally not revised, version, however, did not show significant differences between the interventions in the reported learning gains. That is, they had the same increase in competencies due to the simulation sessions as they had due to the seminar.

Concerning the advantage of simulation training in comparison to lecture-based seminars, the data of training group 0 allow for the tentative conclusion that students do not learn more merely by participating in simulation sessions. If the simulation sessions are instructionally optimized (training group 1) however, students report higher learning due to the simulation sessions than they do due to an also instructionally revised seminar. Students reported more learning specifically in the area of communication. The significant differences between learning gains for training group 2 are not surprising, and allow only for the tentative interpretation that the students benefit from additional skill-building simulation sessions in between the full-scale simulation sessions. It cannot be concluded, however, that this effect is not merely due to the higher quantity of training.

These results can only be considered as tentative, however, since a true control group was missing and since the different methods, interventions, and instruments were not originally designed for this comparison.

4.2.2.3 Investigation of other-reported learning: Training variation effect in instructor ratings

In addition to the students' self-ratings, instructor ratings of the students' simulator performance during the second full-scale simulation session (session b) were collected on a group level.

Hypothesis 14
Training group 2 receives the best performance ratings, followed by training groups 1 and 0, respectively.

4.2.2.3.1 Descriptive and inferential analyses

The descriptive statistics for the instructor ratings of the students' performance are reported in Table 65.

Table 65. *Instructor ratings of students' simulator performance*

Item	**Training group 0 (n = 22)**		**Training group 1 (n = 20)**		**Training group 2 (n = 20)**	
The group…	***M***	***SD***	***M***	***SD***	***M***	***SD***
1. was well prepared.	2.91_a	1.15	$2.50_{a,b}$	.76	2.20_b	.62
2. participated actively.	2.23_a	.81	2.15_a	.75	1.95_a	.39
3. was motivated to learn.	2.14_a	.83	2.30_a	.87	2.00_a	.46
4. was homogenous concerning their level of proficiency.	2.95_a	1.29	3.00_a	.97	2.85_a	.59
5. could approach the patient.	2.14_a	.83	2.30_a	.47	1.83_a	.38
6. cooperated among themselves.	2.32_a	.78	2.30_a	.80	2.17_a	.38

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix E.
Cells in the same row that do not share subscripts are significantly different ($p < .05$) from each other.

The inspection of Table 65 reveals that for three out of the six items (items 1, 2, and 6) the means developed in the hypothesized direction. On the other three items (3, 4, and 5), training group 1 received the lowest ratings, followed by training group 0. Training group 2 thus received the best ratings on each item.

Conducted ANOVAs (see Table 66) revealed that the descriptive pattern was inferentially supported only for item 1 concerning the groups' preparation level. The found effect was small.

Table 66. *Test of group differences among instructor ratings of students' simulator performance*

Source **The group…**	**Sum of squares**	***F***	**Hypo-thesis *df***	**Error *df***	***p***	**Eta²**
1. was well prepared.	5.33	3.42	59	2	*.039**	.09
2. participated actively.	.85	.91	59	2	.407	.03
3. was motivated to learn.	.90	.81	59	2	.449	.03
4. was homogenous concerning their level of proficiency.	.24	.12	59	2	.889	.00
5. could approach the patient.	2.11	2.82	59	2	.068	.09
6. cooperated among themselves.	.26	.27	59	2	.764	.01

Note. * $p < .05$; significant effect in italics.

Scheffé post-hoc comparisons for item 1 showed that the significant overall effect was due to the difference between training group 0 and training group 2 ($p = .041$).

Hypothesis 14 could thus only be partially supported for one item.

4.2.2.3.2 Conclusion

Hypothesis 14 was partially supported only for item 1, concerning the students' level of preparation as perceived by the instructors. The found effect was small (Cohen, 1988). This finding provides further support for the curriculum changes, indicating that the improved seminar and the part-scale simulation sessions prepared the students better for the second full-scale simulation session. Although the ratings for group 1 were better than for group 0 on the descriptive level, this difference did not reach significance, indicating that the improved seminar's effect was not strong enough.

Although two more items showed the expected pattern descriptively, no significant effects could be detected concerning the students' level of participation, their motivation, the homogeneity of their proficiency level, their ability to approach the patient, or their cooperation.

The results have to be interpreted with caution, however. Due to procedural constraints, each aspect could only be assessed by one item and the instructors filled out the questionnaire under time pressure. Furthermore, the interpretability of the results is limited because the ratings could only be collected on the group level.

4.2.3 Level 3: Behavior

On Kirkpatrick's third level, *behavior*, the students' exam scores were investigated as a more objective indicator of the anesthesia skills the students could display in a transfer setting. The following sections present the investigation of the training variation effect on students' OSCE results (4.2.3.1), with the respective descriptive and inferential findings (4.2.3.1.1), as well as the conclusion (4.2.3.1.2) that can be drawn from the results.

4.2.3.1 Investigation of students' OSCE scores: Training variation effect

Exam scores were taken from the regular OSCE every student had to participate in as part of the degree program. Due to this fact, the exam could not be adapted as part of the study and the exam items consequently could not be matched to the questionnaire items. For this reason, the OSCE items were examined separately and analyzed for group differences using an ANOVA. The OSCE scores were reversely coded so that lower values again imply higher competencies, as they do on a school grade rating scale.

Hypothesis 15
Members of training group 2 receive the best OSCE scores, followed by members of training group 1 and training group 0, respectively.

4.2.3.1.1 Descriptive and inferential analyses

The descriptive statistics (see Table 67) show that training group 1 received the poorest ratings for their performance, followed by training group 0. As predicted, training group 2 received the best performance ratings.

Table 67. *Descriptive statistics for the OSCE scores*

	OSCE scores	
Group	***M***	***SD***
Training 0	4.96_a	2.66
Training 1	5.13_a	2.81
Training 2	3.94_b	2.27

Note. $n_{TG0} = 176$, $n_{TG\,1} = 91$, $n_{TG2} = 140$.
Means that do not share subscripts are significantly different ($p < .05$) from each other.

The conducted ANOVA (see Table 68) yielded a significant effect for group ($p < .01$) with a corresponding small effect (Cohen, 1988).

Table 68. *Analysis of variance of the OSCE scores*

Source	**Sum of squares**	***F***	**Hypothesis *df***	**Error *df***	***p***	**Eta^2**
Group	108.99	8.26	2	400	*.000***	.04

Note. ** $p < .01$; significant effect in italics.

Calculated Scheffé post-hoc tests showed that the scores of training group 2 were significantly different ($p < .05$) from training group 1 and training group 0. The scores of the latter two groups did not significantly differ from each other (cf. Table 67).

Hypothesis 15 was thus partially supported by the data.

4.2.3.1.2 Conclusion

The comparison of the three training groups at the last point of measurement showed that training group 2 received significantly better ratings for their OSCE performance than the other groups. This finding provides first evidence that the instructionally revised training in which part-scale simulation sessions laid the foundations for the full-scale simulation session (received by group 2) was better suited to prepare the participants for the OSCE.

4.3 *Individual characteristics*

The following sections contain the results of the investigation of the variables aimed at assessing the students' individual characteristics. Section 4.3.1 reports the results concerning the students' training expectations and perceptions, section 4.3.2 the results of their instrumentality expectancy, section 4.3.3 the results concerning their situation goal orientation, section 4.3.4 the results concerning their readiness for the training, and section 4.3.5 the results concerning their self-efficacy. For the training expectations and perceptions, the instrumentality expectancy, the situational goal orientation, and the students' readiness, the analyses of the respective scales as well as the investigation of the intervention effects are reported. For self-efficacy, no intervention effect was assumed, consequently, only the scale analysis is reported.

4.3.1 Training expectations, perceptions, and expectation fulfillment

As one of the training motivation variables, training expectations and perceptions were assessed (Cannon-Bowers et al., 1995). Following Cannon-Bowers et al.'s (1995) model, the first sections describe the students' training expectations (4.3.1.1), followed by their training perceptions (4.3.1.2). Section 4.3.1.3 then presents the training expectations and perceptions together in form of the students' expectation fulfillment.

4.3.1.1 Training expectations regarding active participation and knowledge application

In the pretest questionnaire, eight items assessed the students' expectations concerning their active participation in the simulation sessions and the opportunity to apply their theoretical knowledge. The instructionally revised seminar contained elements to influence the students' expectations towards their active participation and the application of their theoretical knowledge, which led to the following hypothesis.

Hypothesis 16
Training groups 1 and 2 have higher expectations concerning their active participation in the simulation sessions and the application of their theoretical knowledge in the simulation sessions than group 0.

The following sections first present the examination of the scale structure (4.3.1.1.1) and then the examination of the training intervention effect to test the above hypothesis (4.3.1.1.2).

4.3.1.1.1 Examination of the training expectation scale's structure

As a first step, a factor analysis was run to investigate the scale structure. The data of all groups were pooled for this analysis. The following sections (4.3.1.1.1.1-4.3.1.1.1.6) present the results of the conducted principal components analysis.

4.3.1.1.1.1 Investigation of the training expectation scale's factor extraction

An examination of the assumptions showed that the normality assumption was not met by the data. As Tabachnik and Fidell (1996) pointed out, this is not considered critical in exploratory settings. The examination of the correlation matrix revealed correlations ranging from .30 to .66, all reaching significance ($p \leq .000$). Bartlett's test of sphericity also reached significance ($p \leq .000$). The correlation matrix can thus be considered factorable (Tabachnick & Fidell, 1996), as was further supported by the Kaiser-Meyer-Olkin measure of sample adequacy (*KMO* = .889) (Janssen & Laatz, 1999).

A principal components analysis with orthogonal varimax rotation with Kaiser normalization was subsequently conducted investigating the eight expectation items. 438 data sets were included in the analysis.

The following table (Table 69) contains the components' eigenvalues.

Table 69. *Training expectations factors' eigenvalues and explained variance*

	Initial factor matrix			Rotated factor matrix		
Component	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**
1	4.40	54.94	54.94	4.40	54.94	54.94
2	.90	11.25	66.19			
3	.70	8.79	74.98			
4	.56	7.02	82.00			
5	.46	5.81	87.81			
6	.37	4.66	92.47			
7	.31	3.91	96.38			
8	.29	3.62	100.00			

Note. Principal components method of extraction.

As Table 69 shows, only one factor should be extracted using the Kaiser-Guttman criterion.

A second criterion used to determine the number of factors to be extracted is the screeplot (Figure 10).

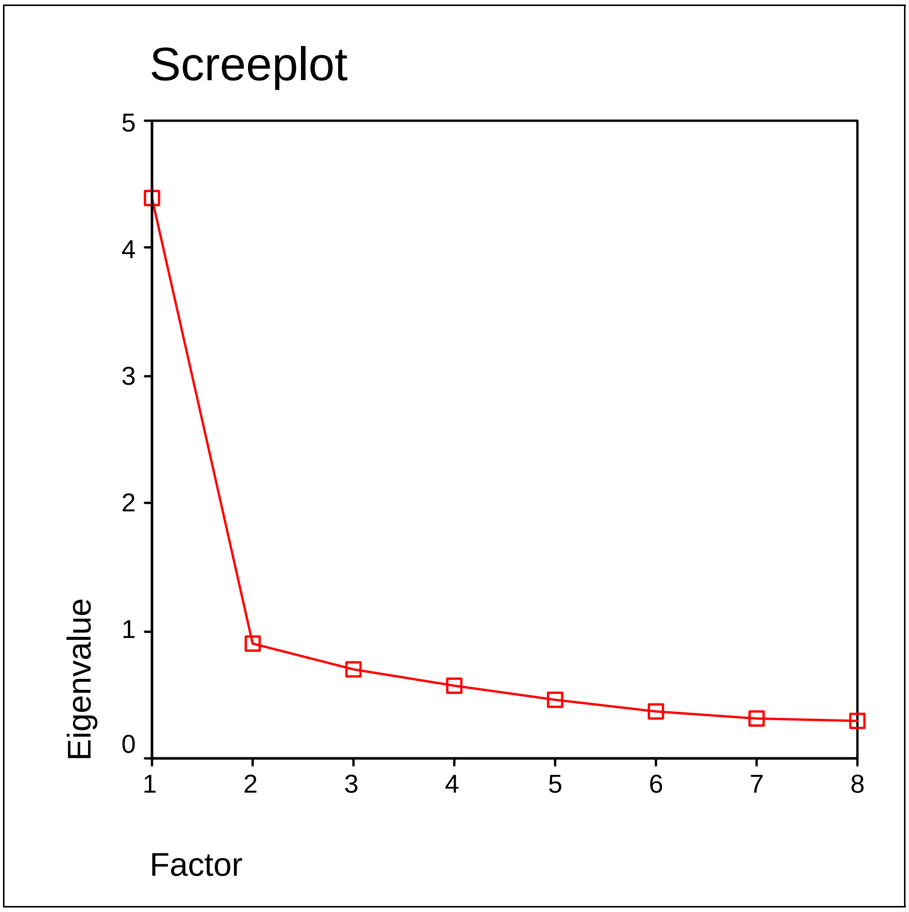

Figure 10. Screeplot of the expectation factors' eigenvalues

The drop in the initial eigenvalues after the first component visible in Table 69, is also reflected in the screeplot (see Figure 10). The slope drastically changes after factor 2.

Taken together, the Kaiser-Guttman criterion as well as the screeplot suggest the extraction of only one factor. As can be seen in Table 69, this factor explains 54.94% of the variance.

4.3.1.1.1.2 Investigation of the training expectation scale's item adequacy: communalities

The following table (Table 70) presents the factor solution for the expectation items, including the variables' factor loadings a_{ij}, and their respective communalities h^2.

The investigation of the communalities h^2 reveals that the amount of variance of the observed variables accounted for by the factors is rather high, ranging from 40% (item 5) to 69% (item 6) and thus indicating homogeneity among the variables (cf. Tabachnick & Fidell, 1996).

All items can thus be included in the final factor solution on account of their communality.

Table 70. *Factor loading matrix of the training expectation items*

Item number	Communalities h^2	Factor loadings a_{ij}
		1
6	.69	.83
8	.63	.79
2	.61	.78
3	.58	.76
7	.53	.73
4	.51	.71
1	.44	.67
5	.40	.63

Note. Method of extraction: Principal components analysis.

4.3.1.1.1.3 Investigation of the training expectation scale's item adequacy: factor loadings

A look at Table 70 shows that the items' factor loadings range from .63 (item 5) to .83 (item 6). Following Comrey and Lee (as cited in Tabachnick & Fidell, 1996) factor loadings above .55 can be considered good (30% of overlapping variance), above .63 (40% over overlapping variance) as very good, and above .71 (50% of overlapping variance) as excellent. In this data set six variables have excellent factor loadings and two have very good factor loadings. All of the loadings can thus be considered acceptable.

Further supporting a one factor solution are the recommendations offered by Guadagnoli and Velicer (1988), taking into account not only the actual factor loading, but also the number of variables loading on each factor. They suggest to only interpret factors if more than four variables load on them with .60 or higher. Since the scale is only comprised of eight items, a single factor solution will be considered acceptable.

4.3.1.1.1.4 Investigation of the extracted training expectation factor's interpretability

After the described statistical considerations, the next important test of the analysis is the interpretability of the resulting factor solution from a conceptual perspective (Tabachnick & Fidell, 1996). For this purpose, the factor solution with the corresponding questionnaire items is presented in Table 71.

Table 71. *Training expectation factor solution with items of the factor and their respective loadings*

Item number	**Questionnaire item** I expect from the HANS simulation sessions to be able to…	**Factor loading**
Factor 1		
6	carry out tasks of an anesthesiologist.	.83
8	try out my acquired competencies in anesthesiology.	.79
2	apply my theoretical knowledge.	.78
3	work independently.	.76
7	realistically monitor the patient.	.73
4	learn from my mistakes.	.71
1	actively participate in the lesson.	.67
5	develop diagnoses in a team.	.63

Note. Method of extraction: Principal components analysis.
Table contains the English translation of the German questionnaire items. For the original German items refer to the questionnaire included in appendix C.

Table 71 shows that the items center around the topic of student activity and application of theoretical knowledge. Items 1, 3, 4, and 5 tap the topic of student activity, whereas items 2, 6, 7, and 8 tap the concept of knowledge application.

4.3.1.1.1.5 Investigation of the training expectations factor's internal consistency

As a test of the factor's reliability, its internal consistency using Cronbach's α was examined (see Table 72).

Table 72. *Cronbach's alpha of the training expectations scale*

	Training expectation scale	
	Number of items	**Standardized alpha coefficient**
Factor 1	8	.88

The resulting alpha coefficient can be considered good (Bortz & Döring, 1995). The analysis also revealed that the alpha coefficient could not be further increased by deleting any of the items.

4.3.1.1.1.6 Conclusion: Final factor solution of the training expectations scale

The results of the principal components analysis did not yield the expected 2-factor solution but a single factor solution, containing both expectation aspects, active participation and knowledge application. Since all of the statistical criteria yielded support for the single factor solution, it will be used for the further analyses.

In contrast to the formula suggested by Tannenbaum (1991) and based on the exploratory nature of this study, the items will, however, not only be merged into a total score, but will also be investigated individually to determine possible differential effects on the items.

4.3.1.1.2 Investigation of the intervention effect on students' training expectations

The following sections investigate the effect of the instructional revision of the seminar on the students' training expectations (*Hypothesis 16*).

4.3.1.1.2.1 Descriptive statistics of the students' training expectations

The following table contains the means and standard deviations for the individual expectation items, as well as for the total scale score.

Table 73. *Descriptive statistics of the students' training expectations*

Item	**Training group 0 (n = 182)**		**Training group 1 (n = 97)**		**Training group 2 (n = 159)**	
I expect from the HANS simulation sessions to be able to…	***M***	***SD***	***M***	***SD***	***M***	***SD***
1. actively participate in the lesson.	1.50	.70	1.57	.71	1.50	.80
2. apply my theoretical knowledge.	1.65	.80	1.56	.75	1.58	.81
3. work independently.	1.98	.89	1.85	.85	1.85	.85
4. learn from my mistakes.	1.29	.53	1.34	.61	1.39	.66
5. develop diagnoses in a team.	1.73	.79	1.70	.83	1.73	.89
6. carry out tasks of an anesthesiologist.	1.55	.64	1.48	.71	1.50	.72
7. realistically monitor the patient.	1.80	.85	1.67	.86	1.62	.73
8. try out my acquired competencies in anesthesiology.	1.76	.87	1.70	.77	1.67	.85
Total score	1.66	.54	1.61	.60	1.61	.59

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix C.

The inspection of Table 73 shows that all of the groups strongly to mostly agreed to the items. A comparison of the different groups' ratings indicates that five items (items 2, 3, 6, 7, and 8) reflect the anticipated pattern: Groups 1 and 2 agree more with the items than group 0. For item 1, training groups 0 and 2 agree more to the item than group 1, for item 5, training group 1 shows the most agreement. For item 4, the pattern is opposite from the expected: group 0 agrees most to the item. This opposite pattern might be due to the item's wording that the students might have perceived as negative.

The total score of the scale again reflects the anticipated pattern, although the differences between the groups are only marginal. Training groups 1 and 2 have the same values and lower values than training group 0, thus showing more agreement with the items than training group 0.

4.3.1.1.2.2 Inferential statistics of the students' training expectations

A MANOVA was used to test for significant differences between the groups. The eight items were entered as dependent variables into a MANOVA, group was entered as the independent variable. Subsequently, the total scale score was entered in a separate ANOVA. The following table (Table 74) shows the results of the multivariate and univariate analyses.

Table 74. *Multivariate and univariate results of the training expectation items*

Source		Pillai-Spur/ Sum of squares	*F*	Hypo-thesis *df*	Error *df*	*p*	Eta^2
Group							
Multivariate		.04	1.04	16	858	.408	.02
Univariate	Item 1	.33	.30	2	435	.742	.00
	Item 2	.72	.57	2	435	.564	.00
	Item 3	1.81	1.21	2	435	.299	.01
	Item 4	.83	1.16	2	435	.314	.01
	Item 5	.05	.04	2	435	.926	.00
	Item 6	.36	.38	2	435	.682	.00
	Item 7	2.73	2.09	2	435	.125	.01
	Item 8	.73	.52	2	435	.598	.00
	Total score[1]	.30	.46	2	438	.632	.00

Note. [1]The total score was not included in the MANOVA. It was entered in a separate ANOVA.

In the MANOVA, neither the multivariate nor any of the univariate analyses reached significance. The ANOVA of the scale's total score also didn't reach significance.

Hypothesis 16 could thus not be supported.

4.3.1.1.2.3 Conclusion

Overall, the results show that the students in all groups expected the opportunity to actively participate in the simulation sessions and to apply their theoretical knowledge. The differences between the groups were only marginal and the found effect sizes only small on the multivariate level. The expected pattern of group 1 and 2 showing more agreement with the items was only found on a descriptive level and only for the scale's total score as well as for five of the items on the item level.

It thus has to be concluded that the instructionally revised seminar did not lead to higher training expectations concerning active participation in the sessions and knowledge application than the seminar in its original version.

It is, however, possible, that the effect was not found due to a ceiling effect since training group 0 already showed very high agreement with the statements.

4.3.1.2 Training perceptions regarding active participation and knowledge application

The posttest questionnaire assessed students' perceptions regarding the opportunity they had received in the simulation sessions to actively participate and to apply their theoretical knowledge. The eight items were parallel to the training expectation items used in the pretest. The instructionally revised simulation sessions aimed at fostering the students' active participation and the application of their theoretical knowledge.

> *Hypothesis 17*
> Training groups 1 and 2 perceived a greater opportunity to actively participate and apply their theoretical knowledge in the simulation sessions than group 0.

The following sections first present the examination of the scale structure (4.3.1.2.1) and then the examination of the training intervention effect to test the above hypothesis (4.3.1.2.2).

4.3.1.2.1 Examination of the training perceptions scale structure

Based on the factor analytic results for the pretest items, the posttest perception scale was also treated as a single factor scale[12].

The following table (Table 75) presents the scale's internal consistency as a test of its reliability.

[12] A conducted principle components analysis of the perception items again yielded a single factor solution.

Table 75. *Cronbach's alpha of the training perceptions scale*

	Original item set	
	Number of items	Standardized alpha coefficient
Factor 1	8	.91

The resulting alpha coefficient can be considered good (Bortz & Döring, 1995). The alpha coefficient could not be further increased by deleting any of the items.

4.3.1.2.2 Investigation of the intervention effect on students' training perceptions

The following sections examine the effect of revision of the simulation sessions on the students' training perceptions regarding the opportunity to actively participate in the training and to apply their theoretical knowledge.

4.3.1.2.2.1 Descriptive statistics of the students' training perceptions

Table 76 displays the means and standard deviations of the students' training perceptions regarding the opportunity to actively participate in the simulation sessions and to apply their theoretical knowledge.

Table 76. *Descriptive statistics of the students' training perceptions*

Item	Training group 0 (n = 177)		Training group 1 (n = 99)		Training group 2 (n = 138)	
In the HANS simulation sessions I was able to…	***M***	***SD***	***M***	***SD***	***M***	***SD***
1. actively participate in the lesson.	1.68_{a}	.77	1.79_{a}	.79	1.60_{a}	.77
2. apply my theoretical knowledge.	2.00_{a}	.88	$1.78_{a,b}$	.80	1.61_{b}	.70
3. work independently.	2.31_{a}	1.03	$2.16_{a,b}$	1.05	1.89_{b}	.87
4. learn from my mistakes.	1.94_{a}	1.03	$1.86_{a,b}$	.97	1.63_{b}	.74
5. develop diagnoses in a team.	2.09_{a}	.95	1.81_{b}	.80	1.56_{b}	.65
6. carry out tasks of an anesthesiologist.	1.77_{a}	.89	1.78_{a}	.78	1.56_{a}	.66
7. realistically monitor the patient.	$2.01_{a,b}$	.98	2.28_{b}	.95	1.92_{a}	.93
8. try out my acquired competencies in anesthesiology.	2.08_{a}	1.03	2.07_{a}	1.02	1.91_{a}	.84
Total score	1.99_{a}	.72	1.94_{a}	.73	1.71_{b}	.58

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row that do not share subscripts are significantly different ($p < .05$) from each other.

Overall, the means indicate that the groups mostly agreed with the statements contained in the items, as indicated by means ranging from 1.56 to 2.31. The further inspection of the table reveals the expected pattern of differences between the group means for five of the eight items. On items 2, 3, 4, 5, and 8, training group 0 shows the least agreement with the items, followed by training groups 1 and 2, respectively. On items 1, 6, and 8, training group 1 shows the least agreement, followed by training group 0 and 2, respectively.
A look at the total scale score also shows the expected pattern: Training group 2 shows the most agreement with the items followed by training groups 1 and 0, respectively.

4.3.1.2.2.2 Inferential statistics of the students' training perceptions

A MANOVA was first used to test for differences between the groups on the individual items. The eight items were entered as dependent variables, group was entered as the independent variable. An ANOVA was also calculated with the total score as the dependent variable. The multivariate und univariate results are presented in the following table (Table 77).

Table 77. *Multivariate and univariate results of the training perception items*

Source		**Pillai-Spur/ Sum of squares**	***F***	**Hypo-thesis *df***	**Error *df***	***p***	**Eta²**
Group							
Multivariate		.17	4.72	16	810	*.000***	.09
Univariate	Item 1	2.00	1.67	2	411	.189	.01
	Item 2	12.07	9.32	2	411	*.000***	.04
	Item 3	13.73	7.11	2	411	*.001***	.03
	Item 4	7.56	4.38	2	411	*.013**	.02
	Item 5	22.16	16.27	2	411	*.000***	.07
	Item 6	4.21	3.35	2	411	*.036**	.02
	Item 7	8.02	4.39	2	411	*.013**	.02
	Item 8	2.80	1.49	2	411	.227	.01
	Total score[1]	6.36	6.86	2	412	*.001***	.03

Note. * $p < .05$, ** $p < .01$; significant effects in italics.
[1]The total score was not included in the MANOVA. It was entered in a separate ANOVA.

The multivariate effect reached significance ($p < .01$), as did the differences on the univariate level ($p < .05$) for six items (items 2, 3, 4, 5, 6, 7). The respective effect sizes are small (J. Cohen, 1988).
To investigate which groups differ from each other, post-hoc Scheffé comparisons were calculated. As Table 76 shows, on four items (items 2, 3, 4, 5) the means of training group 0 differ significantly from the means of training group 2. On item 5, training group 0 also

differed significantly from training group 1 and on item 7, training group 1 differed significantly from training group 2. For three items (items 1, 6, 8) the pairwise comparisons did not yield significant differences between group means.
A conducted ANOVA with the total scale score as the dependent measure also revealed a significant effect for group. Pairwise post-hoc comparisons showed significant differences ($p < .05$) between training group 2 and both of the other groups.

Hypothesis 17 was thus only partially supported.

4.3.1.2.2.3 Conclusion

The results show that there is a significant difference between the three training groups in the extent to which they perceived their opportunities to actively participate and apply their theoretical knowledge in the simulation sessions. The found effect size was small (Cohen, 1988). Training group 2 had better opportunities than training group 0 to apply their theoretical knowledge (item 2), to work independently (item 3), to learn from their mistakes (item 4), and to develop diagnoses as a team (item 5). Training group 1 also perceived a greater opportunity to develop diagnoses as a team (item 5) than training group 0. Concerning patient monitoring (item 7), training group 2 perceived that they had a greater opportunity to realistically monitor the patient than did training group 1. In regards to the perceived opportunity to actively participate in the sessions, to carry out tasks of an anesthesiologist, and to try out their acquired competencies, the groups did not significantly differ from each other.
On the scale level, training group 2 perceived better opportunities to actively participate and apply their knowledge than training group 0 and training group 1.
Based on these data it can be concluded that the new instructional strategy as implemented in the second semester (training group 2) led to improvements on the above mentioned aspects of active participation and knowledge application. In the first semester (training group 1), the instructional revisions did not lead to the expected changes in participant perceptions, except for the aspect of teamwork. It is possible that the instructors had not yet fully adopted the new didactic strategy. It can furthermore not be completely ruled out that the differences in the groups were at least partially influenced by the part-scale simulation training received by training group 2 before they filled out the questionnaire. However, this seems rather unlikely since the instructions specifically referred to the “HANS simulation sessions” and since the questionnaire was filled out immediately after a HANS session.
From a methodological point of view, it is to be noted that the double testing to the items due to the additional test of the total score might have inflated the type I error.

4.3.1.3 Expectation fulfillment regarding active participation and knowledge application

To investigate the extent of the participants' expectation fulfillment, fulfillment scores were calculated based on the formula described in section 3.4.3.1.1.2: The trainees' perception scores (posttest) were subtracted from their pretest training expectation scores. Training expectation fulfillment is thus depicted as a function of pretraining expectations and posttraining perceptions (Cannon-Bowers et al., 1995; Tannenbaum et al., 1991).

Hypothesis 18
Training groups 1 and 2 show higher expectation fulfillment than group 0.

Based on the above mentioned formula, higher expectation fulfillment is thus indicated by smaller absolute values of the fulfillment scores.

4.3.1.3.1 Investigation of the intervention effect on the students' expectation fulfillment

The following sections test the above mentioned hypothesis.

4.3.1.3.1.1 Descriptive statistics of the students' expectation fulfillment

The table below displays the expectation fulfillment scores. Negative mean scores indicate that the trainees had higher expectations concerning the particular training aspect in the pretest than they perceived that aspect as having been realized during the training in the posttest. Positive fulfillment scores indicate that the trainees expected the particular aspect to be realized less in the training than they actually perceived it as having been realized in the training in the posttest.

Table 78. *Descriptive statistics of the expectation fulfillment scores*

Item	**Training group 0 (n = 172)**		**Training group 1 (n = 86)**		**Training group 2 (n = 132)**	
In the HANS simulation sessions I was able to…	***M***	***SD***	***M***	***SD***	***M***	***SD***
1. actively participate in the lesson.	$-.19_a$	.96	$-.23_a$	.90	$-.07_a$	1.01
2. apply my theoretical knowledge.	$-.34_a$	1.06	$-.22_{a,b}$	.87	$-.01_b$	.90
3. work independently.	$-.33_a$	1.17	$-.33_a$	1.13	$-.08_a$	1.08
4. learn from my mistakes.	$-.66_a$	1.09	$-.52_{a,b}$	.99	$-.24_b$	.81
5. develop diagnoses in a team.	$-.37_a$	1.06	$-.12_{a,b}$	1.08	$.13_b$	.83
6. carry out tasks of an anesthesiologist.	$-.22_a$	1.04	$-.30_a$	.84	$-.08_a$	.86
7. realistically monitor the patient.	$-.23_a$	1.18	$-.65_b$	1.14	$-.28_{a,b}$	1.01
8. try out my acquired competencies in anesthesiology.	$-.35_a$	1.29	$-.41_a$	1.06	$-.23_a$	1.07
Total score	$-.33_a$	.80	$-.35_a$	.77	$-.11_b$	.68

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix D.
Means in a row that do not share subscripts are significantly different ($p \leq .05$) from each other.

Table 78 shows that all of the training fulfillment means are negative, except for training group 2's mean for item 5, concerning the development of diagnoses in a team. These negative training fulfillment scores indicate that the trainees' expectations concerning the tapped aspects where slightly higher than their perceptions of the extent to which these aspects had been realized in the training. The scores range from -.66 to .13, indicating overall only small discrepancies between trainees' expectations and their perceptions. A comparison between the different groups' scores shows that training group 2 has the smallest absolute means of the groups except for items 5 and 7. The comparison of training groups 0 and 1 does not yield a uniform pattern: for four items (item 1, 6, 7, 8) training group 0 has smaller absolute means than training group 1, for three items (items 2, 4, 5) training group 0 has bigger absolute means than training group 1, for item 3, the means are the same.
A look at the mean total scale score shows that training group 1 has the greatest negative fulfillment score, followed by training groups 0 and 2, respectively.

4.3.1.3.1.2 Inferential statistics of the students' expectation fulfillment

A multivariate MANOVA was calculated to check for differences between the groups. The results are displayed in the following table (Table 79). The total score was again entered in a separate ANOVA.

Table 79. *Multivariate and univariate results of the expectation fulfillment items*

Source		**Pillai-Spur/ Sum of squares**	***F***	**Hypo-thesis *df***	**Error *df***	***p***	**Eta²**
Group							
Multivariate		.11	2.80	16	762	*.000***	.06
Univariate	Item 1	1.68	.90	2	387	.409	.00
	Item 2	8.44	4.48	2	387	*.012**	.02
	Item 3	5.62	2.19	2	387	.113	.01
	Item 4	12.98	6.71	2	387	*.001***	.03
	Item 5	18.84	9.55	2	387	*.000***	.05
	Item 6	2.91	1.65	2	387	.194	.01
	Item 7	11.00	4.43	2	387	*.012**	.02
	Item 8	1.75	.64	2	387	.529	.00
	Total score[1]	4.65	4.10	2	391	*.017**	.02

Note. * $p < .05$, ** $p < .01$; significant effects in italics.
[1]The total score was not included in the MANOVA. It was entered in a separate ANOVA.

As Table 79 shows, the conducted MANOVA resulted in a significant effect for group. Subsequently calculated univariate analyses showed that the overall effect can be attributed to four significant univariate effects in items 2, 4, 5, and 7.
To determine which groups differed significantly from each other, Scheffé contrasts were used. As can be seen in Table 78, significant differences were found between training group 0 and 2 for items 2 (knowledge application), 4 (learning from mistakes), and 5 (team diagnoses). On item 7 (realistic monitoring), the scores of training group 0 differed significantly from training group 1. The corresponding effect sizes are only small.
The conducted univariate ANOVA of the total scores also revealed a significant difference between the groups with a small effect. Subsequently computed Scheffé contrasts (cf. Table 78) showed that training group 2 differed significantly from training group 0 ($p < .05$) as well as from training group 1 ($p < .05$).

Hypothesis 18 could thus be partially supported.

4.3.1.3.1.3 Conclusion

The overall absolute values show that the discrepancies between the trainees' expectations and their perceptions were only small, not even diverging a whole grade. This means that the students' expectations concerning the investigated training aspects, such as application of theoretical knowledge, were met in the training. The extent of students' expectation fulfillment furthermore depended on the training group they belonged to, as the significant

multivariate effect indicates. Training group 2's expectations concerning the possibility to apply their theoretical knowledge (item 2), to learn from their mistakes (item 4), and to develop diagnoses in a team (item 5) were better met than training group 0's. A small effect was obtained in the analyses (Cohen, 1988). *Hypothesis 18* could thus be partially supported.
Taken together with the previous analyses of the expectations and perceptions (cf. 4.3.1.1.2 and 4.3.1.2), these results provide a first indication that although the revised seminar did not lead to different expectations, the revised simulation sessions better met the students' high expectations, thus leading to better expectation fulfillment scores. It can thus be tentatively concluded that the different curriculum components were better matched in the second semester after instructional changes in the preparatory seminar and the simulation sessions.
Again, it is to be cautiously noted that the double testing increased the possibility of a chance finding.

4.3.2 Instrumentality expectancy

As a motivation variable, students' instrumentality expectancy regarding the simulation sessions were investigated in the pretest.
The following hypothesis was formulated for these variables:

> *Hypothesis 19*
> Training groups 1 and 2 have higher instrumentality expectancies concerning the simulation sessions than group 0.

The following sections first present the examination of the scale structure (4.3.2.1), and then the hypothesis test (4.3.2.2).

4.3.2.1 Examination of the instrumentality expectancy scale's structure

The data of all groups were pooled for the analysis of the scale structure. Since the scale contained only four items, a factor analysis was not conducted.

The calculated reliability analysis yielded a standardized item alpha of .61 (see Table 80).

Table 80. *Cronbach's alpha of the instrumentality expectancy scale*

	Original item set	
	Number of items	**Standardized alpha coefficient**
Scale	4	.61

Although this alpha can still be considered good for a scale of this length (Bortz & Döring, 1995), the item-intercorrelations reflect the heterogeneous content of the items (see Table 81).

Table 81. *Correlation matrix of the instrumentality expectancy items*

Item		**1. Chance in exam**	**2. Applied clinical work**	**3. Better than book**	**4. Better learning opportunity**
1. If I learn a lot during the simulation sessions, I will have a better chance to succeed in the final exam.	Pearson N		*.49*** 437	*.22*** 433	*.22*** 432
2. Through the simulation sessions I will acquire skills that will be useful to me in my applied clinical work.	Pearson N			*.23*** 437	*.33*** 436
3. In respect to the final exam, it is better to participate in the simulation sessions than reading a book on the subject would be.[1]	Pearson N				*.18*** 433
4. I consider the simulation sessions to contain a learning opportunity that I will not have in any other form of instruction.	Pearson N				

Note. ** correlations are significant ($p < .01$, two-sided).
[1] Item 3 was reversely coded. Refer to appendix C for the original item.

As the above table shows, the correlations between the items are rather low, albeit significant ($p < .01$). On the basis of this result and the different aspects the items tap, the scale will be investigated on a multivariate as well as a univariate level.

4.3.2.2 Investigation of the intervention effect on the students' instrumentality expectancy

The following paragraphs investigate *Hypothesis 19* to determine whether the revised seminar had an effect on the students' motivation.

4.3.2.2.1 Descriptive statistics of the students' instrumentality expectancy

The following table (Table 82) presents the means and standard deviations of the instrumentality expectancy scale.

Table 82. *Descriptive statistics of the students' instrumentality expectancy*

Item	Training group 0 (n = 181)		Training group 1 (n = 94)		Training group 2 (n = 154)	
	M	*SD*	*M*	*SD*	*M*	*SD*
1. If I learn a lot during the simulation sessions, I will have a better chance to succeed in the final exam.	1.80	.86	1.91	.76	1.85	.88
2. Through the simulation sessions I will acquire skills that will be useful to me in my applied clinical work.	1.59	.69	1.76	.94	1.66	.75
3. In respect to the final exam, it is better to participate in the simulation sessions than reading a book on the subject would be.[1]	2.28	1.10	2.22	1.04	2.32	1.15
4. I consider the simulation sessions to contain a learning opportunity that I will not have in any other form of instruction.	1.77	.94	1.53	.70	1.78	.88

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
For the original German questionnaire items refer to appendix C.
[1] Item 3 was reversely coded. Refer to appendix C for the original item.

The mean ratings range between 1.53 and 2.32, which indicates that the students overall mostly agreed to the items. The lowest agreement (as indicated by the highest means) was attained for item 3. This might be due to the fact that the item was reversely coded in the original questionnaire (cf. appendix C). A look at the above table (Table 82) further shows that for items 1 and 2, group 0 shows the most agreement, as indicated by the lowest means, followed by training group 2. For items 3 and 4, training group 1 shows the most agreement, followed by training group 0.

4.3.2.2.2 Inferential statistics of the students' instrumentality expectancy

To test *Hypothesis 19*, a multivariate ANOVA was calculated for the items of the *instrumentality expectancy* scale. The individual items were entered as the dependent variable, training group served as the independent variable.

Table 83. *Multivariate analysis of the students' instrumentality expectancy*

Effect	Pillai-Spur	*F*	Hypothesis *df*	Error *df*	*p*	Eta^2
Group	.03	1.62	8	848	.116	.02

As Table 83 shows, the effect for group did not reach significance on the multivariate level, nor did any of the univariate calculations yield a significant effect (see Table 84).

Table 84. *Univariate tests of the students' instrumentality expectancy*

Source	**Item**	**Sum of squares**	***F***	***p***	**Eta^2**
Corrected model	1.	.81	.57	.566	.00
	2.	1.80	1.51	.223	.01
	3.	.53	.22	.805	.00
	4.	4.28	2.81	.061	.01

Note. df = 2, 434.

Hypothesis 19 could thus not be supported by the data.

4.3.2.2.3 Conclusion

Overall, the students perceived the simulation sessions as mostly instrumental for the exam and for their future clinical practice as well as more instrumental than books or other teaching forms. Yet, based on the above presented results, *Hypothesis 19* could not be supported. It can thus not be concluded that the intervention in the seminar led to higher student motivation, as indicated by the students' instrumentality expectancy in regards to the simulation sessions.
Since training group 0 already showed high instrumentality perceptions, it is possible that the seminar intervention was not able to further improve the motivation of the two training groups more because the seminar in its original version had already led to high instrumentality perceptions regarding the simulation sessions.
The non-significant difference could thus be due to a ceiling effect. Future studies should consequently include more reversely coded variables since the reversely coded item 3 showed lower agreement values than the other items.

4.3.3 Situational goal orientation

As another motivation variable, students' situational goal orientation was investigated in the pretest.

Based on the preparatory seminar intervention, the following hypothesis was formulated.

> *Hypothesis 20*
> Training groups 1 and 2 are less performance goal oriented and more learning goal oriented than group 0.

The following sections first cover the examination of the scale structure (4.3.3.1), and then the intervention effect (4.3.3.2).

4.3.3.1 Examination of the situational goal orientation scale's structure

The pretest questionnaire assessed students' goal orientation with 10 items. A principal components analysis with orthogonal varimax rotation with Kaiser normalization was calculated to determine the scale structure. 418 cases were included in the analysis.

4.3.3.1.1 Investigation of the situational goal orientation scale's factor extraction

The following table (Table 85) contains the components' eigenvalues and the amount of variance explained by them.

Table 85. *Situational goal orientation factors' eigenvalues and explained variance*

	Initial factor matrix			Rotated factor matrix		
Component	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**
1	2.65	26.49	26.49	2.58	25.77	25.77
2	1.57	15.72	42.20	1.58	15.82	41.59
3	1.26	12.61	54.81	1.32	13.22	54.81
4	.91	9.10	63.91			
5	.85	8.48	72.38			
6	.69	6.88	79.27			
7	.59	5.93	85.20			
8	.57	5.73	90.93			
9	.50	5.00	95.93			
10	.41	4.07	100.00			

Note. Principal components method of extraction.

As Table 85 shows, the first factor to be extracted has an eigenvalue of 2.58 after rotation, the second factor a smaller eigenvalue of 1.58, followed by the third factor with an eigenvalue of 1.32. The three extracted factors explain 54.81% of variance.
Following the Kaiser-Guttman criterion, the analysis supports the extraction of three factors.

The second criterion to be considered is the screeplot displayed in Figure 11.

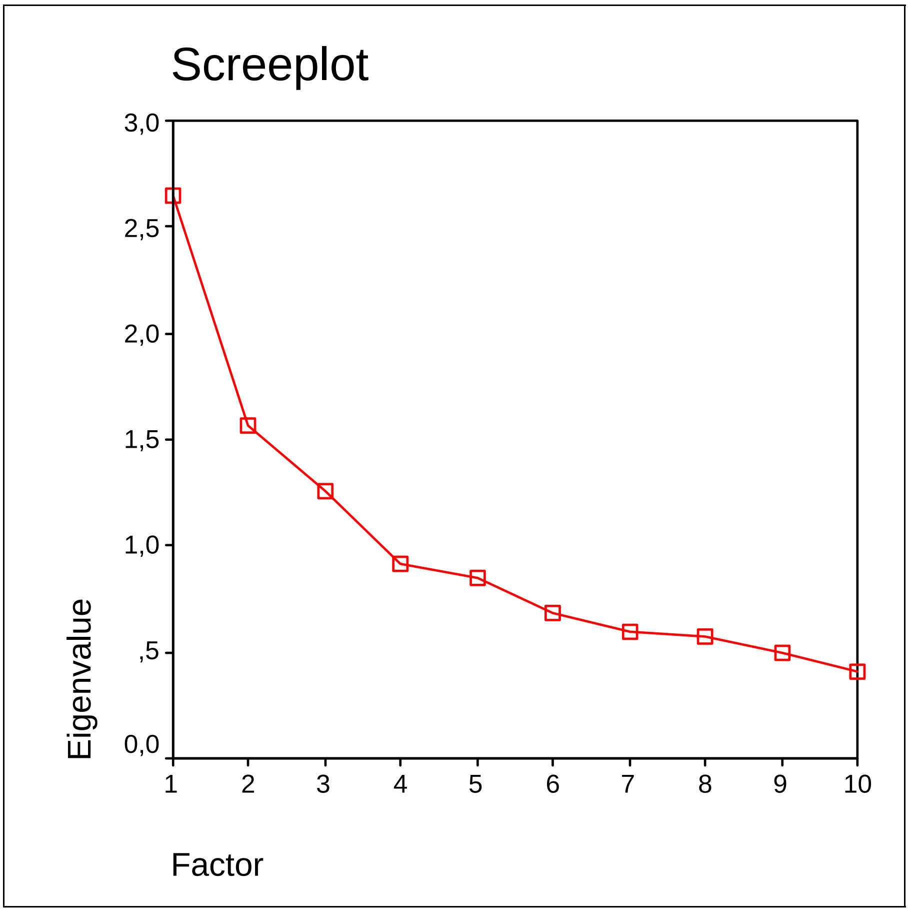

Figure 11. Screeplot of the situational goal orientation factors' eigenvalues

As becomes visible in Figure 11, the slope changes after factor 2 and again after factor 4. The inspection of the screeplot thus suggests the extraction of two or four factors.

4.3.3.1.2 Investigation of the situational goal orientation scale's item adequacy: communalities

Table 86 shows the factor solution for the goal orientation items, including the variables' factor loadings a_{ij} and their respective communalities h^2.

Table 86. *Rotated factor loading matrix of the situational goal orientation items*

Item number	Communalities h^2	Factor loadings a_{ij}		
		1	2	3
6	.64	.80	.05	-.02
4	.64	.79	-.03	.09
2	.54	.72	-.10	.08
9	.50	.68	.21	-.04
7	.33	.54	.00	.20
10	.63	.10	.76	-.23
8	.56	.00	.75	-.03
5	.52	-.02	.60	.40
3	.71	.02	.09	.84
1	.41	.18	-.17	.59

Note. Method of extraction: Principal components analysis. Method of rotation: Varimax with Kaiser-normalization.

The inspection of the items' communalities h^2 shows that the amount of variance accounted for by the factors ranges from 33% (item 7) to 71% (item 3).
Based on their rather low communalities, items 1 and 7 should be further examined to determine if they should be excluded from the final factor solution (cf. Tabachnick & Fidell, 1996).

4.3.3.1.3 Investigation of the situational goal orientation scale's item adequacy: factor loadings

A look at Table 86 further shows that the items' loadings on the respective factors range from .54 (item 7) to .84 (item 3). Based on Comrey and Lee's categorization (as cited in Tabachnick & Fidell, 1996) one item (item 7) thus has a fair loading, two items (items 1 and 5) have good loadings, one has a very good loading (item 9), and six items (items 3, 8, 10, 2, 4, and 6) have excellent loadings. It can further be observed that the items' loadings on the other factors are all low (except item 5). The factor solution thus approaches simple structure (cf. Tabachnick & Fidell, 1996). Consequently, all of the loadings can be considered acceptable.

Following Guadagnoli and Velicer (1988) one should not only consider the factor loadings, but also the number of variables loading on each factor. They suggest to only interpret factors if more than four variables load on them with .60 or higher. This supports factor 1 but poses problems for the extraction of factors 2 and 3.

4.3.3.1.4 Investigation of the extracted situational goal orientation factors' interpretability

The following table (Table 87) presents the scale items and their respective loadings to examine the factor solution from a conceptual perspective (Tabachnick & Fidell, 1996).

Table 87. *Situational goal orientation factor solution with items of the factors and their respective loadings*

Item number	Questionnaire item	Factor loading
Factor 1: Situational performance goal orientation		
6	I hope my performance is up to par.	.80
4	I wonder how my scores will compare to others.	.79
2	I hope I don't make any mistakes.	.72
9	Will I look competent?	.68
7	Will my performance today influence my course grade?	.54
Factor 2: Situational learning goal orientation a)		
10	I'm eager to get started and try to figure things out.	.76
8	I'd like to get a chance to discuss my mistakes with others.	.75
5	I hope there are some tough parts that I really get to work at.	.60
Factor 3: Situational learning goal orientation b)		
3	I hope we won't cover material I already know.	.84
1	Am I going to learn something?	.59

Note. Method of extraction: Principal components analysis.
Table contains the translation of the German questionnaire items. For the German items used in the questionnaire refer to appendix C.

As can be seen in Table 87, the items loading on factor 1 are the *situational performance goal orientation* items, as they were contained in the original scale (cf. Button et al., 1996). The items of the original *situational learning goal orientation* sub-scale load on factors 2 and 3. Factors 2 and 3 are thus also interpretable, although the split of the items into two factors was not anticipated based on previous research results (cf. Button et al., 1996).

4.3.3.1.5 Investigation of the situational goal orientation factors' internal consistency

To test the factors' reliability, their internal consistency was investigated using Cronbach's α (see Table 88).

Table 88. *Cronbach's alpha of the situational goal orientation scale*

	Goal orientation scale		
	Number of items	**Number of cases**	**Standardized alpha coefficient**
Factor 1	5	422	.76
Factor 2	3	432	.52
Factor 3	2	438	.34

The obtained alpha coefficient for factor 1 can be considered good (Bortz & Döring, 1995). With .52, the alpha coefficient for factor 2 is rather low, but has to be put in the perspective of the factor's small item number (Bortz & Döring, 1995). Factor 3's reliability with .34 does not seem acceptable, in spite of the small number of only two items.

4.3.3.1.6 Conclusion: Final factor solution of the situational goal orientation scale

Taken together, the previous analyses suggest the extraction of two factors. Based on the eigenvalues, three factors could have been extracted. But even this three-factor solution did not explain a large amount of variance with only 54.81%. The additional screeplot inspection indicated a slope change after factor 2 and 4. Regarding the items' communalities, items 1 and 7 were questionable, all other communalities satisfactory. Although all of the factor loadings were acceptable in size, the small number of items on factors 2 and 3 poses a problem. Also considering the conceptual interpretability and previous research findings (cf. Button et al., 1996) of the factors, a two-factor solution seems most appropriate, thus excluding items 3 and 1 from further analyses.

Factor 1 consequently contains the five *situational performance goal orientation* items, factor 2 three of the original *situational learning goal orientation* items. The reliabilities of the two factors can be considered acceptable in the light of the small number of items. It has to be noted, however, that this factor solution only explains a rather small amount of variance (41.59%).

4.3.3.2 Examination of the intervention effect on the students' situation goal orientation

In order to test *Hypothesis 20*, a MANOVA was calculated to determine if the revised seminar had an effect on students' situational goal orientation.

4.3.3.2.1 Descriptive statistics of the students' situational goal orientation

The following table (Table 89) presents the descriptive statistics for the students' situational learning and performance goal orientation.

Table 89. *Descriptive statistics of the students' situational goal orientation*

Factor	Training group 0 (n = 184)		Training group 1 (n = 97)		Training group 2 (n = 158)	
	M	***SD***	***M***	***SD***	***M***	***SD***
1 Performance goal orientation	3.77_a	1.02	3.73_a	.97	3.72_a	.98
2 Learning goal orientation	2.68_a	.89	2.97_b	.78	3.00_b	.90

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
Means in a row that do not share subscripts are significantly different ($p \leq .05$) from each other.

The inspection of Table 89 shows that for the performance goal orientation, the differences are very small: Training group 0 agreed least with the performance goal items, followed by training group 1 and 2, respectively. Overall the groups "rather disagreed" with the items as indicated by the means around 3.7.

A look at the learning goal orientation factor reveals that training group 0 shows the highest agreement, followed by training groups 1 and 2, respectively. As indicated by the means between 2.7 and 3.0, the groups "rather agreed" with the learning goal items. The differences between the groups are still rather small but nevertheless larger than on the performance goal factor.

Overall, the means on the learning goal orientation factor are lower than on the performance goal factor. It can thus be stated that all of the groups were more learning goal oriented than performance goal oriented.

4.3.3.2.2 Inferential statistics of the students' situational goal orientation

A MANOVA was calculated with the two factors as dependent and training group as the independent variable. The following table (Table 90) contains the results.

Table 90. *Multivariate analysis of the students' situational goal orientation*

Effect	Pillai-Spur	*F*	Hypothesis *df*	Error *df*	*p*	Eta^2
Group	.03	3.51	4	872	*.007***	.02

Note. ** $p < .01$; significant effect in italics.

The effect for group is significant ($p < .01$) on the multivariate level. The univariate tests (see Table 91) showed that the significant multivariate effect is due to a significant effect on the learning goal orientation factor. The respective effect size was small (Cohen, 1988).

Table 91. *Univariate tests of the students' situational goal orientation*

Source	**Factor**	**Sum of squares**	***F***	***p***	**Eta²**
Group	1. Performance goal orientation	.22	.11	.896	.00
	2. Learning goal orientation	10.43	6.89	*.001***	.03

Note. $df = 2, 436$. ** $p < .01$; significant effect in italics.

Subsequently conducted post-hoc comparisons (Scheffé contrasts) revealed significant differences ($p < .05$) on the learning goal orientation factor between training group 0 and both of the other groups but not in the expected direction (cf. Table 89). Training groups 1 and 2 did not significantly differ in their learning goal orientation.

Hypothesis 20 could thus not be supported.

4.3.3.2.3 Conclusion

On the descriptive level, training group 0 showed the most favorable pattern with the least performance goal orientation of all of the groups and at the same time the highest learning goal orientation. Both of the other training groups were more performance goal oriented and less learning goal oriented than training group 0. The conducted tests showed that the groups did not significantly differ on their performance goal orientation. They did, however, differ in their learning goal orientation. Contrary to *Hypothesis 20* training group 0 was the most learning goal oriented. It has to be concluded that the revised seminar was thus not able to focus the students' situational goal orientation towards less performance goal orientation and more learning goal orientation. It is possible, that the chosen intervention was not strong enough to do so. The students were informed about the goals and benefits of the simulation sessions at the beginning of the seminar. A stronger intervention should induce more student involvement with the topic of goal orientation, for example, through extended discussions or reflection on the topic. Since an intervention of this kind was not possible within the timeframe of the current study, future studies should investigate the latter kind of intervention. It is furthermore possible that the questionnaire was not able to separately assess the situational components of the students' goal orientation, but instead merely or mostly assessed the respective dispositional components, thus reflecting group differences beyond the intervention.

Questions posed during the data collection process let the assumption arise that the students perceived the goal orientation questions as "very strange". It can thus not be ruled out that the students were reluctant to answer the goal orientation questions, rendering the results invalid. From a statistical viewpoint, the presented conclusion can again only be considered tentative due to the unequal cell sizes.

4.3.4 Student readiness

To investigate the students' readiness for the HANS training, five items were administered in the posttest.

Based on the preparatory seminar intervention, the following hypothesis was formulated.

Hypothesis 21
Training groups 1 and 2 were readier for the full-scale simulation training than group 0.

The following sections present the results of the scale structure's examination (4.3.4.1), followed by the results of the training effect's investigation (4.3.4.2).

4.3.4.1 Examination of the readiness scale's structure

To examine the scale structure of the posttest readiness items, a principal components analysis with orthogonal varimax rotation with Kaiser normalization was calculated. 395 cases were available for the analysis.

4.3.4.1.1 Investigation of the readiness scale's factor extraction

The following table (Table 92) contains the components' eigenvalues and the amount of variance explained by them.

Table 92. *Readiness factors' eigenvalues and explained variance*

	Initial factor matrix			Rotated factor matrix		
Component	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**
1	2.67	53.37	53.37	2.41	48.11	48.11
2	1.03	20.59	73.95	1.29	25.84	73.95
3	.74	14.76	88.72			
4	.36	7.22	95.94			
5	.20	4.06	100.00			

Note. Principal components method of extraction.

The first factor to be extracted has an eigenvalue of 2.41 after rotation (see Table 92), the second factor a smaller eigenvalue of 1.29. The two extracted factors explain 73.95% of variance.
Following the Kaiser-Guttman criterion, two factors should thus be extracted.

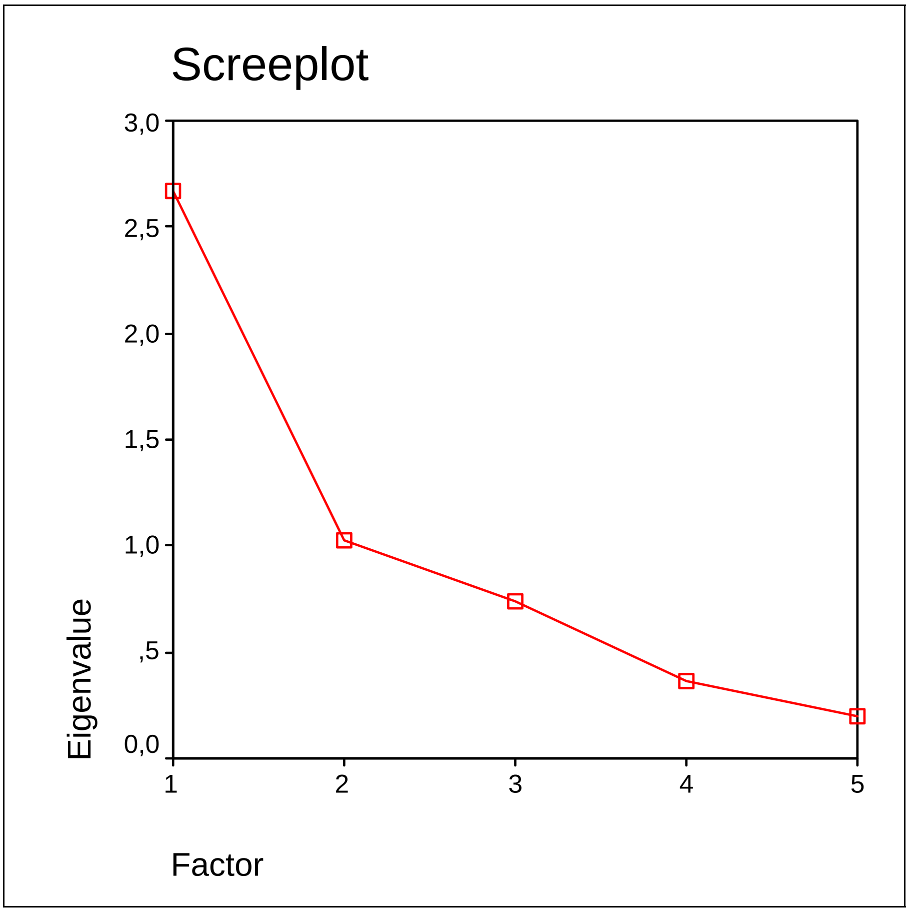

Figure 12. Screeplot of the readiness factors' eigenvalues

The visual inspection of the screeplot (Figure 12) also suggests the extraction of two factors since a slope change can be detected after factor 2.

4.3.4.1.2 Investigation of the readiness scale's item adequacy: communalities

The table below contains the factor solution for the readiness items, including the variables' factor loadings a_{ij} and their respective communalities h^2.

Table 93. *Rotated factor loading matrix of the readiness items*

Item number	Communalities h^2	Factor loadings a_{ij}	
		1	2
2	.86	.92	.14
3	.81	.89	.13
1	.73	.83	.18
5	.76	.02	.87
4	.54	.27	.68

Note. Method of extraction: Principal components analysis. Method of rotation: Varimax with Kaiser-normalization.

A look at the items' communalities h^2 in Table 93 shows that the observed variables' amount of variance accounted for by the factors ranges from 54% (item 4) to 86% (item 2).
Based on their communalities all items can thus be included in the final factor solution (cf. Tabachnick & Fidell, 1996).

4.3.4.1.3 Investigation of the readiness scale's item adequacy: factor loadings

Table 93 also shows that the items' loadings on the respective factors range from .68 (item 4) to .92 (item 2). Their loadings on the respective other factor are at the same time very low. Item 4 thus has a very good loading, the other items have excellent loadings (Comrey and Lee as cited in Tabachnick & Fidell, 1996). Consequently, all items can be included based on their loadings.
Considering the number of variables loading on each factor (cf. Guadagnoli & Velicer, 1988), the extraction of two factors becomes questionable since the scale contained only five items.

4.3.4.1.4 Investigation of the extracted readiness factors' interpretability

For the examination of the factor solution from a conceptual perspective, the following table (Table 94) presents the scale items and their respective loadings.

Table 94. *Readiness scale factor solution with items and their respective loadings*

Item number	Questionnaire item	Factor loading
Factor 1: Expectations		
2	I knew up front what I could expect from the HANS sessions.	.92
3	I was aware of the HANS-sessions' expected results from the beginning.	.89
1	I knew before the HANS-sessions which effects the sessions should have on my skills.	.83
Factor 2: Usefulness of experience		
5	Usually, practice helps me to increase my performance.	.87
4	As far as I know, the skills conveyed in the HANS sessions reflect the examination requirements.	.68

Note. Method of extraction: Principal components analysis.
Table contains the translation of the German questionnaire items. For the German items used in the questionnaire refer to appendix D.

As Table 94 shows, the items loading on factor 1 concern the students' readiness regarding their *expectations*, the items loading on factor 2 regard the students' readiness concerning the *usefulness of experience*.
From a conceptual viewpoint, the scales are thus interpretable.

4.3.4.1.5 Investigation of the readiness factors' internal consistency

To test the factors' reliability, their internal consistency was investigated using Cronbach's α (see Table 95).

Table 95. *Cronbach's alpha of the readiness factors*

	Readiness scale		
	Number of items	Number of cases	Standardized alpha coefficient
Factor 1	3	404	.88
Factor 2	2	401	.40

The obtained alpha coefficient for factor 1 can be considered good (Bortz & Döring, 1995), especially for a scale with only three items. The alpha coefficient for factor 2 is low, as should be expected if the factor consists of only two items (Bortz & Döring, 1995).

4.3.4.1.6 Conclusion: Final factor solution of the readiness scale

The results of the previous factor analysis overall suggest the extraction of two factors, as indicated by the two factors with eigenvalues greater than 1 and a slope change in the screeplot after the second factor. With 73.9%, these two factors furthermore explain a good amount of variance. The communalities of all items were acceptable, as were the factor loadings that approached simple structure. The analysis furthermore yielded two interpretable factors. The items loading on factor 1 deal with the students' readiness concerning their training *expectations*, the items on factor 2 tap the students' readiness in regard to the *usefulness of experience* contained in the training.

The factor analysis has nevertheless to be viewed critically since the original scale comprised only five items, thus yielding factors with a very small number of items, which is also reflected in the resulting low alpha of factor 2.

In the light of the factor analysis' exploratory nature, the two factors will be used in the further analyses.

4.3.4.2 Examination of the intervention effect on the students' readiness

In order to test *Hypothesis 21*, a MANOVA was calculated to investigate if the revised seminar had an effect on students' readiness.

4.3.4.2.1 Descriptive statistics of the students' readiness

The following table (Table 96) contains the descriptive statistics for the students' readiness concerning their training *expectations* and the *usefulness of the experience* contained in the training.

Table 96. *Descriptive statistics of the students' readiness*

Factor	Training group 0 (n = 177)		Training group 1 (n = 98)		Training group 2 (n = 137)	
	M	***SD***	***M***	***SD***	***M***	***SD***
1 Expectations	3.88_a	1.11	3.42_b	.96	3.33_b	1.09
2 Usefulness of experience	2.29_a	.71	2.13_a	.69	2.10_a	.71

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).
Means in a row that do not share subscripts are significantly different ($p \leq .05$) from each other.

As Table 96 shows, training group 0 agreed least with both, the *expectations* and *usefulness of experience* items, followed by training group 1 and 2, respectively. Overall, the groups show more agreement with the *usefulness of experience* items than the expectations items. The means indicate that the groups "rather disagreed" (*M*=3.88) to "rather agreed" (*M*=3.33) with the expectations items. Concerning the *usefulness of experience*, the groups "mostly agreed" with the items as indicated by the group means between 2.29 and 2.10.

4.3.4.2.2 Inferential statistics of the students' readiness

A MANOVA was calculated with the extracted factors as dependent measures and training group as the independent variable. The following table (Table 97) contains the results.

Table 97. *Multivariate analysis of the students' readiness*

Effect	Pillai-Spur	*F*	Hypothesis *df*	Error *df*	*p*	Eta^2
Group	.06	5.96	4	818	*.000***	.03

Note. ** $p < .01$; significant effect in italics.

The effect for group is significant ($p < .01$) on the multivariate level. The univariate tests (see Table 98) showed that the significant multivariate effect is reflected in both univariate tests. The respective analysis yielded small effect sizes (Cohen, 1988).

Table 98. *Univariate tests of the students' readiness*

Source	Factor	Sum of squares	*F*	*p*	Eta^2
Group	1. Expectations	26.84	11.74	*.000***	.05
	2. Usefulness of experience	3.17	3.17	*.043**	.02

Note. df = 2, 409. ** $p < .01$; * $p < .05$ significant effects in italics.

Subsequently conducted post-hoc tests (Scheffé contrasts) yielded significant results ($p < .05$) for the comparisons on factor 1 between training group 0 and both of the other groups (cf. Table 96), training groups 1 and 2 did not significantly differ from each other regarding their ratings on this factor.
The contrasts on factor 2 did not yield significant results.

Hypothesis 21 could thus be supported for factor 1, but could only be partially supported for factor 2, since the univariate tests yielded significant results, but not the post-hoc pair-wise comparisons.

4.3.4.2.3 Conclusion

Hypothesis 21 was supported by the data on the multivariate and univariate level. A small effect was found (Cohen, 1988). Concerning the students' readiness for the training, the results show that training group 0 felt as expected least ready for the training with regard to their expectations (factor 1). Again, as expected, there was no significant difference between training groups 1 and 2. The results thus seem to suggest that the revised seminar was effective in increasing the students' readiness for the training: They knew better what they could expect from the training and what was expected of them.
The results further indicate that the students felt overall more ready for the training with regard to usefulness of the experience in the training than they did in regard to what to expect from the training. On this second factor, training group 0 also felt least ready for the training, again followed by training groups 1 and 2, respectively. On the univariate level, the differences reached significance in the expected direction but did not reach significance in the conducted post-hoc contrasts. It has to be noted, that factor 2 consisted of only two items which limited its reliability.
The results thus seem to suggest that the revised seminar led to a better trainee readiness concerning the usefulness of the experience contained in the training but that the intervention was not strong enough to lead to significant differences on the level of pair-wise comparisons.

4.3.5 Self-efficacy

As another individual characteristics variable, students' task-specific self-efficacy regarding anesthetic tasks was investigated in the pretest.

The seminar intervention was not targeted at this variable, consequently, no intervention hypothesis was formulated and the following sections present the examination of the scale structure (4.3.5.1) and the according descriptive and inferential statistics (4.3.5.2).

4.3.5.1 Examination of the self-efficacy scale's structure

In the pretest questionnaire students' task-specific self-efficacy was investigated by using eight items. A principal components analysis with orthogonal varimax rotation with Kaiser normalization was used to investigate the scale structure. 423 cases were included in the analysis.

4.3.5.1.1 Investigation of the self-efficacy scale's factor extraction

The following table (Table 99) displays the components' eigenvalues.

Table 99. *Self-efficacy factors' eigenvalues and explained variance*

	Initial factor matrix			Rotated factor matrix		
Component	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**	**Eigenvalue**	**% of explained variance**	**Cumulated % of explained variance**
1	3.78	47.25	47.25	3.12	39.02	39.02
2	1.08	13.51	60.76	1.74	21.74	60.76
3	.74	9.20	69.96			
4	.65	8.17	78.12			
5	.63	7.83	85.96			
6	.44	5.47	91.42			
7	.40	4.94	96.36			
8	.29	3.64	100.00			

Note. Principal components method of extraction.

As Table 99 shows, the first factor to be extracted has an eigenvalue of 3.12 after rotation, the second factor a much smaller eigenvalue of 1.74. Together, they explain 60.76% of variance. Following the Kaiser-Guttman criterion strictly, two factors should be extracted, but the extraction of only one factor also seems possible.

A second criterion used to determine the number of factors to be extracted is the screeplot as shown in Figure 13.

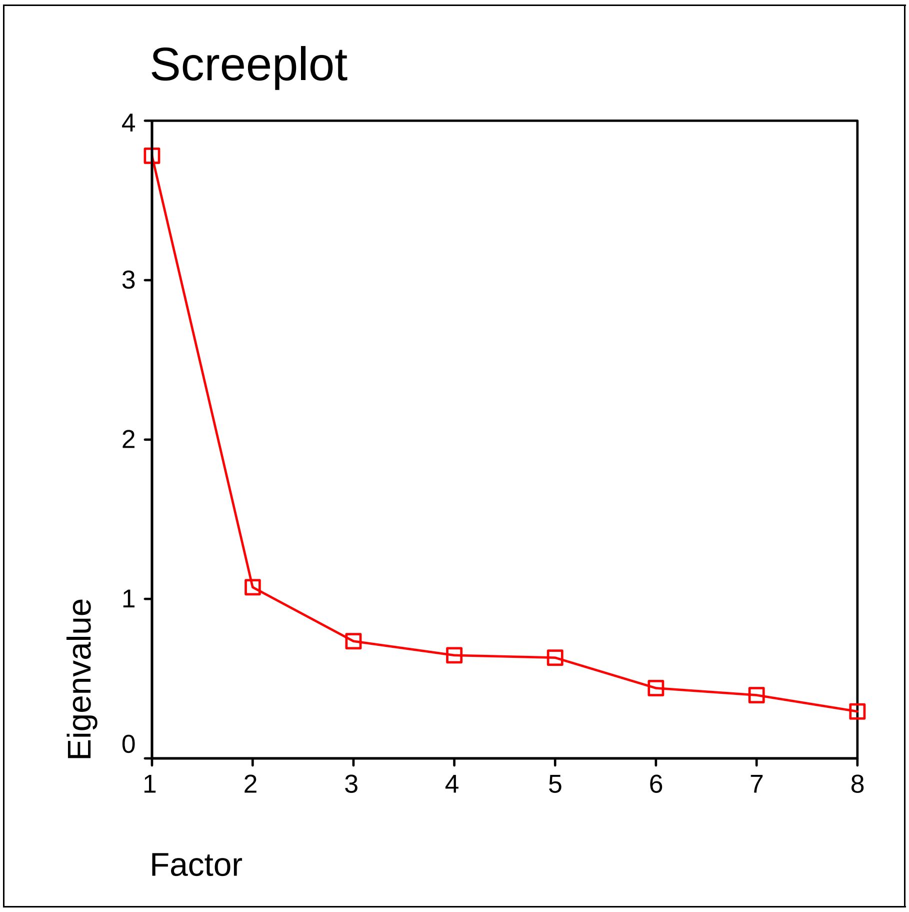

Figure 13. Screeplot of the self-efficacy factors' eigenvalues

Reflecting the results of Table 99, the screeplot shows the sharp drop in the eigenvalues after the first factor. Its slope furthermore changes after factor 2 and again after factor 3. The visual inspection of the screeplot thus also suggests the extraction of one or two factors.

4.3.5.1.2 Investigation of the self-efficacy scale's item adequacy: communalities

The following table (Table 100) contains the factor solution for the self-efficacy items, including the variables' factor loadings a_{ij} and their respective communalities h^2.
The investigation of the communalities h^2 reveals that the amount of variance of the observed variables accounted for by the factors is rather high, ranging from 43% (item 6) to 79% (item 8) indicating homogeneity among the variables (cf. Tabachnick & Fidell, 1996).
Based on their communalities, all items can thus be included in the final factor solution.

Table 100. *Rotated factor loading matrix of the self-efficacy items*

Item number	**Communalities h^2**	**Factor loadings a_{ij}**	
		1	**2**
2	.73	.83	.20
3	.68	.76	.32
4	.62	.75	.24
1	.57	.74	.14
7	.47	.68	.07
8	.79	-.04	.89
5	.57	.39	.65
6	.43	.35	.55

Note. Method of extraction: Principal components analysis. Method of rotation: Varimax with Kaiser-normalization.

4.3.5.1.3 Investigation of the self-efficacy scale's item adequacy: factor loadings

A look at Table 100 further shows that the items' factor loadings range from .55 (item 6) to .89 (item 8). As suggested by Comrey and Lee (as cited in Tabachnick & Fidell, 1996) factor loadings above .55 will be considered good (30% of overlapping variance), above .63 (40% of overlapping variance) as very good, and above .71 (50% of overlapping variance) as excellent. In this data set five variables have excellent factor loadings (2, 3, 4, 1, 8), two have very good factor loadings (7, 5) and item 5 still loads well on factor 2. It can further be observed that the items' loadings on the respective other factor are all low. All of the loadings can thus be considered acceptable and a two-factor solution seems appropriate.

Guadagnoli and Velicer (1988) further recommend taking into account not only the actual factor loading, but also the number of variables loading on each factor. They suggest to only interpret factors if more than four variables load on them with .60 or higher. In the case of the self-efficacy scale, this would support a single factor solution, rather than a two-factor solution.

4.3.5.1.4 Investigation of the extracted self-efficacy factors' interpretability

Following the described statistical considerations, the interpretability of the resulting factor solution should be investigated from a conceptual perspective (Tabachnick & Fidell, 1996).

Table 101. *Self-efficacy factor solution with scale items and their respective loadings*

Item number	Questionnaire item	Factor loading
Factor 1: Simulation session self-efficacy		
2	I know exactly that I am able to accomplish the tasks during the simulation sessions.	.83
3	I remain calm anticipating possible difficulties during the simulation because I can trust my own abilities.	.76
4	I will be able to apply the content of the simulation sessions well.	.75
1	I'm convinced that I will achieve the learning goals of the anesthesia simulation sessions.	.74
7	I know that I will succeed in independently conducting a standard anesthesia.	.68
Factor 2: Anesthetic task self-efficacy		
8	I will not need a long time to deduct and justify diagnoses, such as myocardial infarction, asthma, pulmonary embolism, ...[a]	.89
5	I'm convinced that I will succeed in using my interdisciplinary medical knowledge to form a diagnosis.	.65
6	On average, I'm probably more capable than others to complete anesthetic tasks.[a]	.55

Note. Method of extraction: Principal components analysis.
Table contains the English translation of the German questionnaire items. For the original German items refer to the questionnaire included in appendix C.
[a] Items were reversely worded in the original questionnaire.

Table 101 shows that the items loading on factor 1 center around the students' self-efficacy concerning their success in the simulation sessions. Item 7 has the smallest loading on factor 1 and is the only item on this factor that does not refer to the students' self-efficacy concerning specific tasks or aspects of the *simulation sessions*. Factor 2 comprises three items that tap the students' self-efficacy in regard to *anesthetic tasks* in general.

The conducted principal components analysis thus yielded two interpretable factors, with the restriction of one item (item 7) on factor 1.

4.3.5.1.5 Investigation of the self-efficacy factors' internal consistency

To test the factors' reliability, their internal consistency using Cronbach's α was examined (see Table 102).

Table 102. *Cronbach's alpha of the self-efficacy factors*

	Self-efficacy scale		
	Number of items	**Number of cases**	**Standardized alpha coefficient**
Factor 1	5	426	.84
Factor 2	3	434	.60

The resulting alpha coefficient for factor 1 can be considered good (Bortz & Döring, 1995). With .60, the alpha coefficient for factor 2 is rather low, but can still be considered acceptable in the light of the factor's small item number (Bortz & Döring, 1995).

The analysis also revealed (see Table 103) that the alpha coefficient of factor 1 is not diminished by deleting item 7 that is not fitting with the other items from a conceptual perspective (cf. 4.3.5.1.4).

Table 103. *Cronbach's alpha self-efficacy factor 1 after item deletion*

	Self-efficacy scale		
	Number of items	**Number of cases**	**Standardized alpha coefficient**
Factor 1 excluding item 7	4	427	.84

The investigation of the factors' reliability thus suggests the feasibility of the two-factor solution with the deletion of item 7.

4.3.5.1.6 Conclusion: Final factor solution of the self-efficacy scale

Overall, the results of the principal components analysis suggested the extraction of one or two factors. Taken together, the evidence seems to yield better support for the two-factor solution: two factors with eigenvalues greater than 1.00 can be extracted, a slight drop after factor 2 can be observed in the screeplot and the items have high loadings on only one of the factors. The conceptual investigation of the solution shows that the resulting two factors are interpretable. Factor 1 comprises items dealing with the students' self-efficacy in regard to the simulation sessions, whereas factor 2 contains items tapping aspects of the students' anesthetic task self-efficacy. Based on the conceptual investigation, item 7 will be excluded from further analyses. The two scales furthermore yield good to acceptable reliability scores, even after exclusion of item 7.

Since all of the statistical criteria yielded support for the two-factor solution, it will be used in the further analyses.

4.3.5.2 Descriptive and inferential statistics of the self-efficacy factors

The following table contains the descriptive statistics of the two extracted self-efficacy factors.

Table 104. *Descriptive statistics of the self-efficacy factors*

Factor		**Training group 0 (n = 185)**		**Training group 1 (n = 97)**		**Training group 2 (n = 157)**		**Overall (n = 439)**	
		M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
1	Simulation session self-efficacy	2.89	.76	3.08	.73	2.89	.79	2.93	.77
2	Anesthetic task self-efficacy	3.24	.76	3.21	.71	3.18	.86	3.21	.78

Note. 6-point scale (1 = strongly agree to 6 = strongly disagree).

The means indicate only slight agreement with the items for all the groups concerning their simulation session self-efficacy and even less agreement with the items concerning the anesthetic task self-efficacy. There are only small differences between the means of the different groups.

A subsequently conducted MANOVA did not reveal a significant effect ($p < .05$) for group.

It can thus be concluded that the training groups did not differ concerning their task-specific self-efficacy.

4.4 Relations between the measured variables

The following sections describe the results of the explorative tests of key relations between the measured variables. Section 4.4.1 investigates the relations between the individual characteristics variables and the training outcomes, and section 4.4.2 investigates the relationship between learning and behavior.

4.4.1 Relations between individual characteristics variables and training outcomes

The following sections investigate the main research question concerning the individual characteristics variables' influence on the training outcomes.

4.4.1.1 Individual characteristics variables and reactions

The following hypothesis was formulated in regard to the influence of specific individual characteristics variables on training reactions.

Hypothesis 22
Individual characteristics variables (readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy) predict training reaction scores.

To test the above hypothesis, the respective individual characteristics variables were entered into a multiple regression equation as predictors, the overall training reaction score was entered as the criterion.

Table 105. *Descriptive statistics and intercorrelations for reactions and individual characteristics variables*

Variable			**1.**	**2.**	**3.**	**4.**	**5.**
	M	***SD***			**r_{ij}**		
1. Reactions (DV)	1.86	.70	-				
2. Readiness	3.03	.80	*.35***	-			
3. Expectation fulfillment	-.26	.76	*-.36***	*-.12***	-		
4. Instrumentality expectancy	1.86	.58	*.27***	*.19***	*.11**	-	
5. Self-efficacy	3.10	.67	*.29***	*.21***	.07	*.10**	-

Note. N = 390
* *p* < .05. ** *p* < .01.

The above table (Table 105) shows the correlations between the dependent variable (training reactions) and the predictors (individual characteristics variables). All of the predictors correlate significantly ($p < .01$) with the dependent variable. Overall, these correlations were medium to high. The highest correlation was found between expectation fulfillment and reactions (-.36), the lowest between instrumentality expectancy and reactions (.27).

Table 106. *Summary regression analysis for reactions on individual characteristics variables*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.59	.34	51.00	*.000***

Note. DV= reactions
(N = 390)

As shown in Table 106, the regression analysis yielded a significant R ($F_{(4,385)} = 51.00$, $p < .001$). Overall, the regression equation explained 34% of variance in the reaction scores.

Table 107. *Regression analysis for reactions on individual characteristics variables*

Variable	***B***	***SEB***	***ß***
Intercept	-.12	.17	
Readiness	.18	.04	*.21***
Expectation fulfillment	-.35	.04	*-.38***
Instrumentality expectancy	.30	.05	*.25***
Self-efficacy	.25	.04	*.24***

Note. DV= reactions
$R = .59$. $R^2_{adj} = .34$ (N = 390, $p < .001$)
** $p < .01$.

All of the independent variables contributed significantly to the prediction of the participants' reaction scores (Table 107).

Hypothesis 22 could thus be supported.

4.4.1.2 Conclusion

The conducted analyses showed that about a third of the variability in participants' reactions can be explained by their readiness for the training, the degree of expectation fulfillment, the degree of perceived instrumentality, and their self-efficacy. *Hypothesis 22* could thus be supported. Trainees who have been prepared for the training, that is, know what is expected of them, react more positively to the training. Trainees whose expectations have been fulfilled or exceeded react more favorably to the training. Trainees who perceive the training as instrumental react more favorably to the training as do trainees who believe that they can succeed in the training.

These findings highlight the need for specific pre-training interventions to assure that trainees are prepared for the training content and the goals of the training (readiness), that they form appropriate expectations concerning the methods used in the training and the kind of participation expected from them and that the training meets these expectations (expectation fulfillment), that the trainees understand in which way the KSAs acquired in the training will be beneficial to them (instrumentality expectancy), and that they feel confident that they can succeed in the training (self-efficacy).

Since it cannot be ruled out that part of the correlations are due to answering tendencies, these findings need to be corroborated by further studies.

4.4.1.3 Individual characteristics variables and learning

Concerning the influence of the individual characteristics variables on the learning outcome, two hypotheses were formulated and tested as described in the following sections. *Hypothesis 23* was based on the findings in the respective training literature (Cannon-Bowers et al., 1995; Holton, 2005), *Hypothesis 24* was formulated additionally because this study employs a pre-post-test design other training studies previously oftentimes lacked.

Hypothesis 23
Individual characteristics variables predict learning post-training scores.

Hypothesis 24
Individual characteristics variables predict learning difference scores.

4.4.1.3.1 Individual characteristics variables and learning post-training scores

A standard multiple regression was performed between self-reported post-test anesthesia competency scores ($\text{learning}_{\text{post}}$) as the dependent variable and the individual characteristics variables as the independent variables.

The following tables (Table 108, Table 109, Table 110) display the results of the analyses.

Table 108. *Descriptive statistics and intercorrelations for learning$_{post}$ and individual characteristics variables*

Variable			**1.**	**2.**	**3.**	**4.**	**5a.**	**5b.**	**6.**
	M	***SD***	**r_{ij}**						
1. Learning$_{post}$ (DV)	2.84	.79	-						
2. Readiness	3.03	.80	*.41***	-					
3. Expectation fulfillment	-.26	.76	*-.29***	*-.11***	-				
4. Instrumentality expectancy	1.85	.58	.02	*.19***	*.11***	-			
5a. Perf. goal orientation	3.71	.99	.03	.01	-.08	.00	-		
5b. Learn. goal orientation	2.87	.89	*.09**	*.11**	*.16***	*.31***	*.08**	-	
6. Self-efficacy	3.10	.67	*.41***	*.20***	.07	*.10**	*-.14***	.34	-

Note. N = 389
* *p* < .05. ** *p* < .01.

As the above table (Table 108) shows, the correlations between learning$_{post}$ and the individual characteristics variables were significant except for instrumentality expectancy and performance goal orientation. The correlation between learning$_{post}$ and learning goal orientation was significant but low, the correlations with the other three variables were medium to high.

Table 109. *Summary regression analysis for learning$_{post}$ on individual characteristics variables*

Model	***R***	***$R^2_{adjusted}$***	***F***	***α***
1	.60	.36	36.57	*.000***

Note. DV= learning$_{post}$
(N = 389)

The conducted regression analysis (Table 109) yielded an *R* for regression that was significantly different from zero ($F_{(6,383)} = 36.57$, $p < .001$). The regression equation explained 36% of variance in the learning posttest scores as the dependent variable.

Table 110. *Regression analysis for $learning_{post}$ on individual characteristics variables*

Variable	***B***	***SEB***	***ß***
Intercept	.40	.24	
Readiness	.31	.04	*.31***
Expectation fulfillment	-.28	.04	*-.27***
Instrumentality expectancy	-.05	.06	-.03
Perf. goal orientation	.05	.03	.07
Learn. goal orientation	-.02	.04	-.03
Self-efficacy	.46	.05	*.40***

Note. DV= $learning_{post}$
(N = 389)
* $p < .05$. ** $p < .01$.

As can further be seen in Table 110, only three of the six independent variables contributed significantly to the prediction of the participants' learning scores in the posttest: For readiness ($ß = .31$, $p < .01$) expectation fulfillment ($ß = -.27$, $p < .01$), and self-efficacy ($ß = .40$, $p < .01$).

Hypothesis 23 could thus be supported for three variables.

4.4.1.3.2 Individual characteristics variables and learning difference scores

All of the individual characteristics variables were also entered in a standard regression as predictors for the students' learning gains, that is the difference scores between their self-reported posttest and pretest scores[13] ($learning_{diff}$ = posttest - pretest).
The following tables (Table 111, Table 112, Table 113) contain the results of the analyses.

[13] The pretest raw scores were subtracted from the posttest raw scores in this set of analyses to stay within the correlation pattern. Negative difference scores thus indicate learning gains.

Table 111. *Descriptive statistics and intercorrelations for learning$_{diff}$ and individual characteristics variables*

Variable			**1.**	**2.**	**3.**	**4a.**	**4b.**	**5.**	**6.**
	M	***SD***				**r$_{ij}$**			
1. Learning$_{diff}$ (DV)	-.95	.67	-						
2. Readiness	3.05	.78	*.17***	-					
3. Expectation fulfillment	-.26	.76	*-.33***	-.11	-				
4. Instrumen-tality expec-tancy	1.87	.58	*.10**	.17	.11	-			
5a. Perf. goal orientation	3.72	.99	.03	.02	-.09	-.01	-		
5b. Learn. goal orientation	2.88	.89	-.06	.10	.16	.30	.09	-	
6. Self-efficacy	3.11	.67	-.08	.19	.09	.10	-.13	.35	-

Note. N = 375
* $p < .05$. ** $p < .01$.

Of the six individual characteristics variables only three correlated significantly with the participants' learning difference scores (Table 111). Overall, the correlations were moderate to low. The highest correlation, a negative correlation, was found between the students' expectation fulfillment scores and their learning difference scores. This correlation indicates, that more self-reported learning was correlated with students' exceeded expectations. The second highest correlation was found between students' readiness and their learning difference scores. The higher they had reported their readiness, the more they reported they had learned. The last significant correlation was found between the students' perceived instrumentality of the training and their learning difference scores. Again, the higher they had rated the trainings' instrumentality, the more they reported had learned.

Table 112. *Summary regression analysis for learning$_{diff}$ on individual characteristics variables*

Model	***R***	***R^2$_{adjusted}$***	***F***	***α***
1	.39	.14	10.71	*.000***

Note. DV= learning$_{diff}$
(N = 375)

The conducted regression analysis (Table 112) resulted in an *R* for regression that was significantly different from zero ($F_{(6,368)} = 10.71$, $p < .001$). The regression equation explained 15% of variance in the learning difference scores.

Table 113. *Regression analysis for learning$_{diff}$ on individual characteristics variables*

Variable	*B*	*SEB*	*ß*
Intercept	-1.29	.24	
Readiness	.11	.04	*.13***
Expectation fulfillment	-.28	.04	*-.32***
Instrumentality expectancy	.15	.06	*.13***
Perf. goal orientation	-.01	.03	-.01
Learn. goal orientation	-.03	.04	-.04
Self-efficacy	-.08	.05	-.08

Note. DV= learning$_{diff}$
(N = 375)
** $p < .01$.

Table 113 shows that only three of the six independent variables contributed significantly to the prediction of the participants' learning difference scores.

Hypothesis 24 could thus be supported for expectation fulfillment ($ß = -.32$, $p < .01$), readiness ($ß = .13$, $p < .01$) and instrumentality expectancy ($ß = .13$, $p < .01$).

4.4.1.3.3 Conclusion

The regression results indicate that about a third of the variation in students' posttest learning scores (see 4.4.1.3.1) could be predicted by the assessed individual characteristics variables. More specifically, the students' task-specific self-efficacy, the degree of their expectation fulfillment, and their readiness for the training predicted the students' self-reported anesthesia competencies in the posttest. The predictive function of the students' instrumentality expectancy and their goal orientation also expected in *Hypothesis 23* was not supported by the data.

The test of the individual characteristics variables' function as predictors for the students' learning gains (see Table 113) revealed that overall 15% of the variability in the students' learning gains was predicted by the respective variables. The degree to which the students' expectations were fulfilled, how ready they felt for the training, and the degree to which they perceived the training as instrumental significantly predicted their learning gains due to the training. Students' goal orientation and self-efficacy did not have the predicted (*Hypothesis 24*) significant influence on the learning gains.

The students' readiness for the training and the degree to which their expectations were fulfilled thus predicted both, their self-reported posttest performance and their learning gains in the training.

The interesting finding is that students' task-specific self-efficacy significantly predicted their posttest training performance but not their learning gains. It is thus possible that students with high self-efficacy already possessed rather high competencies before the training. This would then imply the possibility that the self-efficacy merely reflected their competence.

Furthermore interesting is that the perceived instrumentality of the training predicted the learning gains but did not predict the posttest performance. This result provides a first clue that students who perceived the training as instrumental to their goals, namely succeeding in the exam, might for example exert more effort in the training and thus learn more. If one only considers the posttest performance as the learning measure, this relationship does not exist. More studies are thus needed that operationalize learning as learning gains, not merely as posttest performance.

The only two variables that were not significant predictors of either one of the learning outcomes were the goal orientation factors. It is possible, that the data did not support the hypotheses due to measurement issues. The experiences during the data collection process provided some cues that the participants might not have understood the aim of the goal orientation items. Future studies should thus try to use a different goal orientation scale.

4.4.1.4 Individual characteristics variables, learning, and behavior

The next hypothesis deals with the relation between the individual characteristics variables, learning, and behavior.

Hypothesis 25
Learning post-training scores mediate the relationship between individual characteristics variables and behavior.

To investigate this hypothesis, a set of regression analyses was calculated for each of the individual characteristics variables following the steps suggested by Baron and Kenny (1986). For reasons of succinctness, the analyses are only reported up to the step that did not yield the required significant result anymore.

4.4.1.4.1 Test of learning as a mediator between readiness and behavior

As the first step, a regression analysis was calculated to investigate the effect of the independent variable on the mediator.

Table 114. *Summary regression analysis for learning$_{post}$ on readiness*

Model	*R*	$R^2_{adjusted}$	*F*	*α*
1	.39	.15	72.98	*.000***

Note. DV= learning$_{diff}$
(N = 375)

Table 115. *Regression analysis for learning$_{post}$ on readiness*

Variable	*B*	*SEB*	*ß*
Intercept	1.69	.14	
Readiness	.38	.05	*.39***

Note. DV= learning$_{post}$
(N = 412)
** *p* < .01.

The analysis (Table 114, Table 115) yielded a significant *ß* for readiness (*ß* = .39, *p* < .001). The first condition of the mediator effect was thus met.

In the second step, behavior as the dependent variable, was regressed on readiness as the independent variable.

Table 116. *Summary regression analysis for behavior on readiness*

Model	*R*	$R^2_{adjusted}$	*F*	*α*
1	.07	.00	1.99	.160

Note. DV= behavior
(N = 365)

Table 117. *Regression analysis for behavior on readiness*

Variable	*B*	*SEB*	*ß*
Intercept	3.91	.53	
Readiness	.24	.17	.07

Note. DV= behavior
(N = 365)

The regression analysis (Table 116, Table 117) did not support the assumption of readiness as a significant predictor for behavior.
The second condition of the hypothesized mediation relationship was thus not met.

Consequently, the data did not support *Hypothesis 25*.

4.4.1.4.2 Test of learning as a mediator between expectation fulfillment and behavior

As step one, a regression analysis was conducted to investigate the effect of the expectation fulfillment as the independent variable on $learning_{post}$ as the mediator.

Table 118. *Summary regression analysis for $learning_{post}$ on expectation fulfillment*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.30	.08	37.25	*.000***

Note. DV= $learning_{post}$
(N = 395)

Table 119. *Regression analysis for $learning_{post}$ on expectation fulfillment*

Variable	***B***	***SEB***	***ß***
Intercept	2.76	.04	
Expectation fulfillment	-.31	.05	*-.30***

Note. DV= $learning_{post}$
(N = 395)
** $p < .01$.

The analysis (Table 118, Table 119) yielded a significant *ß* for expectation fulfillment that was significantly different from zero ($ß = -.30$, $p < .001$).
The data thus met the first condition for the mediator effect.

In step two, behavior as the dependent variable, was regressed on the independent variable expectation fulfillment.

Table 120. *Summary regression analysis for behavior on expectation fulfillment*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.09	.01	2.67	.103

Note. DV= behavior
(N = 355)

Table 121. *Regression analysis for behavior on expectation fulfillment*

Variable	***B***	***SEB***	***ß***
Intercept	4.53	.15	
Expectation fulfillment	-.30	.18	-.09

Note. DV= behavior
(N = 355)

Expectation fulfillment did not serve as a significant predictor for behavior (Table 120, Table 121).
The second condition of the hypothesized mediation relationship was thus not met.

Consequently, the data for expectation fulfillment did not support *Hypothesis 25*.

4.4.1.4.3 Test of learning as a mediator between instrumentality expectancy and behavior

The first regression analysis was calculated to test the effect of instrumentality expectancy as the independent variable on learning_{post} as the mediator.

Table 122. *Summary regression analysis for* $\textit{learning}_{post}$ *on instrumentality expectancy*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.03	.00	.26	.608

Note. DV= learning_{post}
(N = 396)

Table 123. *Regression analysis for* $\textit{learning}_{post}$ *on instrumentality expectancy*

Variable	***B***	***SEB***	***ß***
Intercept	2.77	.13	
Instrumentality expectancy	.04	.07	.03

Note. DV= learning_{post}
(N = 396)

The analysis did not yield the expected significant *ß* for instrumentality expectancy (Table 122, Table 123).
The data did thus not meet the first condition for the mediator effect.

Consequently, the data for instrumentality expectancy did not support *Hypothesis 25*.

4.4.1.4.4 Test of learning as a mediator between situational performance goal orientation and behavior

As the first step, a regression analysis was conducted to investigate the effect of performance goal orientation as the independent variable on the mediator learning_{post}.

Table 124. *Summary regression analysis for learning$_{post}$ on performance orientation*

Model	***R***	***$R^2_{adjusted}$***	***F***	***α***
1	.03	.00	.35	.556

Note. DV= learning$_{post}$
(N = 394)

Table 125. *Regression analysis for learning$_{post}$ on performance orientation*

Variable	***B***	***SEB***	***ß***
Intercept	2.75	.15	
Performance goal orientation	.02	.04	.03

Note. DV= learning$_{post}$
(N = 394)

The analysis did not result in the expected significant *ß* for performance goal orientation (Table 124, Table 125).

The data did thus not meet the first condition for the mediator effect and did consequently not support *Hypothesis 25*.

4.4.1.4.5 Test of learning as a mediator between situational learning goal orientation and behavior

To examine the proposed mediator effect, the mediator learning$_{post}$ was regressed in a first step on learning goal orientation as the independent variable.

Table 126. *Summary regression analysis for learning$_{post}$ on learning orientation*

Model	***R***	***$R^2_{adjusted}$***	***F***	***α***
1	.09	.01	3.49	.556

Note. DV= learning$_{post}$
(N = 394)

Table 127. *Regression analysis for learning$_{post}$ on learning orientation*

Variable	***B***	***SEB***	***ß***
Intercept	2.60	.13	
Learning goal orientation	.08	.05	.09

Note. DV= learning$_{post}$
(N = 394)

The analysis did not yield the expected significant *ß* for learning goal orientation (Table 126, Table 127).
The data did thus not meet the first condition for the mediator effect.

Consequently, *Hypothesis 25* could not be supported for learning goal orientation.

4.4.1.4.6 Test of learning as a mediator between self-efficacy and behavior

As the first step, a regression analysis was calculated to investigate the effect of self-efficacy as the independent variable on the mediator ($learning_{post}$).

Table 128. *Summary regression analysis for $learning_{post}$ on self-efficacy*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.41	.17	80.47	*.000***

Note. DV= $learning_{post}$
(N = 394)

Table 129. *Regression analysis for $learning_{post}$ on self-efficacy*

Variable	***B***	***SEB***	***ß***
Intercept	1.34	.17	
Self-efficacy	.48	.05	*.41***

Note. DV= $learning_{post}$
(N = 395)
** $p < .001$.

The analysis (Table 128, Table 129) yielded a significant *ß* for self-efficacy ($ß = .41$, $p < .001$).
The first condition for the mediator effect was thus met by the data.

In the second step, behavior as the dependent variable, was regressed on self-efficacy as the independent variable.

Table 130. *Summary regression analysis for behavior on self-efficacy*

Model	***R***	$R^2_{adjusted}$	***F***	***α***
1	.10	.01	3.78	*.053**

Note. DV= behavior
(N = 390)

Table 131. *Regression analysis for behavior on self-efficacy*

Variable	***B***	***SEB***	***ß***
Intercept	3.44	.62	
Self-efficacy	.38	.20	.10*

Note. DV= behavior
(N = 390)
* $p \leq .055$.

As Table 130 and Table 131 show, the second regression equation also yielded a significant result. Self-efficacy contributed significantly to the prediction of the participants' behavior scores ($ß = .10, p \leq .05$).
The data thus also met the second condition of the hypothesized mediation relationship.

In the third step, behavior as the dependent variable was regressed on both, self-efficacy as the independent variable and learning_{post} as the mediator.

Table 132. *Summary regression analysis for behavior on self-efficacy and* $\textit{learning}_{post}$

Model	R	$R^2_{adjusted}$	F	α
1	.11	.01	1.98	.139

Note. DV= behavior
(N = 356)

Table 133. *Regression analysis for behavior on self-efficacy and* $\textit{learning}_{post}$

Variable	***B***	***SEB***	***ß***
Intercept	3.34	.71	
Self-efficacy	.16	.23	.04
$\textbf{Learning}_{post}$	.27	.20	.08

Note. DV= behavior
$R = .11$. $R^2_{adj} = .01$ (N = 356)

As can be seen in Table 132 and Table 133, the mediator (learning_{post}) could not be shown to affect the dependent variable in this equation. That is, the third condition for the mediator effect was not met by the data.

Hypothesis 25 could thus not be supported for self-efficacy.

4.4.1.4.7 Conclusion

The function of learning as a mediator between individual characteristics variables and behavior could not be established by the conducted tests. In five of the six examined relations, the data did not support the second condition of the mediation relationship. Specifically, the individual characteristics variables did not function as significant predictors of behavior, the dependent variable. An exception was self-efficacy. Self-efficacy did significantly predict behavior, but when self-efficacy and learning were entered in the equation in the third step, learning did not function as a significant predictor for behavior.

The individual characteristics variables' lack of influence on behavior might be due to measurement issues concerning either or both types of variables, or other factors, such as external motivational factors before the examination situation, might have exerted a stronger influence than the investigated factors.

4.4.2 Relations between learning and behavior

The following hypothesis concerns the relationship between the training outcomes on level two and three (Kirkpatrick, 1994): learning and behavior.

> *Hypothesis 26*
> The students' self-reported anesthesia skills at the end of the training ($learning_{post}$) predict their transfer performance (behavior).

4.4.2.1 Test of $learning_{post}$ as a predictor of behavior

To test the above hypothesis, $learning_{post}$ was entered in a regression equation as the predictor, the students' OSCE scores (behavior) were entered as the dependent variable (Table 134).

Table 134. *Summary regression analysis for behavior on* $learning_{post}$

Model	R	$R^2_{adjusted}$	F	α
1	.10	.01	3.71	*.055**

Note. DV= behavior
(N = 368)

Table 135. *Regression analysis for behavior on learning$_{post}$*

Variable	***B***	***SEB***	***ß***
Intercept	3.68	.51	
Learning$_{post}$	.33	.17	.10*

Note. DV= behavior
(N = 368)
* $p \leq .055$.

As can be seen in Table 135, the regression model reached significance ($F_{(1,366)} = 3.71$, $p \leq .055$) on a slightly adjusted α-level of .55. The analysis furthermore yielded a significant *ß* for learning$_{post}$ ($ß = .10$, $p \leq .055$).

Hypothesis 26 was thus supported on a slightly adjusted α-level.

4.4.2.2 Conclusion

The results of the regression analysis supported the hypothesis that the students' level of post-training anesthesia skills predict their performance in a subsequent exam, assessing the transfer of their skills. The results were significant on a slightly adjusted α-level of .055. This adjustment was deemed acceptable in the light of the involved measurement issues. Altogether, the posttest level of self-reported anesthesia skills predicted 10% (1% adjusted) of the variability in the transfer measure (OSCE-scores). The results thus provide a first indication that the post-training skills transferred to the application setting, namely the students' exam. The adjusted model was, however, only able to predict 1% of the variability. This small amount might be attributed to several factors. First of all, it is possible that the predictor and the criterion measured different skill subsets since the exam content could not exactly be matched to the questionnaire items. Secondly, the posttest skills were collected as self-report data, whereas the transfer skills were rated by instructors. Thirdly, it is very likely that the students also studied on their own after the training, thus creating another type of intervention. The latter behavior is not as likely to be found in the type of training programs evaluated in the training literature since the transfer measures are not perceived as exams by the participants.
Further studies are needed to corroborate these findings. They should include self-reported transfer measures as well as observer-based learning measures.

5 Discussion

Research in the training area has produced a great amount of knowledge about the design of learning environments, influence variables on training effectiveness, and training evaluation. The resulting models have, however, in many cases not been adopted by the disciplines that are undertaking respective training efforts. Medical simulation training is an example of such an area. The current study was conducted to apply instructional principles as specified in an instructional-psychology training model to the development and evaluation of basic anesthesia simulation training.

The concluding part of this dissertation contains the discussion of the study's results in the light of the presented theoretical aspects and previous research findings. In chapter 5.1 the development of the instructionally-based anesthesia simulation training program is summarized and reflected, followed by the summary and discussion of the evaluation study's major findings (5.2). Chapter 5.3 reviews the limitations of the development and evaluation process. Building upon this information, chapter 5.4 provides the general discussion of the findings and future perspectives for this area of research. The discussion part ends with a conclusion that also explicates implications of the current project for medical practitioners (5.5).

5.1 Development of an instructionally-based anesthesiology simulator-training program

The current project was undertaken to close an existing gap in the medical simulation area: The gap between theoretical training models and the actual simulation trainings conducted in medical school settings, in this study, the University of Heidelberg's medical school. Based on the model for simulator training effectiveness (Schaper, Schmitz et al., 2003) an anesthesiology simulator-training curriculum was thus developed that incorporated theoretical findings in the area of learning and training research. The following paragraphs will reflect upon specific issues that became evident during the design process.

As specified in the model (Schaper, Schmitz et al., 2003), the training development process was started by completing a set of instructional systems design tasks. These tasks comprised the needs analysis with organization, task, and person analyses; the subsequent specification of the learning objectives and the training content to achieve the objectives; and the derivation of evaluation criteria.

The needs assessment was conducted with the help of structured interviews with two SMEs. This process could have been enhanced by employing more thorough methods in the task and person analyses (see Sonntag, 2006b). A student sample could, for example, have been tested in the simulator to better judge the participants' level of knowledge and skills. Student representatives could have furthermore been involved in the complete needs assessment process to also adapt the new training program to their perceived needs and to thus increase acceptance of the development and evaluation process. Additional methods can be found in the work psychology literature (e.g., Sonntag, 1992; 2006) and have already been partially applied to the medical context (Buerschaper, Hofinger, & Harms, 2003).

Based on the results of the needs assessment, the training design process was begun. The design of the revised preparatory seminar and the full-scale simulation sessions reflected an integrated approach to teaching and learning, combining cognitivist and constructivist elements (Reinmann-Rothmeier & Mandl, 2001). The preparatory seminar was restructured and the instructional strategy revised to specifically foster student activity, student motivation, learning goal orientation, readiness, and transfer. It became evident during the design process and the following study that a very close contact must be kept between the training designer and the instructor to assure that the training is implemented as planned. This is especially important if the training designer is not a member of the medical faculty.

The simulation sessions were also designed to represent the integrated approach to teaching and learning with the cognitive apprenticeship approach's instructional principles serving as the guideline (Brown et al., 1989; Reinmann-Rothmeier & Mandl, 2001). Overall, the cognitive apprenticeship guidelines could be easily realized in the simulation sessions. The reflection sessions might have been improved by providing students the opportunity to review their performance on video as this strategy has been shown to improve anesthesiologists' performance in the simulator (Byrne et al., 2002). The question is, however, if this would be feasible in a mandatory curriculum with large numbers of students. In the current setting, it could not be realized due to time restrictions.

It became further evident during the simulation design process that the element of providing practice opportunities (Salas & Cannon-Bowers, 1997) could not be realized in a satisfactory fashion. Only two full-scale simulation sessions were included in the curriculum and due to resource constraints, could not be extended. It would have been desirable to at least allow each student to assume each of the roles in the scenarios (cf. 3.3.2.2.2.1). Furthermore desirable from a design point of view would have been the provision of the opportunity for the students to practice in the simulator on their own after the completion of the initial sessions (cf. Savoldelli et al., 2005). These additional sessions could have supported overlearning. Overlearning means providing learners with continued practice after they have already mastered the task and has been shown to have a positive effect on retention (Baldwin & Ford, 1988; Driskell & Johnston, 1992). Again, this was not possible based on resource restrictions. Future design projects should thus try to secure appropriate resources. The problem of insufficient practice could in this case, however, be partially attenuated by the part-task simulation sessions. In the light of the training outcomes, this issue will also be discussed below.

The major challenge during the design process was the training's setting within the restrictions of the mandatory medical school curriculum. It led to a development process divided into two phases that resulted in study design limitations that will be elaborated below (see 5.3). This experience highlights the importance of a thorough organizational analysis and

the securing of organizational support (Goldstein & Ford, 2002) to clearly identify at the beginning which constraints and affordances are contained in the setting.
In the evaluation design process it also became evident that beyond the summative evaluation of the initial training, additional formative evaluation procedures should be included in the design (cf. Wottawa & Thierau, 1998). The revised instructional elements should have been pretested and feedback loops should have been incorporated to systematically include the instructors' experiences.

In conclusion it can be stated that the model (Schaper, Schmitz et al., 2003) provided suitable guidelines for the development of a revised anesthesiology simulation training program based on theoretical approaches and findings in instructional psychology. Most of the guidelines contained in the model could be easily implemented, even within the restrictions of a mandatory medical school curriculum. Some of the issues mentioned above, such as resource limitations, can possibly be prevented in future development and evaluation projects, others, such as curriculum-linked restrictions, are inherent to mandatory medical school settings.

5.2 Evaluation of the instructionally-based simulator-training program

Concerning the effectiveness of training programs, Arthur, Bennett, Edens, and Bell's recent meta-analysis (2003) showed that organizational training is generally effective. The effectiveness varies, however, as a function of the training method, the skills trained, and the criteria used to assess effectiveness. This result highlights the need to thoroughly evaluate individual training programs and not to rely on general findings in the training or medical education literature. Despite the widespread curriculum reform in medical education, studies investigating the effects of the new training methods on the trained skills are still rare, however (Jünger et al., 2005). For this reason, the instructionally revised anesthesiology training program was evaluated based on the presented training model (Schaper, Schmitz et al., 2003). The following sections first display the training outcomes structured along Kirkpatrick's (1994) evaluation levels (5.2.1-5.2.3), then present the individual characteristics findings (5.2.4), and close with the found relations between specific model elements (5.2.5)[14].

5.2.1 Training reactions

The reaction part of the questionnaire assessed several aspects of the curriculum and investigated if the training design changes lead to different participant reactions. The results are subsumed under two headings: Section 5.2.1.1 summarizes and discusses the students' reaction results to the full-scale simulator, the simulation situation, and the used scenarios.

[14] Note that in the discussion the frame of the german school grade rating scale is left and that indications of "higher" ratings now refer to better outcomes.

Section 5.2.1.2 summarizes and discusses the students' appraisal of the instructors' teaching skills. Section 5.2.1.3 then concludes with an overall discussion of the reaction outcomes.

5.2.1.1 Reactions to the full-scale simulator and scenarios

The first assessed area were the students' *reactions to the simulation and the simulator*. Overall, the results of the reaction data in regard to the simulation and the simulator were mixed. None of the hypotheses was fully supported, three of them were partially supported.

Concerning the participants' reactions to the simulation and the simulator, the results did not yield the expected better ratings of the full-scale simulation training's value for the curriculum by training groups 1 and 2. *Hypothesis 1* could thus not be supported: All of the groups considered the simulation an important component of the curriculum. Concerning their recommendations to use the full-scale simulation training in its respective form in the subsequent semester, training group 0 indicated the most agreement while training group 1 agreed significantly less to this item. A possible explanation for this finding is that the implementation of the new training methods was not yet optimal in the first semester. This explanation is supported by a found increase in the agreement of group 2 as compared to group 1 that did, however, not reach significance.

Concerning their perceptions of the simulator's realism (*Hypothesis 2*) the groups were not expected to differ because the simulator's realism was not affected by the instructional redesign. This hypothesis was partially supported. The groups did not differ in their ratings concerning the helpfulness of the simulator's realism for learning the content. Concerning the motivational effect of the simulator's realism, however, training group 1 again perceived it as significantly less motivating than training group 0. It remains unclear at this point what could have caused training group 1's lower ratings of the simulator's realism's motivational content. Also not targeted by the interventions was the simulator's suitability as a learning tool. *Hypothesis 3* could only be partially supported by the data, as a significant difference was found between the groups on their simulator suitability ratings. This difference was, however not reflected on the pairwise comparison level. On a descriptive level, it was again training group1 that gave the lowest ratings. The reason for this difference can't be detected with the current data base.
In line with the assumptions, the groups did, however, not differ in their ratings of the suitability of the simulator for skill acquisition.

Another aspect that was targeted with the interventions was the students' ability to adjust to the simulation situation. Training groups 1 and 2, who had received the video anchors, were expected to be better able to adjust to the simulation sessions *(Hypothesis 4*). *Hypothesis 4* could, however, not be supported by the data. The effect of the video anchors might not have been strong enough to lead to changes in the participants' ability to adjust to the following simulation situation. Future seminars should test the inclusion of a more thorough reflection phase in which the participants are encouraged to discuss their role and the necessary adjustments in the simulator. In the light of more recent research, the question of the students' ability to adjust to the simulation sessions furthermore seems to lose importance: A study with medical students, final-year anesthesia residents, and anesthesiologists found that the participants' familiarity or comfort with the simulation had little or no effect on their performance (Devitt et al., 2001).

The second area that was assessed were the students' *reactions to the scenarios*.
In *Hypothesis 5,* it was assumed that training group 2 considered the scenarios' difficulty level as most appropriate because they participated in the revised full-scale simulation sessions that provided instructional support and had also participated in the interspersed part-task emergency training practicing their basic subskills (Collins et al., 1989; Murray, 2005). As hypothesized, participants of training group 2 rated the used difficulty level of the scenarios as most adequate for their level of knowledge and significantly more so than training group 0. *Hypothesis 5* could thus partially be supported by the data. This provides first evidence that the intervention of the preparatory seminar and the change in the curriculum sequence was successful: the students felt better prepared for the simulation sessions. The provision of additional support in the course of the redesign of the simulation sessions based on cognitive apprenticeship principles alone did not seem to suffice to change the students' perceptions of the difficulty level. Only with the previously received additional part-task training did they find the scenarios adequately difficult. This finding is especially interesting as it has been pointed out that difficulty reactions have not yet received a lot of research (Brown, 2005).

Hypothesis 6 concerned the degree of motivation the students experienced through the scenarios. It was assumed that training group 2 was most motivated because of the authentic task augmented by the instructional support they received to prevent their being overwhelmed and because they could integrate the previously practiced skills in the full-scale simulation scenarios (Collins et al., 1989; Reinmann-Rothmeier & Mandl, 2001). Training group 1 was supposed to find the scenarios more motivating than training group 0 based on the additional support provided by the instructor (Collins et al., 1989). This hypothesis was not supported by the data. It seems likely that with easier scenarios, the students did not feel as motivated anymore to further work on the anesthesiology content. In regard to the motivational effect of the scenarios, a more adequate difficulty level might have detrimental motivational effects as group 0 reported the highest motivation due to the scenarios, when at the same time they found them least adequate for their level of knowledge. It thus seems very likely that the experience of scenarios that were too difficult for the participants' level of knowledge resulted in more motivation to further work on the subject. Specific thought probably has to be given to the underlying concept of motivation. Whereas constructivist learning approaches seek to improve the students' intrinsic motivation, the medical students in this sample might have experienced more extrinsic motivation after having participated in scenarios they perceived as very difficult. They might have experienced this extrinsic motivation in anticipation of their performance in the upcoming exam. As can be deducted from other research in this area (e.g., Moust & Schmidt, 1995), passing exams is regarded as an important motivational factor. Within the setting of a mandatory medical school curriculum that requires the students to pass exams in a high frequency, the constructivist approach of fostering intrinsic motivation through the provision of authentic tasks might thus not work because the extrinsic situational motivation factors are so strong. Nevertheless, the students' intrinsic motivation should also tried to be fostered in the training design based on theoretical considerations. It has, however, been pointed out in the literature contrasting the two learning theoretical approaches (Reinmann-Rothmeier & Mandl, 2001) that the cognitivist examination culture, that is also found in the medical school setting, does not fit within the constructivist approach. It thus seems possible that for this reason, the expected motivational effects will not be found in this setting.

In *Hypothesis 7,* it was assumed that the training groups would not differ in their ratings of the scenarios' relevancy as the latter aspect was not targeted in the redesign process. This hypothesis could not be supported by the data, as training group 0 perceived the scenarios significantly more relevant than training group 1. One possible explanation for this finding is that in the unstandardized simulation sessions received by group 0, the instructors put more emphasis on the scenarios' relevancy for the students' exams or clinical practice.

Furthermore not supported was *Hypothesis 8* that assumed that training group 2 was overall more satisfied with the full-scale simulation training than groups 1 and 0, respectively. The groups did not differ in their satisfaction with the curriculum. It seems possible that the interventions were not strong enough to exert the expected effect or that the different elements perceived by training group 2 as more positive than by the other groups were outweighed by aspects the other groups were more satisfied with.

5.2.1.2 Student ratings of instructors' teaching skills

In *Hypothesis 9,* it was assumed that the students' reactions to the instructors' teaching skills would reflect the instructor training interventions of the initial didactics session and the refresher session, leading to highest ratings by training group 2, followed by training groups 1 and 0, respectively. The results showed that this hypothesis was supported for four out of the six investigated teaching aspects. The instructor training sessions thus seemed to be specifically successful in changing the instructors' teaching behavior towards better structured lessons, better questioning techniques, more appropriate scaffolding, and better adapting the lesson to the group's needs.

Although visible on the descriptive level, the differences on these aspects did not become significant between training group 0's and 1's ratings, possibly indicating that the instructors did take actions to change their teaching behaviors after receiving the first didactic instruction, but that successful behavior change was only reliably implemented or detectable after the refresher intervention.

No significant change was detected in the instructors' ability to motivate their students as well as in the portrayal of their personal motivation. This finding suggests that the teaching guidelines might not have been specific enough in exemplifying how to better motivate students or that this teaching skill is not as easily implemented as the others. Concerning the portrayal of personal motivation, it is again possible that it was not made clear enough how to portray personal motivation or that the instructors' motivation did not change due to the new teaching approach. Future interventions might have to further emphasize motivational topics and tap the concept of teacher motivation more directly by also using teachers' self-report data to investigate if they would report a change in their motivation but that this change was not detected by the students or if they themselves would also not report a change in their motivation.

Although the use of students as raters for the instructors' teaching skills is quite common (e.g., Mayo et al., 1993), further research should use additional and more objective data sources (cf. also Hartman & Nelson, 1992), such as trained observers, to rate the instructors' skills, employing, for example, behavioral markers.

The found improvements in the teaching skill ratings after the instructor training interventions also emphasize the often-neglected importance of instructor training. Teaching is still too often considered as a skill that solely depends on subject expertise. Other settings, such as school teacher education, faced the same problems but are slowly beginning to devote more attention toward building the necessary teaching skills beyond subject expertise (e.g., Cochran-Smith, 2003). Future studies should thus also continue the investigation of specific teaching skills required in the simulation setting and examine how the instructors can best be trained in these skills.

Based on the findings of this study, it can be concluded that even short didactic interventions for medical instructors lead to changes in their teaching behaviors that are detectable by the students.

5.2.1.3 Overall discussion of the reaction results

Concerning the differential effects expected due to the instructional revision of the curriculum, it needs to be stated that the different training versions did not have the expected effect on the students' reactions. An exception was the scenarios' difficulty level perceived by the students. For this aspect, the hypothesis was partially supported as the training group which received the revised full-scale simulation training including supportive instructional elements based on the cognitive apprenticeship approach with the interspersed part-task training sessions (training group 2) perceived the scenarios' difficulty level as significantly more adequate than the training group which participated in the original version of the training (training group 0).

Beyond the differences that were unexpectedly not found between the groups, the reaction findings fit within the existing body of research. Like in previous research efforts, the students' reactions were generally positive to the simulator and the simulations (e.g., Cleave-Hogg & Morgan, 2002; Freid, 1998; McIvor, 2004), simulation training was found to be valuable (Weller, 2004), a suitable learning tool (Fish & Flanagan, 1996), relevant and realistic (McIvor, 2004), as well as an appropriate application setting for the students' theoretical knowledge (Cleave-Hogg & Morgan, 2002).

From a theoretical point of view, it seems possible that the reactions to the different training types reflect the higher demands constructivist learning environments present for the students participating in the revised simulation curriculum. These higher demands might have led to less favorable reactions. This would explain the findings of the items in which training group 0 gave the best ratings, training group 1 gave significantly worse ratings, and training group 2 did not differ from groups 0 and 1, as the demands experienced by training group 2 were attenuated by the previously received part-task simulation sessions. This pattern was found for three items (2, 6, 3) tapping the concepts of training value, the simulations' motivational content, and the simulations' relevancy, whereas the explanation does not quite seem to fit for the relevancy item.

In the respective literature it has been reported that students do not always accept, that is, react positively, to constructivist learning environments (Perkins & Cunningham, 1992). This might also explain the finding why training group 1 gave the least favorable ratings on so many items: Although it was tried to attenuate this effect by the provision of instructional support in the constructivist learning environment, the demands of the revised simulation sessions might have been the highest for training group 1. For training group 2 they were lower again, based on the instructional support and on the previously received additional part-task training. For this line of argumentation, it would have also been interesting to investigate the students' reactions to the revised seminar. Due to the restrictions in the frequency of measurement, this was unfortunately not possible in the current study.

That a shift towards a more constructivist, specifically a cognitive apprenticeship-type, learning environment did indeed take place is supported by the students' ratings of the instructors' teaching skills. After the instructors had received the training session and the refresher course on teaching, the students perceived the instructors as implementing the teaching behaviors specified in the cognitive apprenticeship approach. The differences in the instructor ratings thus seem to further support the explanation that the simulation sessions based on constructivist principles did not lead to the expected better reaction outcome because of their increased demands on the participants. The found differences in the teaching behavior

might, however, also counter this line of argumentation since the included element of instructional support was aimed at attenuating the demands on the students. Again, it is possible, that the students were only able to profit from this support after they had built their more basic skills. Further studies should thus investigate the different reaction outcomes of purely constructivist simulation training in comparison to constructivist simulation training with instructional support as implemented in this study to clarify these findings. Respective studies investigating this aspect could so far not be identified.
As has been mentioned above, the instructors do not seem to have successfully changed their teaching behaviors after the first training session, only after the refresher session. One explanation is that the instructors were not able to apply the new concepts after the first training session. Another explanation could be based on findings that teaching behaviors are often rooted in strong personal beliefs. The first training session might not have been able to change these beliefs. Watkins (1990), for example, found that instructors have different beliefs about training and that only one set of beliefs, namely, that "training is learning", fostered the instructors' teaching skills development. Only instructors with this belief set actively engaged in learning about learning, were open towards evaluating training programs, and continually reflected on their practice. From a training design point of view, this again shows that it is pivotal for organizations to carefully select instructors, and to establish an appropriate teaching and learning culture (cf. Schaper, Friebe et al., 2003). This is especially important in subject domains, such as medicine, where the need for instructor training is only slowly being recognized (Seropian et al., 2004). Future studies should also investigate the simulation instructors' attitudes towards learning and teaching and evaluate the instructor training to identify its strengths and weaknesses.

From a methodological point of view, it is to be noted that future studies should extend the employed questionnaire and thereby operationalize the investigated concepts by more than the used one or two items per concept, thus increasing the questionnaire's reliability and validity. This could not be accomplished in this first, exploratory study due to time constraints and also due to reasons of acceptance on part of the participants. The current questionnaire primarily investigated aspects that were of interest to the SMEs and that were investigated in other simulation studies. While this is an important aspect that should not be neglected, future studies should try to construct even the reaction part of the questionnaire more theory-oriented. The questionnaire could for example be improved by taking previous research into account which found that participant reactions are a multidimensional construct in which, for example, utility judgments represent an underlying dimension (e.g., Brown, 2005; Holton, 2005; Morgan & Casper, 2000; Tan, Hall, & Boyce, 2003). At the current point of research this is still difficult, however, as the respective research has not yet produced unequivocal results as to the underlying factors of reaction measures (e.g., Alliger et al., 1997; Brown, 2005; Cannon-Bowers et al., 1995; Gordon et al., 2001; Morgan & Casper, 2000; Warr & Bunce, 1995). Brown (2005) furthermore pointed out that is still unclear to what extent reactions are influenced by delivery aspects, content aspects, and trainee characteristics. It is thus also possible that the proposed differences were not found because the questionnaire did not tap the relevant aspects. The further advancement of research in the reaction area seems necessary in the light of recent developments in the theoretical foundation of training reactions in affect research (Brown, 2005). This further theoretical development might also help attenuate the risk of inflated type I errors that comes with the testing of numerous items because it would allow to combine items into subscales and test them with the help of multivariate procedures. The found significant results thus have to be viewed with caution due to the possible inflation of α and corroborated in other studies that either reduce the necessary tests by constructing empirically based scales that can be tested with the help of multivariate procedures or adjust the α-level according to the number of tests. Due to its exploratory nature, it was chosen not to adjust the α-level in this study (Bortz, 1993).

5.2.2 Learning

The following sections summarize and discuss the findings on the learning level. Section 5.2.2.1 presents the self-reported learning findings of the seminar and the simulation sessions, section 5.2.2.2 the instructor-reported learning, and section 5.2.2.3 provides a discussion of the aspects concerning all of the learning measures taken together and in the light of the reaction findings as well.

5.2.2.1 Self-reported learning

The newly constructed anesthesia competency scale's structure was first investigated. The conducted principal components analysis yielded an interpretable four-factor solution for the self-perceived anesthesia competency items, with the factors *diagnostic competencies* (factor 1), *anesthetic competencies and airway management* (factor 2), *drug knowledge and administration* (factor 3), and *communication competencies* (factor 4). These four factors reflect meaningful subskills of the basic competency to conduct anesthesia. The following analyses were thus based on this solution.

In the preliminary scale analysis, the assumed response-shift (*Hypothesis 10*) could not be ascertained. It was thus assumed that the students were able to adequately assess their level of performance in the pretest, most likely because the experience of the preparatory seminar provided them with the necessary information to rate their level of performance. This explanation is further supported by the overall rather low grades the students assigned to their competencies, which adequately reflected their status as beginning students in the area of anesthesiology. In the pretest, they rated their competencies between hardly existent (= 5) and satisfactory (= 3) on a German school grading scale, and in the posttest, between satisfactory (= 3) and mostly existent (= 2).

The ratings of the real pretest data were thus used for the following analyses.

Overall training effect

The hypothesis concerning the learning outcomes stated that training group 2 would show the most improvement in their learning scores, followed by training groups 1 and 0, respectively (*Hypothesis 11*). *Hypothesis 11* was supported by the findings. The overall analysis showed that the groups differed in their self-reported competencies, that the self-reported competencies differed over time, and the effect of time of measurement varied depending on the training the participants received.

To further specify this finding in regard to the two different interventions that were supposed to affect students' learning, *Hypothesis 11* was further differentiated and two different learning outcomes were investigated. The first was the learning due to the seminar, the second was the learning due to the simulation sessions. Learning difference scores were calculated for the seminar's effect and the full-scale simulations' effect.

Learning due to the seminar

Hypothesis 12 assumed that the instructionally revised seminar leads to more learning than the original seminar. The original seminar was a mainly lecture-based seminar that focused on conveying a great amount of factual knowledge. The revised seminar was designed based on an integrated approach to teaching and learning (Reinmann-Rothmeier & Mandl, 2001). It included a problem-based approach to conducting anesthesia and was specifically developed to foster student activity, student motivation, learning goal orientation, readiness, and transfer.

To assess students' learning due to the seminar, and thus test *Hypothesis 12*, the change from the retrospectively assessed pretest 1 ratings, which intended to assess the students' skills before they entered the seminar, to the pretest 2 ratings, which intended to assess the students' skills after the seminar, was investigated.

The conducted analyses showed that the seminar in its revised version did not lead to a significant increase in self-reported learning. *Hypothesis 12* could thus not be supported. A first hint that the revised seminar led to more learning, was, however, found on the descriptive level for the students' *diagnostic competencies* (factor 1) as well as their *drug knowledge and administration competencies* (factor 3).

An explanation for this result can be found on the item level of the questionnaire. The items used to assess the students' *anesthetic competencies and airway management* (factor 2) and *communication competencies* (factor 4) primarily tap competencies that cannot specifically be practiced in a seminar since they involve psychomotor skills (e.g., mask ventilation) as well as cognitive components and communication skills. The used instrument was thus most likely not sensitive enough to detect the change in the seminar.

For future evaluation studies, two areas of improvement seem thus necessary. Firstly, the questionnaire items should be revised and more specifically targeted at the skills that can be practiced and thus acquired in the seminar, and secondly, the cases on paper used in the seminar should more directly address the issue of communication with the patient. Further exercises could also be included in the seminar that deal specifically with communication skills.

From a theoretical point of view, another explanation might be possible for why the expected improvements due to the revised seminar were not found. It has been noted, that a "mixing of different approaches in the form of lectures on the one hand, and independent but co-operative problem-based study on the other, seemed to generate a negative interactive effect. Instead of presenting a complementary mix as intended, the conflicting conceptions of learning underlying lecturing on the one hand, and problem-based learning on the other, seemed to have resulted in divisive competition for students' attention" (Margetson, 2000, p. 299). In the current study this explanation seems for several reasons less likely, however. First of all, the negative interactive effect would have been expected to be noticeable on at least the descriptive level, which wasn't the case, but could of course also have been a chance finding. Secondly, the design of the seminar specifically tried to link the lecture-based presentation section of the seminar with the application-oriented problem-based learning phase of the seminar. It did so by also encouraging student participation in the lecture part of the seminar. Before each new presentation section, the students' prior knowledge was elicited in a group discussion. The results were visualized by the instructor and referred back to during his presentation. Furthermore, the problem-based learning was not conducted in its pure form which involves, for example, self-study phases in between the meetings, instead, the problem-based learning phases were included in the seminar. From a theoretical point of view, it thus seems indeed possible to combine the two approaches, but only if their characteristics are adapted so that the students perceive the two elements not as contrary but as truly complementary. Future studies should try to investigate the students' perceptions in regards to the different instructional elements used in the seminar in more detail.

From a methodological point of view, another cautionary note is necessary: The validity of the results is furthermore limited by the fact that the ratings for pretest 1 were only assessed retrospectively. This had to be done because data could only be collected at two points in time, due to the setting's restrictions. Participants' questions and comments during the data collection indicated that they did not always know how to correctly use the retrospective rating scales based on the written instructions alone. It is thus possible that some participants did not accurately fill in the retrospective pretest.

It can so far merely be concluded that there are only first preliminary pieces of evidence that the instructionally revised seminar led to more learning than the traditional seminar in those competencies that can also be practiced on a cognitive level, namely *diagnostic competencies* and *drug knowledge and administration competencies*.

Learning due to the simulation sessions

Hypothesis 13 stated that the instructionally revised simulation training leads to greater improvements in participants' self-perceived competencies than the original simulation training, with participants in training group 2 showing the greatest improvements, followed by participants in training group 1 and training group 0, respectively.

Training group 0 received an unstructured version of the simulation sessions that was not standardized: Learning goals had not been bindingly set, the scenarios were not sequenced according to their difficulty level, and the instructors pursued their individual way of teaching. Training groups 1 and 2 received an instructionally revised version of the simulation sessions that reflected an integrated approach to teaching and learning. The scenarios were developed based on the learning goals identified in the needs assessment phase, scripts were developed to assure standardization, and the instructional strategy was based on cognitive apprenticeship elements. The instructors were trained to model key tasks, to coach the students, support them where necessary (scaffolding), fade their support, encourage the students to articulate their thought processes, and guide the students' reflection after the completion of the scenario. Training group 2 further received the curriculum elements in a different order. In the simulation part of the curriculum, they first received full-task simulation session a), then received the part-task training sessions, followed by full-task simulation session b). The post-test was again collected after simulation session b) but consequently included the part-task training, which it didn't include for the other two groups.

The results showed that training group 2 reported significantly more learning on every factor than both of the other groups, thus supporting the first assumption of *Hypothesis 13*. Due to the above described design it can, however, only be concluded that this finding shows that instructionally revised full-scale simulation training and part-task simulation training together lead to more learning than the original full-scale simulation training and also to more learning than the instructionally revised full-scale simulation training.

Due to the confounding of training time and intervention type in the design, it cannot be determined if the found effect is merely due to the greater amount of training, or if it can also be attributed to the new sequence of the respective elements that allowed the students to first practice the complete task, then practice relevant subtasks which they could finally integrate again in the last complete-task simulation session. Since group 0 and group 1 did not differ significantly, it seems likely that the amount of training was primarily responsible for the found difference between training group 1 and training group 2. Based on the theoretical elaborations, it seems, however, also likely, that the students in group 1 were not able to fully profit from the revised full-scale simulation sessions because they didn't possess the necessary subskills, such as mask ventilation or intubation skills. Due the interspersed part-task training, training group 2, on the other hand, might have benefited more from the second simulation session that then enabled them to integrate the previously practiced subskills.

The effect of the new sequence on students' learning can in this study thus only be supported from a theoretical point of view and should be tested in future studies.

Contrary to the second assumption contained in *Hypothesis 13*, training group1 did not report significantly more learning than training group 0. These results mean that an effect for the instructional redesign of the full-scale simulation sessions based on an integrated approach to learning could not be detected.

This could have been the case due to several factors. First of all, it cannot completely be ruled out, that the redesign truly is not effective in improving student learning. This might be the case because the simulator and the simulation in themselves provide such a strong experience that the additional instructional changes could not further improve the outcome. Secondly, it is possible, that the changes could not be adequately implemented in the first semester (training group 1) by the instructors. This explanation is supported by the students' ratings of the instructors' teaching skills: A significantly higher rating was on most items only found in the second semester (training group 2) before which the instructors had received a refresher session in addition to the training session received before semester 1 (training group 1). Thirdly, it seems possible that the students were not able to profit from the changes implemented in the first semester. That is, although the new design was supposed to offer more support to the students than the original version, this support might still not have been sufficient due to a lack of basic skills. They might have only been able to adequately profit from the new form of the simulation sessions after they had received more training to build their basic skills, namely the part-task emergency training (training group 2). Unfortunately, this could not be tested in the current design, but finds support in other studies in the literature that showed that constructivist learning environments are often overwhelming for beginning learners (e.g., Jonassen et al., 1993). As mentioned above, it was tried to attenuate this effect by incorporating the cognitive apprenticeship elements, but this might not have been sufficient due to the lack of basic skills.

From a training design point of view, it further needs to be noted that full-scale simulation sessions of only one hour seem a very short time. Especially in the light of the number of participants, which was with eight rather large. This led to the situation that not every student was able to assume every role in the scenarios. Based on these restrictions, even a small increase in the self-reported competencies seems remarkable. This time frame reflects, however, other medical simulation training sessions, in which residents are, for example, taught procedural surgical skills (Blum, Powers, & Sundaresan, 2004) or anesthesia non-technical skills (Yee et al., 2005) with only a one-hour session. The need to study long-term applications of the simulation technology and curricula rather than single applications has already been emphasized (Cooper & Taqueti, 2004). This seems especially important as these short interventions most likely reflect resource constraints rather than training design considerations. In long-term training programs, it could also be investigated if overlearning, that is, providing learners with continued practice after they have already mastered the task, also leads to the positive effects on retention found in other research areas (e.g., Baldwin & Ford, 1988; Driskell & Johnston, 1992).

Comparison of learning gains due to the seminar and due to the simulation sessions

Due to previous findings in the literature that reported advantages of simulation training over lecture-based methods (e.g., Good et al., 1992; Ostergaard et al., 1997; Steadman et al., 2006), the learning gains from the seminar were exploratively contrasted with those of the simulation sessions. The results showed that the simulation sessions led to a greater learning gain for training groups 1 and 2 than the seminar did. The significant difference between the gains for training group 1 was, however, only due to the communication factor (factor 4).

Training group 0 that received both interventions in their original, instructionally not revised, version, however, did not show significantly higher learning gains due to the simulation sessions than due to the seminar.

Concerning the advantage of simulation training in comparison to lecture-based seminars, the data of training group 0 call for the conclusion that students do, contrary to the research findings mentioned above, not learn more by participating in simulation sessions. The question remains if the present finding does not merely reflect a time effect. Since this comparison was not a research question, the design allowed only for the intraindividual comparison of the two different learning gains. It is possible that the first learning gain was larger because at that point the students' learning curve might still have been steeper. The seminar furthermore lasted three hours and each student had the chance to actively participate in all of the exercises. The simulation sessions, on the other hand, only lasted one hour each and the students did not have a chance to assume all of the scenario roles.

The students who participated in the instructionally optimized simulation sessions (training group 1) however, reported higher learning due to the simulation sessions than they did due to an also instructionally revised seminar, specifically in the area of communication. The students thus did seem to profit more from the interactive set-up of the simulation sessions than they did from the seminar. The found significant differences between learning gains for training group 2 are not surprising, and allow only for the tentative interpretation that the students benefit from additional skill-building simulation sessions in between the full-scale simulation sessions. It cannot be concluded, however, that this effect is not merely due to the higher quantity of training. From a methodological point of view, it again has to be remembered that the questionnaire items were targeted more at the simulation sessions than at the seminar, as was explicated above. It can thus not be excluded that this finding was a questionnaire artifact.

Based on the previously reported literature, it remains unclear, however, why the simulation sessions did not lead to consistently more self-reported learning in training groups 0 and 1.

These results can only be considered as tentative, however, since a true control group was missing and since the different methods and interventions were not originally designed for this comparison.

Taking these limitations into consideration, it can overall be concluded that the comparison provides another piece of support for the need to provide simulator training based on sound instructional principles and methods as it can then lead to more self-reported learning than a seminar.

5.2.2.2 Other-reported learning: Instructor ratings of students' group performance

The self-reported learning measures were supplemented by instructor ratings of the students' performance in the simulator during the second full-scale simulation session. On a group level, it was investigated whether the training variation effect could also be detected by ratings of students' performance given by the instructors. It was hypothesized that training group 2 receives the best performance ratings, followed by training groups 1 and 0, respectively (*Hypothesis 14*). Again, *Hypothesis 14* could be partly supported in so far as training group 2 received significantly higher ratings than training group 0 concerning their level of preparedness. The instructor ratings for the three training groups did, however, not differ significantly on the other five aspects of the groups' performance that were assessed: No significant effects could be detected concerning the students' level of participation, their motivation, the homogeneity of their proficiency level, their ability to approach the patient, or their cooperation. For the aspects of the groups' active participation and intragroup cooperation, however, the results showed the expected tendency at least on the descriptive level. On every item, training group 2 furthermore received the best ratings by the instructors.

These results provide additional support for the curriculum changes implemented for training group 2, indicating that the improved seminar and the part-scale simulation sessions prepared the students better for the second full-scale simulation session. Again, it cannot empirically be determined if these findings do not merely reflect the higher amount of training received by training group 2 before they entered the second full-scale simulation session. Although the ratings for group 1 were better than for group 0 on the descriptive level, this difference did not reach significance, indicating that the improved seminar's and simulation session's effect was not strong enough to lead to significant changes.

The results have to be interpreted with caution, however. Due to procedural restraints, each aspect could only be assessed by one item and the instructors filled out the questionnaire under time pressure, thus limiting the validity of the ratings. Furthermore, the results' interpretability is limited because the ratings could only be collected on the group level. This holistic measure of student performance was thus not very sensitive and the findings might have been further influenced by the distribution of roles the students assumed during the scenarios as not every student could assume each of the roles in the simulation. Research has shown that large amounts of variance are attributed to the interaction between trainee and scenario. This means, that the students' performance can vary substantially between scenarios (Weller et al., 2005), a variance that could not be detected by this measure. Weller and colleagues' (2005) study showed that about 12-15 scenarios are required to reliably rank trainees on their ability to manage simulated crises. The reliability can be increased by increasing the number of scenarios instead of the number of judges. It thus seems very likely that the instructor ratings in the current study were not reliable. Weller and colleagues' study (2005) furthermore showed that the students' self-assessments correlated modestly but significantly with the external judges' ratings. In the present study, the self- and other reported ratings were not correlated because the items were not equivalent and because the instructor ratings could only be collected on the group level.

The ratings are additionally limited by the posttest-only, no control-group, design of the instructor ratings. Although this design does provide a measure of change, it cannot unequivocally be determined whether the observed changes are due to the training or to other factors. Lastly, it has to be noted that the instructors were not blind to the interventions.

Further studies should thus aim at collecting observational data on the individual level in a pre-posttest-design with raters blind to the conditions. This can, however, only be achieved with smaller student groups which are unfortunately not feasible in a natural curriculum setting as the present one.

5.2.2.3 Overall discussion of the learning results

To put all of the learning results in perspective, a few more methodological issues have to be discussed and the learning results have to be viewed in the light of the collected demographic data, as well as the reaction outcomes.

The first methodological issue concerns the use of self-report measures to assess learning. Overall, research on the validity of self-assessments produced inconclusive results. While there are studies that have supported the validity of self-assessments (e.g., Fox & Dinur, 1988), there is not yet enough research on medical students' ability to assess their own performance in the simulator. Existing evidence on physicians' self-ratings indicates, however, that self-ratings are inaccurate and correlate insufficiently with external ratings. This implies that the physicians lack the ability to identify their own deficiencies (Weller et al., 2005). While MacDonald (2003) could show that undergraduate medical students' self-ratings became more accurate with increasing skill expertise, medical personnel's lacking ability to accurately assess their own performance seems to be a robust finding in several other studies (cf. Boud, 1995; Gordon, 1991). These studies also found that increasing experience did not lead to an increase in the self-ratings' accuracy and that poor performers tend to overrate their performance (Weller et al., 2005). The underlying mechanism that was assumed was that self-ratings are dependent on relative stable self-concepts of general ability. It is thus possible, that the medical students were also not able to adequately assess their knowledge and performance. Although it is still possible that they might not have been able to do so on a detailed level, the overall rather low ratings of their own performance seem to provide a first indication that they validly judged their level of competence. The validity of self-reports can, however, be increased by several techniques. Among them are the emphasis of accuracy, self-awareness, and repeated use of the self-report measure (Hartman & Nelson, 1992). Such measures seem especially useful as the self- as well as the other-ratings in this study were further limited by the small number of scenarios the students could participate in (Weller et al., 2005): The instructor ratings were based on only one scenario, the students could integrate their experiences of at least two full-scale simulation scenarios in their self-ratings. Findings in other areas showed that performance varies across scenarios, implying that large numbers of scenarios and long test times are necessary to obtain reliable results of a trainee's performance (Weller et al., 2005). Weller et al. (2005) found that 10-12 scenarios and three to four external raters would be necessary to achieve acceptable levels of reliability for performance ratings in the simulator, because, although their scenarios were of equal difficulty, their trainees' performance nevertheless varied substantially across the three scenarios. These findings corroborated Boulet et al.'s (2003) results who also concluded that a large number of scenarios is required to obtain reliable ratings of undergraduates' performance since their performance varies between scenarios and their performance in one scenario was not a good predictor of their performance in following scenarios. For this study, it can therefore not be excluded that neither the instructors nor the students were able to reliably rate their performance in the simulator, since they did not participate in a sufficient number of scenarios. This aspect should be taken into account in interpreting the findings of this evaluation study but also poses a great challenge for ratings obtained as part of the students' medical school examinations. Future studies should thus employ appropriate measures to increase the validity of self-reports and also use trained raters observing the students' performance in multiple scenarios as another source of information concerning the individuals' skills. Again, this might only be feasible in smaller scale laboratory studies and not in naturalistic settings of mandatory curricula. These small-scale studies might also allow for the collection of another kind of objective learning indicators with the help of the simulator, such as logging lung ventilation data, or drug administration times (see e.g., McIvor, 2004).

The second methodological issue that has to be considered is the lack of a no-intervention control group. To better deal with this limitation the Internal Referencing Strategy (IRS) could have been employed. It incorporates both relevant and nonrelevant (untrained) material into the pre- and posttests (Haccoun & Hamtiaux, 1994). Using this evaluation strategy, training is considered effective, when changes on the trained material are greater than changes on the untrained material. In the light of the missing no-intervention control group, this would have provided further evidence that the detected changes can be attributed to the training and not to other factors, such as maturation or instrumentation. Again, this strategy could not be employed in the current study as the number of items that could be used was very limited due to time constraints.

A third related methodological issue concerns the question of criterion deficiency. Since the number of items that could be used in the self-report questionnaire and even more so in the other-reported questionnaire, was very limited, it is possible that the selected items did not capture all of the relevant aspects. Although this seems unlikely in the light of the conducted needs assessment and the derived learning objectives on which the items were based, future studies should try to extend the needs assessment to determine more learning objectives, as has already been mentioned above, and to also assess the learning objectives by separate subscales instead of only individual items.

A fourth issue pertains to the level of specificity with which the students' competencies were investigated. The aim of this study was the initial assessment of rather broad aspects. Future studies should build on these results and include the investigation of more detailed concepts, for example the assessment of specific diagnostic strategies, and deal with the question with which training designs these strategies can be taught most effectively (cf. Gräsel et al., 1994).

A fifth issue relates to the number of conducted tests on the same items. Due to the possibility of an inflated type I error, the found significant results have to be viewed with caution and corroborated in other studies that either reduce the necessary tests or adjust the α-level according to the number of tests (Bortz, 1993). Due to its exploratory nature, it was chosen not to adjust the α-level in this study (Bortz, 1993).

Putting the learning results in perspective with the students' prior experience data collected in the demographics section of the questionnaire, the above learning findings are limited by the fact that training group 2 had more clinical experience than the other two groups. That is, in training group 2, there were too many participants with experience and too few participants without experience. Although it cannot definitely be excluded that this influenced the results, other research findings suggest that the difference in experience does not exert an influence on the simulation performance. Schmidts and colleagues (1998) specifically investigated if different experiences with medical internships would lead to differences in performance after simulation training. They did not find differences in performance after a simulator course between groups with or without internship experience. Morgan and Cleave-Hogg (2002b) also found that students' clinical experience did not have a predictive value in performance evaluations in the simulator. It is furthermore to be noted that training group 2 did not rate their competencies as best in the pretest, thus also corroborating that the difference in previous experience was not a major influence factor.

When viewing the learning outcomes in the light of the reaction data, further support for the usefulness of the restructured curriculum received by training group 2 is given by the latter group's reaction ratings to the curricular sequence and coherence. They strongly to mostly agreed that the sequence of their curriculum was appropriate, that the elements were well coordinated, that the contents of the part-task simulation training prepared them well for the second complete-task simulation session, and that they profited more from the second full-scale simulation session than from the first one due to the experiences they gained during the part-task training. Training group 2 furthermore rated the difficulty level as most appropriate of all of the groups which provides another indication for the usefulness of the curriculum change with the part-task training sessions in between the full-scale sessions.

Taking these different measures of student learning and the discussed limitations together, it can be stated that the results provided several first indications that the instructional revision of the simulation training program in its second version, with the added instructional support during the simulation sessions based on the cognitive apprenticeship approach and the part-task training interspersed in the full-scale simulation sessions, was most effective. Training group 2 reported a higher increase in their anesthesia competency than the other groups, they learned more in the simulation sessions than in the seminar, and they were seen as better prepared by the instructors.

The findings thus provide a much-needed corroboration of the existing literature on student learning with the help of anesthesia simulations (e.g., Morgan, Cleave-Hogg, McIlroy, & Devitt, 2002)

5.2.3 Transfer

On the third outcome level, behavior (Kirkpatrick, 1994), training transfer was assessed. Since in the current medical school setting it was neither possible nor applicable to measure “transfer to the job”, transfer was assessed as performance after training (Cannon-Bowers et al., 1995). Although ultimately transfer to the job is of interest, it is not as important in this higher education setting, since the students will not immediately work after training but will further stay at the university. Organizational variables that affect transfer can thus neither be influenced nor controlled. Transfer after training was measured within the regular OSCE examinations the students had to complete as part of the mandatory medical school curriculum. In these examinations, students’ anesthesiology skills were assessed with the help of theoretical questions as well as tasks to be performed on a part-scale simulator. The use of simulators for assessment purposes has been validated in a number of studies (Boulet et al., 2003).

It was hypothesized that training group 2 would receive the best OSCE ratings, followed by training groups 1 and 0, respectively (*Hypothesis 15*). This hypothesis was again based on the instructional redesign of the preparatory seminar and the simulation sessions (training 1), as well as on the changes in the curriculum sequence (training group 2).

Hypothesis 15 could be partially supported by the data as training group 2’s OSCE ratings were significantly better than the ratings of both of the other groups. Training group 1 did not differ significantly from training group 0.

This finding thus provides evidence for the superiority of the instructionally-based revision of the training program as implemented for training group 2. Following the revised preparatory seminar, this training program provided students with the chance to first practice the complete task in full-scale simulation session a), followed by part-task simulation sessions in which they could more thoroughly practice specific subtasks, and to then conclude the program with another complete-task full-scale simulation session. The instructional revision of the preparatory seminar and the full-scale simulation sessions alone did not yield the expected improvements. An explanation for this finding might again be, that the students in training group 1 were not able to adequately profit from the revisions because they did not possess the adequate basic skills yet. In the light of the consistently poorer ratings of training group 1, however, it cannot be definitely excluded that these poorer findings were caused by a variable not assessed in the evaluation study. That transfer did take place and an expected difference was found is especially remarkable as transfer often does not occur because the new KSAs are still too fragile and thus likely break down in application settings, such as exams, that are distracting and stressful (Patrick, 2003).

Reviewing the methodology, a few remarks seem warranted. The first concerns the reliability of the transfer ratings. As has been mentioned above, Weller and colleagues' study (2005) showed, that about 12-15 scenarios are needed to reliably rank trainees on their ability to manage simulated crises. The reliability can be increased by increasing the number of scenarios instead of the number of judges. In the light of these findings, the OSCE ratings' reliability seems questionable. Other studies, on the other hand, have found specifically OSCE ratings a reliable and valid measure (Cohen et al., 1997; Donnelly et al., 2000; Sloan et al., 1995).

Secondly, concerning the transfer measure's validity, it has to be noted, that the criterion validity of transfer assessment increases with the number and specificity of criteria used (Ford & Weissbein, 1997). It has been pointed out that the measurement of transfer is methodologically difficult (Baldwin & Ford, 1988; Bergmann & Sonntag, 1999). This difficulty manifests itself in the criterion problem (Baldwin & Ford, 1988). The value of transfer data is often limited due to criterion measures that are deficient or contaminated. This can also not be excluded for the present study. The OSCE ratings were deficient and contaminated in so far as the exact aspects assessed in the self-report questionnaires could not be used in the OSCE since its content could not be controlled as part of the study. The OSCE ratings thus did not contain all of the self-report aspects and also did contain aspects not assessed in the self-report questionnaires. Future studies should thus try to assess transfer with the help of other criteria as well; specifically criteria that are closely linked to the identified learning objectives.

Thirdly, it is again possible, that the better OSCE results of training group 2 were also influenced by the groups' more extensive prior clinical experience. In the light of the above mentioned research findings (Cleave-Hogg & Morgan, 2002; Schmidts, 1998), it seems, however, unlikely that the prior experience was the main determinant of the results.

If the transfer findings are viewed in line with the learning results, the same rank order of the groups emerges as in the posttest: The best posttest competency ratings were attained by training group 2, followed by training groups 0 and 1, respectively. Concerning the inferential results, training group 2 had significantly higher learning gains due to the simulations than both of the other groups. Added to this learning result is now the transfer result that training group 2's mean still differs significantly from the other groups' means in transfer test. Although, based on the design limitations, the argument that training group 2's higher learning might have been due not only to the additional training they received but also due to the revised sequence of curricular elements, remains grounded on theoretical elaborations, the transfer findings seem to provide indirect support for it. It can thus be tentatively concluded that the instructionally improved training in which part-scale simulation sessions prepare the trainees for the more complex full-scale simulation session leads to more learning and that the attained skills remain over time and transfer to new tasks. Again, these results need further validation since the transfer was rated on a different scale and did not include the previously collected self-report data.

Overall, the transfer finding fits well within other research in the simulation area that also found that the skills acquired in simulation training can be transferred to other settings (Nyssen et al., 2002; Wayne et al., 2005), such as the OSCE (e.g., Schmidts, 1998). Other studies at the University of Heidelberg have furthermore shown that problem-based practical skills and communication training in other medical areas leads to better OSCE results than traditional teaching (Jünger et al., 2005). The results of this study also corroborate findings in the constructivist research area that could show that students who had learned in constructivist settings were more likely to transfer their acquired knowledge (Bransford et al., 1990; Cognition and Technology Group at Vanderbilt, 1992).

The transfer finding should be corroborated by future studies that also attend to the following issues.
First of all, future studies should also employ an OSCE pretest with equivalent items to enhance the design and better exclude alternative explanations for the findings.
Secondly, self-report data should also be collected at the transfer point of measurement since previous research (Tziner et al., 1991) has pointed out that self-report measures and observations made by others often do not correlate in the expected magnitude. It can thus not be concluded that the self-report measures would have reflected the OSCE findings. Viewed from the other perspective, future studies should also try to include more objective data to assess learning. These suggestions are also emphasized in the area of CRM training, where it has been suggested to employ a multimeasure methodology that uses all available data sources (observational, archival, etc.) to assess transfer due to the difficulties inherent in the assessment process (Nullmeyer & Spiker, 2003). The multimeasure methodology would also allow to more thoroughly investigate the students' transfer curves. With this study's design it could not lastly be determined how the students' skill level developed after the posttest.
This does however, still not solve the fourth point: Future studies should therefore also further investigate the scoring methods used in the OSCE, as it has been stated recently, that "overall, there is little agreement on which, if any, scoring method is most appropriate for quantifying performance on simulation exercises" (Murray et al., 2004, p. 1085).

Fifthly, transfer curves should be examined in future studies. Baldwin and Ford (1988) pointed out the necessity to also consider the maintenance curves of transfer. This concept was not employed in the present study but should be investigated in future research to answer the question of how long the students keep up their acquired skills.
From a training design point of view, future studies should lastly try to determine if post-training interventions, such as goal setting or self-management training, can further enhance training transfer (e.g., Richman-Hirsch & Mercer, 2001).

5.2.4 Individual characteristics

Whereas traditional instructional approaches have solely dealt with instructional designs and with how to pace and structure material, the shift to more constructivist approaches has provided the insight that the learners themselves are important in the learning process. They choose to learn, they transform the presented material and construct knowledge in their minds. Thus, individual characteristics have more and more been recognized as important determinants of training effectiveness. The review of the medical simulation literature showed, however, that individual learner characteristics and their influence on training outcomes have so far not been investigated. In addition to the evaluation of the training outcomes, this study was thus also designed to explore individual characteristics that have been identified as important in the training literature and to investigate if they are also influential in the simulation training setting and if they can be influenced by corresponding interventions. Overall, five influence variables were examined: training expectations and perceptions and the resulting expectation fulfillment (5.2.4.1), instrumentality expectancy (5.2.4.2), situational goal orientation (5.2.4.3), student readiness (5.2.4.4), and self-efficacy (5.2.4.5). For all of these individual characteristics, except self-efficacy, an intervention effect was assumed and will be reported accordingly. The individual characteristics section closes with an overall discussion (5.2.4.6).

5.2.4.1 Training expectations, perceptions, and expectation fulfillment

5.2.4.1.1 Training expectations

As specified in the training model and as based on findings in the respective training literature (e.g., Cannon-Bowers et al., 1995), it is important that trainees enter the training program with realistic expectations concerning, for example, the used training methods. Realistic expectations help trainees to adapt to what they will encounter in the training and thus attenuate possible disruptions caused by unexpected demands (Cannon-Bowers, Rhodenizer et al., 1998). This seemed especially important in the area of simulation training, as this training method exerts specific didactic demands on the trainees that they might not have previously encountered. The students' training expectations were thus examined in the pretest and it was investigated whether the students' expectations concerning their active participation and application of knowledge in the simulation sessions could be further increased by respective interventions included in the instructionally revised seminar (*Hypothesis 16*). To guide the students' expectations in the revised seminar, the instructor specifically explicated the didactic demands of the simulation sessions and also presented the video anchors so that the students could form an adequate concept of what they would have to do in the simulator.

The results showed that overall, the training expectations concerning the needed active participation and application of knowledge were high to very high. These high expectations were already found for training group 0. On the descriptive level, the hypothesized higher expectations of training groups 1 and 2 due to their participation in the seminar were visible, they could, however, not be asserted inferentially. *Hypothesis 16* could thus not be supported.

Three explanations for this non-significant finding emerge. First of all, it seems possible that training group 0 received information about the demands of the simulation sessions through students who had participated during the previous semester. It is secondly possible, that the questions exerted a strong demand effect on the students due to their wording. The items might have been worded so that students had to agree, even without knowing what to expect from the sessions. Although this demand effect was not mentioned in the pilot test, future studies should check for a possible acquiescence effect by also including control items referring to expectations that will not be realized in the simulation sessions, such as "I expect to watch the instructor perform all of the tasks". Due to time constraints, respective items could not be included in the present study. Thirdly, it is possible, that the chosen intervention in the revised seminar was not strong enough to lead to the expected effect.

5.2.4.1.2 Training perceptions

To determine if the students also perceived the opportunity to actively participate and apply their knowledge during the simulation sessions, items parallel to the expectation items were used in the posttest to assess students' training perceptions in regard to these aspects. Based on the redesign of the simulation sessions, it was assumed that training groups 1 and 2 would have higher perceptions than training group 0 (*Hypothesis 17*).

The conducted tests showed that the groups also had high perceptions regarding their opportunities to actively participate and apply their knowledge. Overall, these perceptions were not quite as high as the expectations formulated in the pretest. The hypothesis that training groups 1 and 2 would perceive higher opportunities than training group 0 could be partially supported as the inferential tests of the total score yielded a significant difference between training group 2 and both of the other groups in the expected direction. Although training group 1 also had higher perceptions than training group 0, this difference was too small to reach significance.

It is possible that these results again reflect the implementation difficulties of the training changes in group 1. This explanation seems likely in the light of the instructor ratings in which most of the items show the same pattern and which also investigated didactic aspects of the simulation sessions. Not excluded can furthermore be the explanation that the significantly higher ratings of the opportunity to actively participate and to apply their knowledge stem from the interspersed part-task emergency training group 2 had received before filling out the posttest. This explanation seems less likely, however, as the instructions explicitly referred to the full-scale simulation sessions.

5.2.4.1.3 Expectation fulfillment

Based on the results of the training expectation and perception ratings, the students' expectation fulfillment score was calculated according to the procedure used by Cannon-Bowers et al. (1995) and Tannenbaum (1991). It was assumed that training groups 1 and 2 show more expectation fulfillment than training group 0 because their expectations were led to be more realistic and, at the same time, the corresponding opportunities were implemented in the revised full-scale simulation sessions (*Hypothesis 18*).

All of the groups had higher total expectation scores than they had perception scores, although the differences were only small. The conducted analyses showed that in the total scores, training group 2 showed the significantly best expectation fulfillment, that is, their expectations were best met by the training program resulting in the smallest difference between their expectations and perceptions. As was seen in the section on their expectations, they also had the descriptively highest expectations together with training group 1. This indicates that although the revised seminar was not able to lead to the anticipated significantly higher training expectations, the revised simulation session fulfilled the expectations better for training group 2. *Hypothesis 18* could thus be partially supported. The same possible explanations consequently apply as in the training perceptions section.

Based on these results, it can be concluded that overall, the trainees' expectations were realistic and could be further improved through respective training design measures.

To check for instrumentation effects, future studies should also include control items concerning training aspects that were not realized in the training program. Due to time constraints, according items could not be included in the present study. Future studies should furthermore also investigate another expectation fulfillment aspect in the questionnaire, namely, training desires. Tannenbaum et al. (1991) operationalized expectation fulfillment as a combination of expectations, desires, and perceptions of the training. The desire aspect could provide important information as to whether the assessed aspects are relevant to the students. Tannenbaum et al. (1991) pointed out that although different trainees have the same expectations concerning training, the valence of these expectations might differ: Two trainees might expect to be actively involved in the training, but whereas the first trainee might desire this active involvement, the other might be afraid of it. When they then have to become active in the training, the first is pleased, the latter is not pleased, but the expectations of both have been met. It is thus possible that more specific effects could not be detected with this questionnaire.

5.2.4.2 Instrumentality expectancy

Among the different motivational theories, the valence-instrumentality-expectancy framework (Lawler & Suttle, 1973; Vroom, 1964) has specifically been recommended for the use in training motivation research (e.g., Colquitt et al., 2000). In this study, only the instrumentality-expectancy part of the framework was operationalized. The redesign of the preparatory seminar also targeted the increase of student motivation: The instructor informed the students of the simulation sessions' instrumentality for the OSCE at the end of the semester as he introduced the learning objectives. It was thus hypothesized that training groups 1 and 2 would be more motivated than training group 0 (*Hypothesis 19*).

The results showed that, overall, the students perceived the simulation sessions as mostly instrumental for the exam, for their future clinical practice, and more instrumental than books or other teaching forms. *Hypothesis 19* could, however, not be supported, as the groups did not differ in their motivation ratings. It can thus not be concluded that the intervention in the seminar led to higher student motivation, as indicated by the students' instrumentality expectancy in regards to the simulation sessions.

From a training design point of view, it is first of all possible that the intervention was not strong enough to influence the students' motivation. The information aimed at improving their motivation was interwoven with the presentation of the learning objectives and not separately visualized. Future studies should thus extend this intervention by, for example, explicitly visualizing it.

The non-significant finding could, however, also have methodological reasons. It might be due to a ceiling effect, since training group 0 already showed high instrumentality perceptions. Future studies should thus also include control items to investigate possible instrumentation effects. From a methodological point of view, it furthermore needs to be pointed out, that due to time constraints, valence ratings of the outcomes were not collected. Future studies should thus also investigate the valence aspect of the expectancy theory. Previous research has shown that individuals who considered training outcomes as desirable had better training outcomes (Mathieu et al., 1992). An early study (Lawler & Suttle, 1973) has shown, however, that weighting the expectancy items by valence items did not increase the predictability of behaviors. Nevertheless, future studies should expand the used ratings to more completely operationalize the expectancy-instrumentality-valence framework. In a university setting, it seems, however, plausible that success in the following exam is a valued outcome since it is the prerequisite to continue the medical school program. This latter argument might also be an explanation why the intervention could not further improve the students' motivation ratings.

5.2.4.3 Situational goal orientation

Another individual characteristics variable that has received wide research attention is goal orientation (Dweck, 1986). Following Cannon-Bowers et al.'s suggestion (1998) students' situational goal orientation (Button et al., 1996) was chosen as a pre-practice condition to be influenced by the revised preparatory seminar. The instructor encouraged the students during the presentation of the learning contract to, for example, suspend disbelief during the simulations and to view the simulations as an opportunity to build their skills instead of trying to demonstrate them or trying to be better than the other students participating in the simulation. It was consequently hypothesized that training groups 1 and 2 are more learning goal oriented and less performance goal oriented than training group 0 (*Hypothesis 20*).

The students' situational goal orientation was investigated with a scale developed by Button (1996). The examination of the scale structure revealed that the situational performance goal orientation factor could be replicated in the present study. The situational learning goal orientation factor could only be partly replicated. Two of its items had to be excluded from the further analyses since they loaded on a third factor.

The investigation of the intervention effect revealed that, overall, the students showed rather little agreement with the items. No difference between the groups could be detected concerning their situational performance goal orientation. On their situational learning goal orientation, the groups did differ, but in the direction contrary to the assumed. Training groups 1 and 2 showed a significantly weaker learning goal orientation than training group 0. *Hypothesis 20* could thus not be supported by the data.

From an intervention point of view, two explanations for these findings seem likely. First of all, it is possible that the intervention at the beginning of the seminar was not strong enough to lead to the desired effect which would explain why the groups did not differ on their performance goal orientation, but not why training groups 1 and 2 had a lower learning goal orientation than group 0. Based on VandeWalle and colleagues' (2001) suggestion, future seminars should thus try to additionally support learning goal orientation by modeling learning goal orientations and by discussing which types of effort are especially useful in the given learning domain.

An explanation for the opposite situational learning goal orientation effect could be a possible interaction between the learning goal intervention and the used transfer anchors. It is possible that the presented transfer anchors might have led to a decrease in students' learning goal orientation because they emphasized the simulation setting's application character more than its skill building character. This explanation becomes less likely, however, if the performance goal orientation values are also taken into consideration, because they would have been expected to increase as an effect of the used transfer anchors.

A more likely explanation for the finding stems from the methodological point of view. First of all, the students seemed to have problems with the scale. This became evident in several comments while they filled out the questionnaires, as well as in the overall rather low agreement with the items when compared to the other levels of agreement with the rest of the questionnaire items. The comments indicated that the students perceived the questions as "very strange". It thus seems likely that the scale did not validly assess the students' situational goal orientation. The reliable and valid assessment of their learning goal orientation seems especially questionable in the light of only three remaining items. Future studies should thus review the adequacy of the instrument for the use with medical students.

From a methodological point of view, it is furthermore possible that the scale did not primarily assess the students' situational goal orientation, but their dispositional goal orientation, on which they might have differed independent of the interventions. Future studies should thus also explicitly assess the students' dispositional goal orientation and include a pretest for the situational components.

Future studies could furthermore use another conceptualization of the goal orientation construct. Because the situational goal orientation construct was investigated with the help of Button's scale (1996), the performance goal construct was not further differentiated into avoiding goal orientation and proving goal orientation as newer research findings would suggest (VandeWalle et al., 2001). It was decided to use Button's scale because it is targeted at learning experiences in general and conceptualized to assess situational goal orientation. VandeWalle's scale (1997) was not deemed appropriate because it is targeted at work experiences and furthermore assesses domain trait goal orientation.

Future studies thus seem warranted that develop and investigate new measures of *situational* goal orientation that are partitioned into the three by now established factors learning orientation, performance approach/prove orientation, and performance avoidance orientation (Heimbeck, Frese, & Sonnentag, 2003; Senko & Harackiewicz, 2005; VandeWalle, 1997; Zweig & Webster, 2004).

5.2.4.4 Student readiness

Among the individual characteristics that have been identified as important influence variables in training settings is also trainee readiness (e.g., Cannon-Bowers, Rhodenizer et al., 1998). Trainee readiness is a broad concept that refers to the extent to which individuals are prepared to participate in training (Holton, 2005). Based on this concept, the instructionally revised preparatory seminar was designed to increase students' readiness for the full-scale simulation sessions: The students were informed of the simulation sessions' goals, of what was expected from them during the sessions, and of how the simulation sessions' content related to the OSCE. This information was conveyed in form of instructor presentation as well as the video anchor.

Based on these changes, it was assumed that training groups 1 and 2 would feel readier for the full-scale simulation training than training group 0 (*Hypothesis 21*).

Hypothesis 21 was supported by the data on the multivariate and univariate level for both factors of *expectations* and *usefulness of experience*. Concerning the usefulness of experience,

the students did, however, not differ on the conducted post-hoc comparisons. The results thus seem to suggest that the interventions applied in the revised seminar were effective in increasing the students' readiness for the training (cf. Cannon-Bowers, Rhodenizer et al., 1998): They knew better what they could expect from the training and what was expected of them. Concerning the usefulness of experience contained in the training, the results seem to suggest that the revised seminar also led to a better trainee readiness but that the intervention was not strong enough to lead to significant differences on the level of pair-wise comparisons.
From a methodological point of view it has to be pointed out, that the number of items was small, with only three and two items per factor. This led to an unsatisfactory reliability of factor 2. Future studies should thus try to extend this scale.
Methodologically, it furthermore seems questionable if the readiness scale and the training expectation scale measured different constructs. As has been mentioned, both concepts are broad and are discussed from different angles in different contexts (e.g., Cannon-Bowers, Rhodenizer et al., 1998; Holton, 2005; Robertson & Downs, 1989). Training expectations have in this study been operationalized as expectations concerning the training method or didactics. This reflects one of the aspects assessed in Cannon-Bowers et al.'s (1995) model. The preparedness scale was based on Holton's (2005) model. It also assessed a form of trainees' expectations, but on a more general level, concerning the training's effects. Future studies should thus further investigate these concepts and examine their construct validity.

5.2.4.5 Self-efficacy

As another individual characteristics variable, the students' task-specific self-efficacy was assessed (Bandura, 1977). The self-efficacy construct contains cognitive, emotional, and motivational aspects (Bandura, 1997) and has been found very influential in the training context (Colquitt et al., 2000; Tannenbaum & Yukl, 1992).
In this study, no hypothesis concerning the training effect on self-efficacy was formulated since the training interventions did not target the students' self-efficacy.
The investigation of the scale specifically developed for this training yielded a two-factor solution that tapped the students' self-efficacy in regard to the *simulation sessions* and their self-efficacy concerning *anesthetic tasks*. The groups did not differ concerning their self-efficacy ratings on these factors.
From a training design point of view, future studies could also try to incorporate interventions improving the students' self-efficacy. Trainers could promote self-efficacy, for example, by means of behavior-role-modeling or verbal persuasion (see Gist & Mitchell, 1992).
From a methodological point of view, the measure could also be differentiated in future studies into self-efficacy magnitude, strength, and generality (cf. Bandura, 1977) to more completely assess the construct.

5.2.4.6 Overall discussion of the individual characteristics results

Overall, five individual difference variables were investigated. Four of them were supposed to be influenced by the instructional revisions in the seminar or the simulation sessions. Of these four, the hypothesized effect could only be ascertained for two variables. First of all, the revisions to the simulation training curriculum were effective in influencing the students' expectations and increasing their expectation fulfillment. The program as implemented in the second semester led to better expectation fulfillment scores. Secondly, the students' preparedness increased due to the revised seminar.

The study thus provided first indications that different individual characteristics identified in the general training literature could be influenced through respective simulation curriculum design changes.

To corroborate and further extend these findings, future studies could have two foci. A first set of studies should also address additional individual variables and antecedents, for example, other narrow concepts such as locus of control, but also wider concepts, such as conscientiousness, as they have been shown to be significant predictors of training motivation and effectiveness (Colquitt et al., 2000; Noe, 1986). A second set of studies should concentrate on fewer aspects but investigate them more thoroughly. That is, the questionnaire should be validated and it should be determined if the chosen operationalizations of the variables could be varied, for example, instrumentality expectancy as a primary operationalization of training motivation. It could furthermore be tried to better differentiate training expectations and trainee readiness from a conceptual point of view. In the same line of argumentation, the interventions should be investigated on a more detailed level and for each variable to be influenced, different interventions should be designed and compared. It is to be expected that these more fine-grained studies would also provide further insights as to why the other assumed effects could not be ascertained in this study.

5.2.5 Relations between the measured variables

In addition to the evaluation of the training outcomes, this study was also designed to exploratively assess possible relations between the measured variables. Based on the findings, the following sections discuss the relations between the individual characteristics variables and training reactions (5.2.5.1), individual characteristics variables and learning (5.2.5.2), individual characteristics, learning, and behavior (5.2.5.3), as well as the relationship between learning and behavior (5.2.5.4), and close with an overall discussion of the investigated relations (5.2.5.5).

5.2.5.1 Individual characteristics variables and training reactions

Based on the findings in the training literature (Cannon-Bowers et al., 1995; Holton, 2005), it was assumed that a set of individual difference variables would influence participants' training reactions. *Hypothesis 22* stated that students' readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy predict their training reaction scores.

Hypothesis 22 was supported by the findings. The conducted analyses indicated that about a third of the variability in participants' reactions can be explained by individual characteristics variables. Trainees who have been better prepared for the training (readiness), that is, know what is expected of them, react more positively to the training. Trainees whose expectations have been fulfilled or exceeded (expectation fulfillment) react more favorably to the training. Trainees who perceive the training as instrumental (instrumentality expectancy) react more favorably to the training, as do trainees who believe that they can succeed in the training (self-efficacy).

These findings fit well within the existing body of research in the general training literature and replicate the findings of Cannon-Bowers et al. (1995). Cannon-Bowers and colleagues found that self-efficacy, expectation fulfillment, and pre-training motivation were positively related to training reactions.

From a training design point of view, these findings imply the need for specific pre-training interventions in the area of simulation curricula. Because, as Cannon-Bowers et al. (1995, p. 161) pointed out, "no matter, how well a training system is designed, it may not be maximally effective" when it doesn't match or take into account the trainees' individual characteristics. Even though medical school programs are not designed to primarily evoke the students' positive reactions, this outcome level is nevertheless gaining importance because it is also included in externally conducted university rankings. The respective pre-training interventions should thus be designed to assure that trainees are prepared for the training content and the goals of the training (readiness), that they form appropriate expectations concerning the methods used in the training and the kind of participation expected from them and that the training meets these expectations (expectation fulfillment), that the trainees understand in which way the KSAs acquired in the training will be beneficial to them (instrumentality expectancy), and that they feel confident that they can succeed in the training (self-efficacy).

Since it cannot be ruled out that part of the correlations are due to answering tendencies, these findings need to be corroborated by further studies.

5.2.5.2 Individual characteristics variables and learning

Based on the training literature (e.g., Cannon-Bowers et al., 1995; Holton, 2005), the next set of relations that was tested in this study was the relationship between the individual characteristics variables and the learning outcome. It was assumed that the collected individual characteristics variables would predict participants' self-reported anesthesia competency learning scores (*Hypotheses 23* and *24*). Two learning outcomes were differentiated: Learning as training posttest scores, as used in most training studies (e.g., Alliger et al., 1997; Cannon-Bowers et al., 1995) as well as learning as the difference score between pre- and posttest ratings.

The conducted analyses showed that overall, about one third of the learning posttest scores were predicted by the individual characteristics variables. Of the six investigated variables, the students' readiness, their expectation fulfillment, and their self-efficacy significantly predicted their self-reported anesthesia competency posttest scores. Students who were readier, whose expectations were more fulfilled or exceeded, and who felt more confident that they would succeed in the training, reported higher posttest anesthesia competencies. *Hypothesis 23* could thus be supported for these variables. These findings fit well into the body of existent research (e.g., Cannon-Bowers et al., 1995; Holton, 2005).

Contrary to the assumption in *Hypothesis 23*, the students' instrumentality expectations and their goal orientation did not predict their anesthesia competency posttest scores.

Concerning the predictive value of the individual characteristics variables for the students' learning gains (*Hypothesis 24*), the conducted analyses revealed that 15% of the variability in the learning gains was predicted by the individual characteristics variables. More specifically, the increase in the students' self-reported anesthesia competency was significantly predicted by the students' expectation fulfillment, readiness, and instrumentality expectancy. The more their expectations were fulfilled or exceeded, the more ready they were, and the more motivated they were, the higher the learning gains on their self-reported anesthesia competencies. *Hypothesis 24* could thus be supported for three out of the six investigated variables. The goal orientation variables and the self-efficacy variable did not have the expected predictive influence.

Reviewing these results for the two learning outcomes together, the students' readiness for the training and the degree to which their expectations were fulfilled thus predicted both, their self-reported posttest performance and their learning gains in the training.

The interesting finding is that students' task-specific self-efficacy significantly predicted their posttest training performance but not their learning gains, that is, their increase in anesthesia competency. It is thus possible that students with high self-efficacy already possessed rather high competencies before the training. This would then imply the possibility that the self-efficacy ratings merely reflected their competence. Research findings that rather consistently report the self-efficacy's positive influence on learning might mostly have operationalized learning as posttest scores only instead of the difference scores (e.g., Cannon-Bowers et al., 1995; Chen et al., 2005; Multon et al., 1991). This argument seems to gain further support from a recent article by Heggestad and Kanfer (2005) that has questioned the validity of self-efficacy ratings as predictors of learning. Based on their studies, the authors could show that self-efficacy ratings were mostly unrelated to current performance when levels of past performance were controlled. They concluded that "in training contexts in which there is direct feedback regarding the level of one's performance, assessments of self-efficacy are not capable of predicting current performance when past performance is included in the model" (p. 95).

Furthermore interesting is that the perceived instrumentality of the training predicted the learning gains but did not predict the posttest performance. This result provides a first clue that students who perceived the training as instrumental to their goals, namely succeeding in the exam, might for example exert more effort in the training and thus learn more. If one only considers the posttest performance as the learning measure, this relationship does not exist. It remains unclear, however, why the findings did not support literature findings that show a positive influence of training motivation on posttest performance (Cannon-Bowers et al., 1995; Colquitt et al., 2000; Holton, 2005). There is however, also research, that found that valence-instrumentality-expectancy measures are not highly predictive of performance (Schwab, Olian-Gottlieb, & Heneman, 1979).

The goal orientation factors were the only two variables that did not significantly predict either one of the learning outcomes. It is possible, that the data did not support the hypotheses because of two methodological problems. First of all, the non-significant findings might be due to the only two-dimensional operationalization of the construct. This seems especially likely in the light of VandeWalle and colleagues' (2001) statement that the reason for the equivocal findings concerning goal orientation might result from an operationalization of the construct in only two dimensions. But even with only two dimensions, the learning goal orientation could have been expected to predict learning. It is thus likely, that there is a second reason why the effect was not found. It might lie in problems the students had with the questions. As has been explicated above, the participants might not have understood the purpose of the goal orientation items or might not have deemed the items relevant. Future studies should thus try to operationalize the construct in a different way.

5.2.5.3 Individual characteristics, learning, and behavior

Another relationship that was explored, was the relationship between the individual characteristics variables, learning, and behavior, that is, the students' transfer performance in the OSCE. Deducted from existing training models (Baldwin & Ford, 1988; Cannon-Bowers et al., 1995; Holton, 2005) was the hypothesis that the individual characteristics variables also influence transfer, but that this relationship is mediated by the students' self-reported posttest scores (*Hypothesis 25*). This hypothesis could not be supported by the data because the individual characteristics variables did not predict behavior, except for self-efficacy. For self-efficacy, however, learning did not function as a significant predictor of the students' OSCE performance anymore.

It seems possible that the operationalization of the transfer measure was too distal a criterion to find the assumed effects of the individual characteristics variables. This seems likely, as both, the individual characteristics variables and the learning ratings were self-report measures, but transfer was not assessed through self-reported data, but through the students'

OSCE performance. Future studies should thus additionally collect self-reported data for the behavior outcome level and investigate if the hypothesized effect is then found.

Considering the specific medical school setting another emerging explanation is that other variables that were not investigated might have exerted a much stronger effect. Among those other variables could have been external motivational variables because of the examination setting's demands. Previous training research does in most cases not seem to have taken place in mandatory settings in which the success in the transfer measure directly influenced the students' possibility to remain in the program.

In the light of the reported literature, it also seems possible that the underlying relationship is not that of a mediated one, but a more complex one that couldn't be detected with the used analyses. Further influences could have been exerted, for example, by the training design itself, thus suggesting moderation relationships (e.g., Tannenbaum et al., 1991). Holton's (2005) revised model suggests that some individual characteristics, such as goal orientation, directly influence the students' learning and through learning also their transfer performance. Additionally, these characteristics influence learning through motivation to improve work through learning, that again also directly influences behavior (cf. also Mathieu et al., 1992). Other possible effects might be effects related to post-training regulation processes (Chen et al., 2005) that further specify the influence of the collected individual variables. These regulation processes seem especially worthwhile investigating as it seems likely that students will engage in further self-directed learning in preparation of the exam.

These elaborations show that the underlying relationships are far from being clearly established even on a conceptual level. It is thus necessary to conduct further studies that specifically examine these underlying relationships with the help of appropriately complex methods, such as path analytic methods.

5.2.5.4 Learning and behavior

The last relationship that was investigated was the relationship between learning and behavior: It was assumed that the self-report posttest scores predict the students' transfer results in the OSCE (*Hypothesis 26*).

This hypothesis could only be supported on a slightly adjusted α level of .055. The finding thus provides a first indication that the competencies acquired in the training could be transferred by the students to the examination setting.

Although the self-reported posttest scores were only able to predict 1% of the variability in the distal transfer measure in the adjusted model, this finding seems nevertheless remarkable due to the measurement criteria used and the inherent criterion deficiency and contamination issues: On the one hand, the behavior criterion was contaminated since it used a different type of ratings, namely ratings by medical school faculty, and also measured content and skills that were not directly assessed within the self-report learning questionnaire. On the other hand, the behavior criterion was deficient since it did not include all of the exact aspects examined in the self-report learning measure.

In the light of these limitations, the finding fits well within the existing body of literature that reported correlations between learning and performance measures (e.g., Cannon-Bowers et al., 1995; Chen et al., 2005). The reported extent of these correlations varies, however: In their meta-analysis, Alliger and Janak (1997) found low correlations between Kirkpatrick's (1994) learning and behavior levels, Colquitt and colleagues (2000) found moderate to large correlations between learning and transfer. This highlights the need to design appropriate evaluation studies for training programs that assess training effectiveness on as many outcome levels as possible.

5.2.5.5 Overall discussion of the relations between the measured variables

Due to the time constraints and design limitations, this study could only provide a first overview of possible influence variables on the outcomes of medical simulation training and selected relationships between them. Other possible relationships were not tested in order not to overdetermine the model. The selection of relationships to be tested was based on findings in the respective literature. It was thus, for example, not tested if the students' reactions to the training predicted their learning, as for example, a recent meta-analysis found no relationship between reactions and other training outcomes (Arthur et al., 2003).

The found results show that individual characteristics variables partly influence students' reactions to the training program, as well as their learning. Their influence on training transfer through learning could not be ascertained in this study. Learning was found, however, to predict transfer on a slightly adjusted α-level.

These findings provide first indications that the simulation training's participants' individual characteristics do affect training outcomes. So far this has not received attention in the area of simulation training. The results thus imply that training researchers and practitioners should first of all include individual characteristics measures in their evaluation efforts and secondly, also try to influence the respective variables through according interventions, as has been detailed in the present study.

From a methodological point of view, it furthermore needs to be noted, that the accuracy of the parameter estimates depends on whether or not the assumptions of the regression analysis have been met. This also includes the assumption that the independent variables have been measured without error (Tabachnick & Fidell, 1996), an assumption that seems likely to have been violated in the present study because the instruments were newly developed and could not be previously tested for their reliability and validity. It is furthermore to be pointed out, that regression solutions are very sensitive to the combinations of variables included in them (Tabachnick & Fidell, 1996). This again highlights the need to further research this field and to investigate different sets of individual characteristics variables. It is furthermore to be noted that the conducted regression analyses can of course only reveal relationships among variables but not causality.

As has been stated in the research question, the primary aim of this study was not the validation of the relationships within the training model. Schaper et al.'s (2003) model strives to include research findings on influence variables but at the same time tries to provide a pragmatic model for training developers. This study thus investigated only parts of the overall possible relations between the different elements contained in the model. Future studies are needed to gain further insight in all of the relationships as this can provide a broader understanding of training effectiveness (Alliger et al., 1997). To investigate the relations between the different influence variables on a more research oriented level, Schaper et al.'s model should be revised based on models such as Colquitt and colleagues' (Colquitt et al., 2000) to clearly specify expected relationships among the model's elements. This would then allow to also test for mediator and moderator relationships with the help of path analytic strategies. With these more advanced and more detailed theoretical propositions and analytic strategies, more complex relationships might have been detected. Future studies could thus also investigate if the assumed influences could not be ascertained in the current study because of different underlying relationships. Chuang and colleagues (2005), for example, found that self-efficacy, correlates with learning partly through the mediation of training motivation. As the different models (Cannon-Bowers et al., 1995; Colquitt et al., 2000; Holton, 2005) proposed in the training literature show, the exact relation between the variables assessed in these models is far from unequivocally established. Many more relationships could have been investigated in the present study and it is likely that existing relationships were also not detected because they were more complex than was assumed in this explorative setting. An example could also be mediating relationships between trait-like and state-like individual variables. Chen and colleagues (2000) found, for example, that state-like individual variables mediated relationships between trait-like individual variables and learning. It should furthermore be investigated if goal orientation functions as a moderator in the sense that it interacts with training conditions (cf. Heimbeck et al., 2003). More experimental studies with a smaller but more detailed focus would be useful for this type of investigation. It further has to be considered, that, as Holton (1996) points out, training models or studies that are not fully specified, that is, that do not incorporate all of the intervening variables, produce results that might be misleading. Only a fully specified model allows to ultimately determine problems with the intervention. These studies would need to assess a large number of additional individual characteristics variables, such as basic personality traits (extraversion, neuroticism, and openness to experience) that have, for example, been identified as predictors for training reactions (Brown, 2005).

In the light of all the different existing training models, however, it seems unlikely to expect a fully determined model in the near future.

5.3 Limitations

The presented findings have to viewed with caution due to a set of limitations the study faced as a result of the applied setting. Some of them have already been mentioned in the respective sections. The following paragraphs will summarize the study's major limitations.

The first set of limitations pertains to the study design. The first and major limitation was the already mentioned study design limitation that led to a confoundation of the amount of training received and the type of intervention received on the learning outcome for training group 2. To accurately assess the effects of the revised seminar and the two kinds of simulation training including the different curricular elements' sequences, more points of measurement would have been needed: A first pretest would have been needed before the students participated in the preparatory seminar. A second posttest would have been needed after the training groups participated in the part-task emergency training. To further investigate student learning and retention, the self-report measure could have also been used again at the transfer measurement point of time. This would have been of special interest because the OSCE questions and tasks could not be parallel to the self-report questionnaire and could thus not be used as a third point of measurement. Due to the design of the study in two phases and the other evaluation efforts at the medical school, these tests could unfortunately not be implemented.

Secondly, the design suffered from the differences in group sizes. The results thus have to be interpreted with caution and should be corroborated by further studies.

Thirdly, a no-intervention control group was missing, which implies that the threats of maturation, history, and testing cannot lastly be excluded.

Fourthly, the instructors could not be blinded to the training group the students belonged to, which might have had an effect on their ratings of the students' performance in the second full-scale simulation session. The medical faculty assessing transfer was, however, blind to the training group the students belonged to.

Another set of limitations concerns the development of the used instruments. Due to the applied nature of this study and the resulting time restrictions, the used instruments had to be developed and used with the same sample of students. A stricter questionnaire development would have started with a larger number of items, administered these items to a randomly selected group of students and then assessed the psychometric qualities of these items using the appropriate factor analytical as well item analytical methods. Only the most reliable and valid items would have been used in the final questionnaires employed in the evaluation study. Due to the time constraints in this applied setting, the number of items per scale was overall rather limited. If the questionnaire is to be used in further evaluation efforts, the scale development needs to be extended to include larger sets of variables and the analysis of further scale properties.

These limitations should be kept in mind when interpreting the results and also when designing future studies in the field of simulation training. Some of them can be more easily solved in laboratory studies, such as different group sizes and preliminary questionnaire development endeavors, than in applied studies that take place within the framework of mandatory medical school settings. The review of these limitations again emphasizes the importance of taking a systems approach to the development, delivery, and evaluation of training. Only within this systems approach can the organizational restrictions be identified and from the beginning tried to be either influenced or coped with.

5.4 General discussion and future perspectives

This study developed a medical simulation training on the basis of a theoretically founded instructional model and also followed the call of several authors (e.g., Bradley, 2006; Issenberg et al., 2005; Morgan & Cleave-Hogg, 2001; Weller, 2004) to provide simulation training evaluations that assess training effects on more outcome levels and employ stricter designs. Although the study suffered from the discussed limitations, it also possessed the following strong points: it provided several points of measurement, assessed the complete cohort of students in three semesters, was conducted in a field setting, and measured the training's effects on three outcome levels, combined self-report and other-report data, and fully embedded simulation training in a curriculum (Kneebone, 2003).

Taken together, the results of this study corroborate findings in the literature showing that simulation training is a viable method to improve students' anesthesia skills and that simulation training can be effectively incorporated in problem-oriented curricula, as has also been recently shown by Winston and Szarek (2005). The study yielded positive findings on the first outcome level, reactions, on which no differences between the different training groups could be ascertained expect for their perception of the scenarios' difficulty level. On the second outcome level, assessing student learning, it could not be ascertained that the revision of the preparatory seminar led to more learning. The learning results furthermore indicated that training group 2 who received the instructionally revised full-scale simulation training interspersed with the part-task skill building training learned more than the other groups. This self-reported learning result was supported by the preparedness ratings of the instructors that showed that training group 2 was better prepared for the second full-scale simulation session than the other groups. On the third evaluation level, transfer, on which the students' OSCE performance was assessed, training group 2 again received better ratings than the other training groups. Concerning the students' individual characteristics, it was shown that their readiness for the training program as well as their expectation fulfillment can be positively influenced by respective intervention measures. Concerning the relations between the assessed variables it was found that the students' readiness, expectation fulfillment, instrumentality expectancy, and self-efficacy predicted their training reactions. It could not be

ascertained that learning mediated the relationship between the students' individual characteristics and their transfer results. The students' self-reported learning results could be shown to predict their transfer ratings on a slightly adjusted α-level.
The results show that the study extended previous research in the comparison of different instructional methods within the simulation framework that were derived from an instructional model based on learning theoretical considerations and specifically adapted to the simulation setting. It furthermore extended previous research in the consideration of individual characteristics variables and their influence on training outcomes.

Based on the presented results and their shortcomings and limitations, several further research suggestions emerge. First of all, the generalizability of these findings to other universities should be tested by collecting data on the basis of this model at different universities. Namely, the used reaction and individual characteristics questionnaire sections should be used in other medical schools to replicate the study and to thus provide data sets that can be used to compare programs and their effects on the respective variables.
Secondly, a line of future research in this specific setting would be the validation of the used self-report and other-report instruments. The self-report questionnaire should be revised based on the presented findings and further scale development procedures undertaken. The OSCE should be further examined and opportunities for tailoring it more closely to the learning objectives and thus the self-report instruments should be investigated.
Thirdly, another line of future research could be composed of two different types of further studies. First of all, to improve theory building in the field of instructionally based simulation research, future studies should look at a smaller section of the model and investigate this section in more detail in a laboratory setting. This would offer the opportunity to investigate more focused interventions and their effects on specified variables. Training motivation could, for example, be operationalized in different ways and different training programs could be designed to determine which interventions lead to which motivational effects; or it could be investigated if the inclusion of video-supported reflection further improves student learning (cf. Byrne et al., 2002). As became evident in the description of the revised seminar, several instructional elements, such as the transfer anchors, were intended to serve multiple purposes, and one purpose was tried to be influenced with more than one instructional element also. Future studies should test these elements separately and thus investigate their effects in a more fine-grained approach. Such studies could also be used to more thoroughly investigate the underlying relationships between the variables specified in the model. Secondly, studies in applied settings could be conducted that further broaden the scope to search for other potential influence variables. A complete research program in the area of basic anesthesia simulation training could thus combine fine-grained laboratory studies that concentrate on select variables with large-scale field studies that investigate overall effects in the field setting.
From a methodological point of view, future studies should fourthly strive to validate the model by using structural equation modeling to test the relationships assumed in the model.

The field of anesthesia simulation research thus still offers vast possibilities to conduct studies that further the understanding of the specific effects simulation trainings have on various training outcomes.

5.5 Conclusion

The previously described project tackled the two major problems of current medical simulation training: Its lack of a sound instructional theoretical basis and its lack of more rigorous evaluation studies. The study provided a first set of design, delivery, and evaluation guidelines as well as results for an instructionally based simulation training curriculum to develop students' basic anesthesia skills. Since training effectiveness varies as a function of the training method, the skills trained, and the criteria used to assess effectiveness (Arthur et al., 2003) this study can only provide a first piece to close the existing gap. Medical practitioners are thus encouraged to use these guidelines and results to evaluate their own simulation training programs. Beyond this basic recommendation, the project contains a set of additional implications for medical practitioners and researchers based on the discussed results.

First of all, on the most general level, the research project showed that interdisciplinary teams consisting of medical school personnel and instructional psychologists are effective in the area of medical simulation training. The combination of their training backgrounds provided the needed synergy to initially redesign the training program. If a cooperation between the two disciplines cannot be established, it seems advisable to at least provide didactic training for the medical instructors to create an awareness of the relevant elements in training design and delivery.

Secondly, the model's usefulness as a general guideline for the development, delivery, and evaluation of simulation-based anesthesiology training programs could be established, even for settings with a number of organizational constraints. The provided theoretical information on learning processes and their influence variables as well as the evaluation information provided the necessary background to redesign the program and evaluate its effectiveness. The model thus provides information that is becoming increasingly important: In the German Higher Education system evaluation efforts are slowly becoming more influential as can be seen in the university rankings that have begun to be published within the last years. In these rankings the students' satisfaction is also rated in addition to more objective criteria, such as publication quotas. This implies that curriculum designers have to invest considerable efforts in gaining a deeper insight into the relationship between training design and evaluation features and the effectiveness of training programs as well as in improving their curricula on all of the specified outcome levels.

Thirdly, the conducted study showed on an organizational level that a planned development process has beneficial outcomes for the involved faculty, as the development process based on the model encourages their involvement and leads to curricular elements that are better matched to each other. A development process based on the model encourages the involved faculty to take the time to reflect on their curriculum and on the department's teaching practices.

Finally, the conducted project emphasized the importance of securing appropriate resources for training programs, their design, and their evaluation. As Salas and Cannon-Bowers (2001) pointed out, "it is important to note that improved training comes at a cost" (p. 472).

Although simulation training seems to be gaining acceptance in healthcare education, it has not yet reached what could be considered the "'tipping point' of widespread adoption" (Cooper & Taqueti, 2004, p. i16). This "tipping point" will only be reached with the help of continuous efforts of grounding simulation training in didactic theory and providing sound evidence for its effectiveness. Based on these efforts simulators can then be "properly" used and fully unfold their educational value as has been called for by Hays and Singer (1989).

6 References

Abrahamson, S., Denson, J. S., & Wolf, R. M. (1969). Effectiveness of a simulator in training anesthesiology residents. *Journal of Medical Education, 44*, 515-519.

Ackermann, P. L., & Kyllonen, P. C. (1991). Trainee characteristics. In J. E. Morrison (Ed.), *Training for performance: principles of applied human learning* (pp. 193-230). West Sussex: John Wiley & Sons.

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: a review of literature on its outcomes and implementation issues. *Academic Medicine, 68*(1), 52-81.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's level of training criteria: Thirty years later. *Personnel Psychology, 42*, 331-342.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1987). Skill acquistion: Compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Arnold, C. N. (2003). Dermatologie Praktikum - Online-Fortbildung in der Medizin. In U. Dittler (Ed.), *E-Learning: Einsatzkonzepte und Erfolgsfaktoren des Lernens mit interaktiven Medien* (2nd ed., pp. 175-190). München: Oldenbourg.

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234-245.

Arthur, W. J., Tubre, T., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology, 23*(3), 275-285.

Ashton, R., Kennedy, C., & Kirkland, L. (2004). A simulated airway management curriculum. *Medical Education, 38*(5), 552-552.

Ausubel, D. P. (1960). The use of advance organisers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*, 267-272.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*, 65-89.

Baldwin, I., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.

Baldwin, T. T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal skills training. *Journal of Applied Psychology, 77*, 147-154.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*(1), 63-105.

Baldwin, T. T., & Magjuka, R. J. (1997). Training as an organizational episode: pretraining influences on training motivation. In K. J. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 99-127). Mahwah, NJ: Erlbaum.

Baldwin, T. T., Magjuka, R. J., & Loher, B. T. (1991). The perils of participation: Effects of choice of training on trainee motivation and learning. *Personnel Psychology, 44*, 51-65.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Banks, J. (1998). Principles of simulation. In J. Banks (Ed.), *Handbook of simulation*. New York: John Wiley & Sons.

Barach, P. (2000). Patient safety curriculum. *Academic Medicine, 75*(5), 551-552.

Barling, J., & Beattie, R. (1983). Self-efficacy beliefs and sales performance. *Journal of Organizational Behavior Management, 5,* 41-51.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bartlett, C. J. (1982). *Teaching scale developed for Division of Behavioral and Social Sciences.* College Park, MD: University of Maryland.

Beaubien, J. M., & Payne, S. C. (1999, 20 -22 April). *Individual goal orientation as a predictor of job and academic performance: A meta-analysis review and integration.* Paper presented at the 14th Annual Meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Bell, B. S., & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology, 87*(3), 497-505.

Belmont, J. M. (1989). Cognitive strategies and strategic learning. The socio-instructional approach. *The American psychologist, 44*(2), 142-148.

Bendorf, M. (2001). *Bedingungen und Mechanismen des Wissenstransfers.* Wiesbaden: Deutscher Universitätsverlag.

Berg, D., Raugi, G., Gladstone, H., Berkley, J., Weghorst, S., Ganter, M., & Turkiyyah, G. (2001). Virtual reality simulators for dermatologic surgery: measuring their validity as a teaching tool. *Dermatol Surg, 27*, 370-374.

Bergmann, B., & Sonntag, K.-H.(2006). Transfer: Die Umsetzung und Generalisierung erworbener Kompetenzen in den Arbeitsalltag. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 179-205). Göttingen: Hogrefe.

Bergmann, B., & Sonntag, K.-H. (1999). Transfer: Die Umsetzung und Generalisierung erworbener Kompetenzen in den Arbeitsalltag. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (2. erweiterte und überarbeitete Auflage, pp. 287-312). Göttingen: Hogrefe.

Berkenstadt, H., Ziv, A., Barsuk, D., Levine, I., Cohen, A., & Vardi, A. (2003). The use of advanced simulation in the training of anesthesiologists to treat chemical warfare casualties. *Anesthesia and Analgesia, 96*(6), 1739-1742.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Blum, M. G., Powers, T. W., & Sundaresan, S. (2004). Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Ann Thorac Surg, 78*(1), 287-291.

Blum, R. H., Raemer, D. B., Carroll, J. S., Sunder, N., Felstein, D. M., & Cooper, J. B. (2004). Crisis resource management training for an anaesthesia faculty: a new approach to continuing education. *Medical Education, 38*(1), 45-55.

Blyth, P., Stott, N. S., & Anderson, I. (2006). Virtual reality simulators in orthopaedic surgery. What do the surgeons think? *Journal of Bone Joint Surgery British Volume, 88-B*(SUPP_II), 320.

Bond, W. F., Deitrick, L. M., Arnold, D. C., Kostenbader, M., Barr, G. C., Kimmel, S. R., et al. (2004). Using Simulation to Instruct Emergency Medicine Residents in Cognitive Forcing Strategies. *Academic Medicine, 79*(5), 438-446.

Borich, G. (1989). *Air Force instructor evaluation enhancement: Effective teaching behaviors and assessment procedures.* Brooks AFB, TX: Training Systems Division.

Borich, G. (1992). *Effective teaching methods*. New York: Merrill.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Heidelberg: Springer.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation*. Berlin: Springer.

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Boulet, J. R. P. D., Murray, D. M. D., Kras, J. M. D., Woodhouse, J. B. N., McAllister, J. M. D., & Ziv, A. M. D. (2003). Reliability and Validity of a Simulation-based Acute Care Skills Assessment for Medical Students and Residents. *Anesthesiology, 99*(6), 1270-1280.

Boyle, K. A., & Klimoski, R. J. (1995). *Toward an understanding of goal orientation in a training context.* Paper presented at the 10th Annual Conference for the Society of Industrial and Organizational Psychology, Orlando, FL.

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*(3), 254-262.

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. J. Spiro (Eds.), *Cognition, Education, Multimedia: Exploring ideas in high technology* (pp. 115-141). Hillsdale, NJ: Lawrence Erlbaum.

Breland, B. T., & Donovan, J. J. (2005). The role of state goal orientation in the goal establishment process. *Human Performance, 18*(1), 23-53.

Brett, J. F., & VandeWalle, D. (1999). Goal orientation and specific goal content as predictors of performance outcomes in a training program. *Journal of Applied Psychology, 84*, 863-873.

Briggs, L. (1977). *Instructional design: Principles and applications*. Englewood Cliffs, NJ: Educational Technology Publications.

Brosius, F. (2002). *SPSS 11*. Bonn: MITP-Verlag.

Brown, A. L., & Palinscar, A. S. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 1*(2), 117-175.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

Brown, K. G. (2005). An Examination of the Structure and Nomological Network of Trainee Reactions: A Closer Look at "Smile Sheets". *Journal of Applied Psychology, 90*(5), 991-1001.

Buerschaper, C., Hofinger, G., & Harms, C. (2003). Problemlösefähigkeiten in der Anästhesie - Anforderungen und Trainingsziele. In T. Manser (Ed.), *Komplexes Handeln in der Anästhesie* (pp. 208-228). Lengerich: Pabst Verlag.

Bühl, A., & Zöfel, P. (2005). *SPSS 12: Einführung in die moderne Datenanalyse unter Windows* (9th ed.). München: Pearson Studium.

Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: a conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes, 67*(1), 26-48.

Byrne, A. J., & Jones, J. G. (1997). Responses to simulated anaesthetic emergencies by anaesthesists with different durations of clinical experience. *British Journal of Anaesthesia, 78*, 553-556.

Byrne, A. J., Sellen, A. J., Jones, J. G., Aitkenhead, A. R., Hussain, S., Gilder, F., et al. (2002). Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: A multicenter study. *Anaesthesia, 57*, 169-182.

Campbell, J. P. (1988). Training design for performance improvement. In J. P. Campbell, R. J. Campbell & Associates (Eds.), *Productivity in organizations* (pp. 177-216). San Francisco: Jossey-Bass.

Cannon-Bowers, J. A., Burns, J., Salas, E., & Pruitt, J. (1998). Advanced technology in decision-making training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 365-374). Washington, D.C.: American Psychological Association.

Cannon-Bowers, J. A., Rhodenizer, L., Salas, E., & Bowers, C. A. (1998). A framework for understanding pre-practice conditions and their impact on learning. *Personnel Psychology, 51*, 291-320.

Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: A model and initial empirical investigation. *Military Psychology, 7*(3), 141-164.

Caplow, J. A. H., Donaldson, J. F., Kardash, C. A., & Hosokawa, M. (1997). Learning in a problem-based medical curriculum: students' conceptions. *Medical Education, 31*, 440-447.

Casey, C. (1996). Incorporating cognitive apprenticeship in multi-media. *Educational Technology Research and Development, 44*(1), 71-84.

Chen, G., Gully, S. M., Whiteman, J.-A., & Kilcullen, R. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology, 85*(6), 835-847.

Chen, G., Thomas, B., & Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology, 90*(5), 827-841.

Chiaburu, D. S., & Tekleab, A. G. (2005). Individual and contextual influences on multiple dimensions of training effectiveness. *Journal of European Industrial Training, 29*(8), 604-626.

Chopra, V. (1996). Anaesthesia simulators. *Baillere's Clinical Anesthesiology: Safety and risk management in anesthesia, 10*(2), 297-315.

Chopra, V. (1998). What can you do with a simulator? In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 39-49). New York: Plenum.

Chopra, V., Gesink, B., De Jong, J., Bovill, J. G., Spierdijk, J., & Brand, R. (1994). Does training on an anesthesia simulator lead to improvement in performance? *British Journal of Anaesthesia, 73*, 293-297.

Chuang, A., Liao, W.-C., & Tai, W.-T. (2005). An investigation of individual and contextual factors influencing training variables. *Social Behavior and Personality, 33*(2), 159-174.

Clark, R. E., & Voogel, A. (1985). Transfer of training principles for instructional design. *Education Communication and Technology Journal, 33*, 113-123.

Cleave-Hogg, D., & Morgan, P. J. (2002). Experiential learning in an anaesthesia simulation centre: analysis of students' comments. *Medical teacher, 24*(1), 23-26.

Cochran-Smith, M. (2003). Learning and unlearning: the education of teacher educators. *Teaching & Teacher Education, 19*(1), 5-28.

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(6), 2-10.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper Series as an example of anchored instruction: Theory, program, description, assessment data. *Educational Psychologist, 27*(3), 291-315.

Cognition and Technology Group at Vanderbilt. (1993a). Designing learning evironments that support thinking: The Jasper Series as a case study. In T. M. Duffy, J. Lowyck, D. Jonassen & T. M. Welsh (Eds.), *Designing environments for constructive learning* (pp. 9-36). Berlin: Springer.

Cognition and Technology Group at Vanderbilt. (1993b). Toward integrated curricula: Possibilities from anchored instruction. In M. Rabinowitz (Ed.), *Cognitive Science: Foundations of instruction* (pp. 33-55).

Cohen, I. T., Shay, J., Norden, J., & Tome, J. (1997). Reliability of an objective structured clinical exam. *Anesthesiology, 87*(3A), A945.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction. Essays in the honour of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678-707.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company.

Cooper, J. B., & Taqueti, V. R. (2004). A brief history of the development of mannequin simulators for clinical education and training. *Quality & Safety in Health Care, 13*(Suppl 1), i11-i18.

Dalley, P., Robinson, B., Weller, J., & Caldwell, C. (2004). The Use of High-Fidelity Human Patient Simulation and the Introduction of New Anesthesia Delivery Systems. *Anesthesia & Analgesia, 99*(6), 1737-1741.

Davies, D., & Jacobs, A. (1985). "Sandwiching" complex interpersonal feedback: Reactions to alternate methods. *Small Group Behavior, 16*, 387-396.

Davies, J. M. (2001). Medical applications of crew resource management. In E. Salas, C. A. Bowers (Eds.), *Improving teamwork in organizations: Applications of resource management training.* (pp. 265-281): US: Lawrence Erlbaum Associates, Publishers.

De Corte, E. (2001). Instructional psychology. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 7569-7573): Elsevier Science Ltd.

Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik, 39*, 223-238.

Deering, S., Poggi, S., Macedonia, C., Gherman, R., & Satin, A. J. (2004). Improving resident competency in the management of shoulder dystocia with simulation training. *Obstetrics and Gynecology, 103*, 1224-1228.

Dennis, K. A., & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. *International Journal of Aviation Psychology, 8*, 261-276.

Devitt, J. H., Kurrek, M., & Cohen, M. (1998). Can a simulator-based evaluation be used to assess anesthesiologists? *Anesthesiology, 89*(3A), A1172.

Devitt, J. H., Kurrek, M. M., Cohen, M. M., & Cleave-Hogg, D. (2001). The validity of performance assessments using simulation. *Anesthesiology, 95*(1), 36-42.

Devitt, J. H., Kurrek, M., Cohen, M. M., Fish, K., Fish, P., Noel, A. G., et al. (1998). Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anesthesia and Analgesia, 86*, 1160-1164.

Diefendorff, J. M. (2004). Examination of the roles of action-state orientation and goal orientation in the goal-setting and performance process. *Human Performance, 17*(4), 375-395.

Dixon, N. M. (1990). The relationship between trainee responses on participation reaction forms and posttest scores. *Human Resource Development Quartely, 1*, 129-137.

Donnelly, M. B., Sloan, D., Plymale, M., & Schwartz, R. (2000). Assessment of residents' interpersonal skills by faculty proctors and standardized patients. *Academic Medicine, 75*, S93-S95.

Driskell, J. E., & Johnston, J. H. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*, 615-622.

Duffy, T. M., & Jonassen, D. H. (1991). Constructivism: New implications for instructional technology? *Educational Technology, 31*(5), 7-12.

Duntemann, G. H. (1994). Principal components analysis. In M. S. Lewis-Beck (Ed.), *Factor Analysis and related techniques* (pp. 157-245). London: SAGE Publications.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1046.

Dweck, C. S., & Leggett, E. L. (1988). A social cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.

Ehrenstein, A., Walker, B., Czerwinski, M., & Feldman, E. (1997). Some fundamentals of training and transfer: Practice benefits are not automatic. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 31-60). Washington, D.C.: American Psychological Association.

Ellis, H. C. (1965). *The transfer of learning*. New York: Mac Millan.

Ellis, S., & Davidi, I. (2005). After-Event Reviews: Drawing Lessons From Successful and Failed Experience. *Journal of Applied Psychology, 90*(5), 857-871.

Euliano, T. Y. (2000). Teaching respiratory physiology: clinical correlation with a human patient simulator. *Journal of Clinical Monitoring and Computing, 16*(5-6), 465-470.

Facteau, J. D., Dobbins, G. H., Russel, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21*, 1-25.

Fischer, F., Gräsel, C., Mandl, H., Gärtner, R., & Scriba, P. C. (1994). *Konzeption und Entwicklung eines fallbasierten, computerunterstützten Lernprogramms in der Medizin (THYROIDEA)* (Forschungsbericht No. 35). München: Ludwig-Maximilians-Universität.

Fish, M. P., & Flanagan, B. (1996). Use of a realistic patient simulator to teach anesthesia to medical students. *Anesthesiology, 85*(3A), A1041.

Fisher, S. L., & Ford, K. J. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.

Flanagan, B., Nestel, D., & Joseph, M. (2004). Making patient safety the focus: Crisis Resource Management in the undergraduate curriculum. *Medical Education, 38*(1), 56-66.

Flechter, G. C. L., McGeorge, P., Flin, R. H., Glavin, R. J., & Maran, N. J. (2002). The role of non-technical skills in anaesthesia: a review of current literature. *British Journal of Anaesthesia, 88*(3), 418-429.

Flexman, R. E., & Stark, E. A. (1987). Training simulators. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1012-1038). New York: Wiley.

Ford, J. K., Quinones, M. A., Sego, D. J., & Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-527.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*(2), 218-233.

Ford, J. K., & Weissbein, D. A. (1997). Transfer of training: An updated review and analysis. *Performance Improvement Quarterly, 10*(2), 22-41.

Forrest, F. C., Taylor, M. A., Postlethwaite, K., & Aspinall, R. (2002). Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists. *British Journal of Anaesthesia, 88*(3), 338-344.

Fowlkes, J. E., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *International Journal of Aviation Psychology, 8*, 209-222.

Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology, 41*(3), 581-592.

Fredericksen, T. R., & White, B. Y. (1989). An approach to training based upon principles of task decomposition. *Acta Psychologica, 71*, 89-146.

Freid, E. B. (1998). Integration of the human patient simulator into the medical student curriculum. In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 15-21). New York: Plenum.

Gaba, D. M. (1992). Improving anesthesiologists' performance by simulating reality. *Anesthesiology, 76*, 4491-4494.

Gaba, D. M. (1998). Research techniques in human performance using realistic simulation. In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 93-102). New York: Plenum.

Gaba, D. M. (2000). Simulation-based crisis resource management training for trauma care. *American Journal of Anesthesiology, 27*(4), 199-200.

Gaba, D. M., & DeAnda, A. (1988). A comprehensive anesthesia simulation environment: Recreating the operating room for research and training. *Anesthesiology, 69*(3), 387-394.

Gaba, D. M., Fish, K. J., & Howard, S. K. (1994). *Zwischenfälle in der Anästhesie: Prävention und Management*. Lübeck: Gustav Fischer Verlag.

Gaba, D. M., Howard, S. K., Fish, K. J., Smith, B. E., & Sowb, Y. A. (2001). Simulation-based training in anesthesia crisis resource managment (ACRM): A decade of experience. *Simulation & Gaming, 32*(2), 175-193.

Gagné, R. M., Briggs, G., & Wagner, W. W. (1992). *Principles of instructional design*. New York: Harcourt Brace Jovanovich.

Garden, A., Robinson, B., Weller, J., Wilson, L., & Crone, D. (2002). Education to address medical error - a role for high fidelity patient simulation. *N Z Med J, 115*, 133-134.

Gerstenmaier, J., & Mandl, H. (2001). *Methodologie und Empirie zum Situierten Lernen* (No. 137). München: Ludwig-Maximilians-Universität.

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & H. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications*. New York: Academic Press.

Girzadas, D. V., Jr., Harwood, R. C., Clay, L., Rzechula, K., & Caris, J. (2006). High-fidelity Simulation: Assessing Resident Competency in Patient Care. *Academic Emergency Medicine, 13*(5), S190-191.

Gist, M. E. (1997). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review, 12*(3), 472-485.

Gist, M. E., Bavetta, A. G., & Stevens, C. K. (1990). Transfer training method: Its influence on skill generalization, skill repetition, and performance level. *Personnel Psychology, 43*, 501-523.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review, 17*, 183-211.

Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74*(6), 884-891.

Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and posttraining intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Psychology, 44*, 837-861.

Goedhart, H., & Hoogstraten, J. (1992). The retrospective pretest and the role of pretest information in evaluative studies. *Psychological Reports, 70*(3), 699-704.

Goettl, B. P., Yadrick, R. M., Connolly-Gomez, C., Regian, W. J., & Shebilske, W. L. (1996). Alternating task modules in isochronal distributed training of complex tasks. *Human Factors, 38*, 330-346.

Goldstein, I. L. (1993). *Training in organizations: Needs assessment, development, and evaluation*. Pacific Grove, CA: Brooks/Cole.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.

Good, M. L. (2003). Patient simulation for training basic and advanced clinical skills. *Medical Education., 37*(1), 14-21.

Good, M. L., Gravenstein, J. S., Mahla, M. E., White, S. E., Banner, M. J., Carovano, R. G., et al. (1992). Can simulation accelerate the learning of basic anesthesia skills by beginning residents? [abstract]. *Anesthesiology, 77*, A1133.

Goodrow, M. S., Rosen, K. R., & Wood, J. (2005). Using cardiovascular and pulmonary simulation to teach undergraduate medical students: Cases from two schools. *Seminars-in-cardiothoracic-and-vascular-anesthesia, 9*(4), 275-289.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*, 387-405.

Gordon, J. A. (2000). The human patient simulator: acceptance and efficacy as a teaching tool for students. The Medical Readiness Trainer Team. *Academic Medicine, 75*(5), 522.

Gordon, J. A., Oriol, N. E., & Cooper, J. B. (2004). Bringing good teaching cases "to life": A simulator-based medical education service. *Academic Medicine, 79*(1), 23-27.

Gordon, J. A., Wilkerson, W. M., Shaffer, D. W., & Armstrong, E. G. (2001). "Practicing" medicine without risk: Students' and educators' responses to high-fidelity patient simulation. *Academic Medicine, 76*(5), 469-472.

Gordon, M. J. (1991). A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine, 66*, 762-769.

Gräsel, C. (1997). *Problemorientiertes Lernen. Strategieanwendung und Gestaltungsmöglichkeiten*. Göttingen: Hogrefe.

Gräsel, C., & Mandl, H. (1993). Förderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. *Unterrichtswissenschaft, 21*, 355-369.

Gräsel, C., Mandl, H., Fischer, M., & Gärtner, R. (1994). Vergebliche Designermüh? Interaktionsangebote in problemorientierten Computerlernprogrammen. *Unterrichtswissenschaft, 22,* 312-333.

Gräsel, C., Prenzel, M., & Mandl, H. (1993). Konstruktionsprozesse beim Bearbeiten eines fallbasierten Computerlernprogramms. In C. Tarnai (Ed.), *Beiträge zur empirischen pädagogischen Forschung* (pp. 55-67). Münster: Waxmann.

Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology, 29*(4), 462-482.

Grevenik, A., & Schaefer, J. (2004). From Resusci-Anne to Sim-Man: The evolution of simulators in medicine. *Critical Care Medicine, 32*(2), S56-57.

Guadagnoli, E., & Velicer, W. F. (1988). Relationship of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265-275.

Haccoun, R., & Hamtiaux, T. (1994). Optimizing knowledge tests for inferring learning acquisition levels in single group training evaluation designs: The internal referencing strategy. *Personnel Psychology, 47*, 593-604.

Hall, R. E., Plant, J., Bands, C., Kang, J., & Hall, C. (2004). A Randomized Controlled Trial Comparing the Efficacy of Training Paramedic Students Endotracheal Intubation on a Patient Simulator vs. Human Subjects. *Academic Emergency Medicine, 11*(5), 492.

Hall, R. E., Plant, J. R., Bands, C. J., Wall, A. R., Kang, J., & Hall, C. A. (2005). Human Patient Simulation Is Effective for Teaching Paramedic Students Endotracheal Intubation. *Academic Emergency Medicine, 12*(9), 850-855.

Hammond, N. (1993). Learning with Hypertext: Problems, principles, and prospects. In C. McKnight, A. Dillon & J. Richardson (Eds.), *Hypertext - A psychological perspective* (pp. 51-70). New York: Ellis Horwood.

Hartman, S. L., & Nelson, M. S. (1992). What we say and what we do: self-reported teaching behavior versus performances in written simulations among medical school faculty. *Academic Medicine, 67*(8), 522-527.

Hayden, S. R., & Panacek, E. A. (1999). Procedural competency in emergency medicine: the current range of resident experience. *Academic Emergency Medicine, 6*, 728-735.

Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: bridging the gap between reality and training*. New York: Springer.

Heggestad, E. D., & Kanfer, R. (2005). The predictive validity of self-efficacy in training performance: Little more than past performance. *Journal of Experimental Psychology: Applied, 11*(2), 84-97.

Heimbeck, D., Frese, M., & Sonnentag, S. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology, 56*(2), 333-361.

Helmreich, R. L., & Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Brookefield: Ashgate.

Helmreich, R. L., & Schäfer, H.-G. (1994). Team performance in the operating room. In M. S. Bogner (Ed.), *Human error in medicine* (pp. 225-253). Hillsdale, NJ: Lawrence Erlbaum.

Helmreich, R. L., & Schäfer, H.-G. (1998). Turning silk purses into sows' ears. In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 1-8). New York: Plenum.

Hendricks, C. C. (2001). Teaching causal reasoning through cognitive apprenticeship: What are results from situated learning? *Journal of Educational Research, 94*(5), 302-311.

Hesketh, B. (1997). W(h)ither dilemmas in training for transfer. *Applied Psychology: An international review, 46*(4), 380-386.

Hesselfeldt, R., Kristensen, M. S., & Rasmussen, L. J. (2005). Evaluation of the airway of the SimMan full-scale patient simulator. *Acta Anaesthesiology Scandinavia, 49*(9), 1339-1345.

Hicks, W. D. (1984). The process of entering training programs and its effects on training outcomes. *Dissertation Abstracts International, 44*, 3564B.

Hicks, W. D., & Klimoski, R. J. (1987). Entry into training programs and its effects on training outcomes: A field experiment. *Academy of Management Journal, 30*, 542-552.

Hilkey, J. H., Wilhelm, C. L., & Horne, A. M. (1982). Comparative effectiveness of videotape pretraining versus no pretraining on selected process and outcome variables in group therapy. *Psychological Reports, 50*, 1151-1159.

Hoiberg, A., & Berry, N. H. (1978). Expectations and perceptions of Navy life. *Organizational Behavior and Human Performance, 21*, 130-145.

Holladay, C. L., & Quiñones, M. A. (2003). Practice Variability and Transfer of Training: The Role of Self-efficacy Generality. *Journal of Applied Psychology, 88*(6), 1094-1103.

Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Holton, E. F. (2005). Holton's evaluation model: New evidence and construct elaborations. *Advances in Developing Human Resources, 7*(1), 37-54.

Holyoak, K. J., & Thargard, P. R. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295-335.

Holzman, R. S., Cooper, J. B., & Gaba, D. M. (1993). Participant responses to realistic simulation training in anaesthesia crisis resource management. *Anesthesiology, 79*, A1112.

Holzmann, R., Cooper, J., Gaba, D., Philip, J., Small, S., & Feinstein, D. (1995). Anesthesiacrisis resource management: Real-life simulation training in operating room crises. *Journal of Clinical Anesthesia, 7*(8), 675-687.

Honebein, P. C., Duffy, T. M., & Fishman, B. J. (1993). Constructivism and the design of learning environments: Context and authentic activities for learning. In T. M. Duffy, J. Lowyck, D. H. Jonassen & T. M. Welsh (Eds.), *Designing environments for constructive learning* (pp. 87-108). Heidelberg: Springer.

Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education, 50*(4), 200-204.

Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education, 53*(4), 207-210.

Hoover, P. B., Steele-Johnson, D., Beauregard, R. S., & Schmidt, A. M. (1999, April). *Learning and performance goal orientation interactions with dynamic task complexity.* Paper presented at the Annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4*(1), 93-106.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*(2), 144-150.

Howard, G. S., Dailey, P. R., & Gulanick, N. A. (1979). The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement, 3*(4), 481-494.

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G., & Sarnquist, F. H. (1992). Anesthesia crisis resoure management training: Teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine, 63*(9), 763-770.

Ilgen, D. R., & Davis, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology, 49*(3), 550-565.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349-371.

Inzana, C. M., Driskell, J. E., Salas, E., & Johnston, J. H. (1996). The effects of preparatory information on enhancing performance under stress. *Journal of Applied Psychology, 81*, 429-435.

Issenberg, S. B., McGaghie, W. C., Hart, I. R., Mayer, J. W., Felner, J. M., Petrusa, E. R., et al. (1999). Simulation technology for health care professional skills training and assessment. *JAMA, 282*(9), 861-866.

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher, 27*(1), 10-28.

Ivancic, K., & Hesketh, B. (2000). Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics, 43*(12), 1966-1984.

Janssen, J., & Laatz, W. (1999). *Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basissystem Version 8 und das Modul extakte Tests*. Heidelberg: Springer.

Jentsch, F., & Bowers, C. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *The International Journal of Aviation Psychology, 8*, 243-260.

Jonassen, D. H., Mayes, T., & McAleese, R. (1993). A manifesto for a constructivist approach to uses of technology in higher education. In T. M. Duffy, J. Lowyck, D. H. Jonassen & T. M. Welsh (Eds.), *Designing environments for constructive learning* (pp. 231-247). Heidelberg: Springer.

Jones, B. F. (1992). Cognitive designs in instruction. In *Encyclopedia of Educational Research* (Vol. 1, pp. 166-177). New York: McMillan.

Jünger, J., Schafer, S., Roth, C., Schellberg, D., Friedman, B.-D., M., & Nikendei, C. (2005). Effects of basic clinical skills training on objective structured clinical examination performance. *Medical Education, 39*(10), 1015-1020.

Kanfer, R. (1991). Motivational theory and industrial and organizational psychology. In D. Dunnette, L. M. Hough & M. Leaetta (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., pp. 75-170). Palo Alto: Consulting Psychologist Press.

Keller, J. M., & Kopp, T. W. (1987). An application of the ACRS model of motivational design. In C. M. Reigeluth (Ed.), *Instructional design theories in action: Lessons illustrating selected theories* (pp. 289-320). Hillsdale, NJ: Lawrence Erlbaum.

Kemp, J. E., Morrison, G. R., & Ross, S. M. (1998). *Designing effective instruction* (2nd ed.). Upper Saddle River, NJ: Merrill.

Kim, J.-O., & Mueller, C. W. (1994). Factor Analysis: Statistical methods and practical issues. In M. S. Lewis-Beck (Ed.), *Factor Analysis and related techniques* (pp. 75-155). London: SAGE Publications.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training and Development, 13*, 3-9.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco: Berret-Koehler.

Klauer, K. J. (1993). Trainingsforschung: Ansätze, Theorien, Ergebnisse. In K. J. Klauer (Ed.), *Kognitives Training* (pp. 15-63). Göttingen: Hogrefe.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Kneebone, R. (2003). Simulation in surgical training: educational issues and practical implications. *Medical Education, 37*, 267-277.

Knuth, R. A., & Cunningham, D. J. (1993). Tools for constructivism. In T. M. Duffy & D. H. Jonassen (Eds.), *Designing environments for constructive learning* (pp. 63-188). Berlin: Springer.

Koele, P., & Hoogstraten, J. (1988). A method for analyzing retrospective pretest/posttest designs: I. Theory. *Bulletin of the Psychonomic Society, 26*(1), 51-54.

Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback: component analysis of a behavioral safety program. *Journal of Applied Psychology, 65*(3), 261-270.

Koonce, J., & Bramble, W. (1998). Personal computer-based flight training devices. *International Journal of Aviation Psychology, 8*, 277-292.

Kox, W. J. (2000). *Die Geschichte der Simulation in der Anästhesiologie* [online]. Available: www.ukrv.de/ch/anaest/simulato.htm

Kozlowski, S. W. J., Gully, S. M., Brown, K. G., Salas, E., Smith, E. M., & Nason, E. R. (2001). Effects of Training Goals and Goal Orientation Traits on Multidimensional Training Outcomes and Performance Adaptability. *Organizational Behavior and Human Decision Processes, 85*(1), 1-31.

Kozlowski, S. W. J., & Salas, E. (1997). A multilevel organizational systems approach for the implementation and transfer of training. In K. J. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 247-287). Mahwah, NJ: Erlbaum.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 331-376). San Francisco: Jossey-Bass.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(3), 311-328.

Kraiger, K., & Jung, K. M. (1996). Linking training objectives to evaluation criteria. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 151-175). Washington, D.C.: American Psychological Association.

Krapp, A., & Weidenmann, B. (2006). Entwicklungsförderliche Gestaltung von Lernprozessen. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 84-107). Göttingen: Hogrefe.

Kurrek, M. M., & Fish, K. J. (1996). Anaesthesia crisis resource management training: an intimidating concept, a rewarding experience. *Canadian Journal of Anaesthesia, 43*, 430-434.

Lake, C. L. (2005). Simulation in cardiothoraic and vascular anesthesia education: Tool or toy? *Seminars-in-cardiothoracic-and-vascular-anesthesia, 9*(4), 265-273.

Laker, D. R. (1990). Dual dimensionality of training transfer. *Human Resource Development, 1*, 209-223.

Lampotang, S., Öhrn, M. A. K., & van Meurs, W. L. (1996). A simulator-based respiratory physiology workshop. *Academic Medicine, 71*, 526-527.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*, 212-247.

Law, L.-C., & Wong, K. M. P. (1996). Expertise und Instructional Design. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung* (pp. 115-147). Opladen: Westdeutscher Verlag.

Lawler, E. E. (1973). *Motivation in work organization*. Monterey, CA: Brooks/Cole.

Lawler, E. E., & Suttle, J. L. (1973). Expectancy theory and job behavior. *Organizational Behavior and Human Performance, 9*, 482-503.

Lee, A. C. (1998). Using simulators for medical students and anesthesia resident education. In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 23-28). New York: Plenum.

Lee, C., Tinsley, C., & Bobko, P. (2003). Cross-cultural variance in goal orientations and their effects. *Applied Psychology, 52*(2), 272-297.

Leifer, M. S., & Newstrom, J. W. (1980). Solving the transfer of training problem. *Training and Development Journal, 34*(2), 42-46.

Leinhardt, G. (1993). On teaching. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 4, pp. 1-54). Hillsdale, NJ: Erlbaum.

Lemke, S. G. (1995). *Transfermanagement*. Göttingen: Verlag für Angewandte Psychologie.

Leplat, J. (1989). Simulation and simulators in training: Some comments. In *Developing skills with information technology*. Chichester: John Wiley & Sons.

Levinson, W., Gordon, G., & Skeff, K. (1990). Retrospective versus actual pre-course self-assessments. *Evaluation & the Health Professions, 13*(4), 445-452.

Liu, E. H., & Lee, T. L. (2005). The teaching of anaesthesia in Singapore. *Annals of the Academy of Medicine, Singapore, 34*(6), 140-142.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal-setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Lowyck, J., & Elen, J. (1991). Wandel in der theoretischen Fundierung des Instruktionsdesigns. *Unterrichtswissenschaft, 19*(3), 218-237.

Loyd, G. E. (2004). A course for anaesthesiology residents using simulation. *Medical Education, 38*(5), 571-572.

Lupien, A. E. (1998). Simulation in nursing anesthesia education. In L. C. Henson & A. C. Lee (Eds.), *Simulators in anesthesiology education* (pp. 29-37). New York: Plenum.

Lussi, C., Grapengeter, M., & Schüttler, J. (1999). Simulatortraining in der Anästhesie: Möglichkeiten und Stellenwert. *Der Anaesthesist, 48*, 433-438.

Lyon, H. D., Healy, J. C., Bell, J. R., O'Donnell, J. E., Shulth, E. K., Wigton, R. S., et al. (1990). *PlanAlyzer. Cases on Hematology*. Hannover: Dartmouth Medical School.

MacDonald, J., Williams, R. G., & Rogers, D. A. (2003). Self-assessment in simulation-based surgical skills training. *The American Journal of Surgery, 185*, 319-322.

Mager, R. F. (1994). *Lernziele und Unterricht*. Weinheim: Beltz.

Mandl, H., Gruber, H., & Renkl, A. (1997). Situiertes Lernen in multimedialen Lernumgebungen. In L. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (pp. 167-178). Weinheim: Beltz.

Margetson, D. B. (2000). Depth of understanding and excellence of practice: the question of wholeness and problem-based learning. *Journal of Evaluation in Clinical Practice, 6*(3), 293-303.

Marshall, R. L., Smith, J. S. M., Gorman, P. J., Krummel, T. M., Haluck, R. S. M. D., & Cooney, R. N. (2001). Use of a Human Patient Simulator in the Development of Resident Trauma Management Skills. *Journal of Trauma Injury Infection & Critical Care July, 51*(1), 17-21.

Martocchio, J. J. (1992). Microcomputer usage as an opportunity: the influence of context in employee training. *Personnel Psychology, 45*, 529-551.

Marx, R. D. (1982). Relapse prevention for managerial training: A model for maintenance of behavior change. *Academy of Management Review, 7*, 433-441.

Mathieu, J. E., & Martineau, J. W. (1997). Individual and situational influences on training motivation. In K. J. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 193-221). Mahwah, NJ: Erlbaum.

Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: implications for training effectiveness. *Personnel Psychology, 46*, 125-147.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1990, April). *A causal model of individual and situational influences on training effectiveness measures.* Paper presented at the 5th annual conference of the Society for Industrial and Organizational Psychology, Miami, FL.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 35*(4), 828-847.

Maturana, H., & Varela, F. J. (1987). *The tree of knowledge*. Boston: Shambhala.

Mayo, P., Donnelly, M. B., Nash, P. P., & Schwartz, R. W. (1993). Student perceptions of tutor effectiveness in a problem-based surgery clerkship. *Teaching and Learning in Medicine, 5*(4), 227-233.

McCloy, R., & Stone, R. (2001). Science, medicine, and the future. *British Medical Journal, 323*, 912-915.

McDonald, J., Flanagan, B., Keast, P., & Gray, T. (2001). *Simulation training for rural health practitioners: A transformative approach* [online]. Available: http://www.nrha.net.au/ nrhapublic/publicdocs /conferences/6thnrhc/mcdonald2paper.htm

McGehee, W., & Thayer, P. W. (1961). *Training in business and industry*. New York: John Wiley & Sons.

McIvor, W. R. (2004). Experience with medical student simulation education. *Critical Care Medicine, 32*(2), S66-69.

Michel, M. S., Knoll, T., Köhrmann, K. U., & Alken, P. (2002). The URO Mentor: development and evaluation of a new computer-based interactive training system for virtual life-like simulation of diagnostic and therapeutic endourological procedures. *BJU International, 89*, 174-177.

Mitchell, T. R. (1982). Motivation: New directions for theory, research, and practice. *Academy of Management Review, 7*, 80-88.

Mobley, W. H., Hand, H. H., Baker, R. L., & Meglina, B. M. (1979). Conceptual and empirical analysis of military recruit training attrition. *Journal of Applied Psychology, 64*, 10-18.

Morgan, P. J., & Cleave-Hogg, D. (2000). Evaluation of medical students' performance using the anesthesia simulator. *Medical Education, 34*, 42-45.

Morgan, P. J., & Cleave-Hogg, D. (2001). Simulation technology: A global survey of education, evaluation, and research applications. *Journal of education in perioperative medicine* [online], *3*(1).

Morgan, P. J., & Cleave-Hogg, D. (2002a). Comparison between medical students' experience, confidence and competence. *Medical Education, 36*(6), 534-539.

Morgan, P. J., & Cleave-Hogg, D. (2002b). Comparison between medical students' experience, confidence and competence. *Journal of education in perioperative medicine* [online], *3*(1).

Morgan, P. J., & Cleave-Hogg, D. (2002c). A worldwide survey of the use of simulation in anesthesia. *Canadian Journal of Anesthesia, 49*(7), 659-662.

Morgan, P. J., Cleave-Hogg, D., Byrick, R. J., & Devitt, J. H. (1998). Performance evaluation using the anesthesia simulator. *Anesthesiology, 89*(3AS), 67A.

Morgan, P. J., Cleave-Hogg, D., DeSousa, S., & Tarshis, J. (2003). Identification of gaps in the achievement of undergraduate anesthesia educational objectives using high-fidelity patient simulation. *Anesthesia & Analgesia., 97*(6), 1690-1694.

Morgan, P. J., Cleave-Hogg, D., Guest, C. B., & Herold, J. (2001). Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Canadian Journal of Anesthesia, 48*(3), 225-233.

Morgan, P. J., Cleave-Hogg, D., McIlroy, J., & Devitt, J. H. (2002). Simulation technology: A comparison of experiential and visual learning for undergraduate medical students. *Anesthesiology, 96*(1), 10-16.

Morgan, R. B., & Casper, W. (2000). Examining the factor structure of participant reactions to training: A multi-dimensional approach. *Human Resource Development Quartely, 11*, 301-317.

Morris, R. W., & Thomas, J. (1976). Simulation in training: Part 1. *Industrial Training International, 11*(3), 66-69.

Morrison, J. E. (1991). *Training for performance: Principles of applied human learning*. Chichester: John Wiley & Sons.

Moust, J. H. C., & Schmidt, H. G. (1995). Facilitating small-group learning: a comparison of student and staff tutors' behavior. *Instructional Science, 22*, 287-301.

Mueller, M. P., Christ, T., Dobrev, D., Nitsche, I., Stehr, S. N., Ravens, U., et al. (2005). Teaching antiarrhythmic therapy and ECG in simulator-based interdisciplinary undergraduate medical education. *British Journal of Anaesthesia*, *95*(3), 300-304.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.

Murray, D. (2005). Clinical simulation: Measuring the efficacy of training. *Current Opinion in Anesthesiology, 18*, 645-648.

Murray, D., Boulet, J., Ziv, A., Woodhouse, J., Kras, J., & McAllister, J. (2002). An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. *Medical Education, 36*(9), 833-841.

Murray, D. J., Boulet, J. R., Kras, J. F., McAllister, J. D., & Cox, T. E. (2005). A simulation-based acute skills performance assessment for anesthesia training. *Anesthesia & Analgesia, 101*(4), 1127-1134.

Murray, D. J., Boulet, J. R., Kras, J. F., Woodhouse, J. A., Cox, T., & McAllister, J. D. (2004). Acute care skills in anesthesia practice: A simulation-based resident performance assessment. *Anesthesiology, 101*(5), 1084-1095.

Murray, W. B., Schneider, A. J., & Robbins, R. (1998). The first three days of residency: an efficient introduction to clinical medicine. *Academic Medicine, 73*, 595-596.

Myers, J., Mayrose, J., Ellis, D., Anand, A., & Kesavadas, T. (2005). Virtual reality simulation model for emergency airway management training. *Academic Emergency Medicine, 12*(8), 794.

Niegemann, H. M. (1995). Zum Einfluß von "modelling" in einer computerunterstützten Lernumgebung: Quasi-experimentelle Untersuchung zur Instruktionsdesign-Theorie. *Unterrichtswissenschaft, 23*(1), 75-87.

Noe, R. A. (1986). Trainees' attributes: neglected influences on training effectiveness. *Academy of Management Review, 13*, 65-78.

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: test of a model. *Personnel Psychology, 39*, 497-523.

Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 12*(3), 239-249.

Nullmeyer, R. T., & Spiker, V. A. (2003). The importance of crew resource management behaviors in mission performance: Implications for training evaluation. *Military Psychology, 15*(1), 77-96.

Nyssen, A.-S., Larbuisson, R., Janssens, M., Pendeville, P., & Mayne, A. (2002). A comparison of the training value of two types of anesthesia simulators: Computer screen-based and mannequin-based simulators. *Anesthesia and Analgesia, 94*(6), 1560-1565.

Öhrn, M. A. K., van Meurs, W. L., & Good, M. (1995). Laboratory classes: replacing animals with a patient simulator. *Anesthesiology, 83*, A128.

Olson, C. L. (1979). Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin, 86*, 1350-1352.

Ostergaard, D., Jensen, P. F., Jacobsen, J., Larsen, P. D., Lippert, A., Lindekaer, A. L., et al. (1997). Does training in a full-scale anesthesia simulator improve residents' performance? *Anesthesiology, 87*(3A), A940.

Owen, H., & Follows, V. (2006). GREAT simulation debriefing. *Medical Education, 40*(5), 488-489.

Owen, H., Follows, V., Reynolds, K. J., Burgess, G., & Plummer, J. (2002). Learning to apply effective cricoid pressure using a part task trainer. *Anaesthesia, 57*(11), 1098-1101.

Patrick, J. (2003). Training. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology. Human factors in transportation* (pp. 397-434). Mahwah, NJ: Lawrence Erlbaum Associates.

Perkins, D. N., & Cunningham, D. J. (1992). In defense of extremism. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 157-165). Hillsdale, NJ: Lawrence Erlbaum.

Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*, 792-802.

Prince, C., & Jentsch, F. (2001). Aviation crew resource management training with low-fidelity devices. In E. Salas & C. A. Bowers (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 147-164). Hillsdale, NJ: Lawrence Erlbaum.

Prophet, W. W., & Boyd, H. A. (1970). *Device-task fidelity and transfer of training: Aircraft cockpit procedures training*. Alexandria, VA: Human Resources Research Organization.

Quinones, M. A. (1995). Pretraining context effects: training assignment as feedback. *Journal of Applied Psychology, 80*(2), 226-238.

Quinones, M. A. (1997). Contextual influences on training effectiveness. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: applications of psychological research* (pp. 177-199). Washington, D.C.: American Psychological Association.

Quinones, M. A., Ford, J. K., Sego, D. J., & Smith, E. M. (1995). The effects of individual and transfer environment characteristics on the opportunity to perform trained tasks. *Training Research Journal, 1*, 29-48.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: Elsevier.

Reigeluth, C. M. (1979). In search of a better way to organize instruction: The elaboration theory. *Journal of Instructional Development, 2*, 8-15.

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 5-29). Mahwah, NJ: Lawrence Erlbaum.

Reigeluth, C. M., & Stein, F. S. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional design theories and models: An overview of their current status* (pp. 335-382). Hillsdale, NJ: Erlbaum.

Reinmann-Rothmeier, G., & Mandl, H. (1999). *Unterrichten und Lernumgebungen gestalten* (Forschungsbericht No. 60). München: Ludwig-Maximilians-Universität.

Reinmann-Rothmeier, G., & Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie* (pp. 603-648). Weinheim: Beltz.

Reinmann-Rothmeier, G., Mandl, H., & Prenzel, M. (1994). *Computerunterstützte Lernumgebungen: Planung, Gestaltung und Bewertung*. Erlangen: Publicis-MCD-Verlag.

Renkl, A. (1997). *Lernen durch Lehren: Zentrale Wirkmechanismen beim kooperativen Lernen*. Wiesbaden: Deutscher Universitätsverlag.

Resnick, L. B. (1987). Learning in school and out. *Educational Reseacher, 16*(9), 13-20.

Resnick, L. B., Williams, S. M., & Hall, M. (1998). Learning organizations for sustainable educational reform. *Daedalus, 127*(4), 89-118.

Revenstorf, D. (1976). *Lehrbuch der Faktorenanalyse*. Stuttgart: Kohlhammer.

Reznek, M., Smith-Coggins, R., Howard, S., Kiran, K., Harter, P., Sowb, Y., et al. (2003). Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine. *Academic Emergency Medicine., 10*(4), 386-389.

Richardson, A. R., & Waller, D. (2005). The effect of feedback training on distance estimation in virtual environments. *Applied Cognitive Psychology, 19*(8), 1089-1108.

Richman-Hirsch, W. L., & Mercer, W. M. (2001). Posttraining interventions to enhance transfer: The moderating effects of work environments. *Human Resource Development Quarterly, 12*(2), 105-119.

Robbins, S. B., Langley, R., Le, H., Davis, D., Lauver, K., & Carlstrom, A. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin, 130*(2), 261-288.

Robertson, I. T., & Downs, S. (1989). Work sample tests of trainability: A meta-analysis. *Journal of Applied Psychology, 74*, 402-410.

Rogers, P. L., Jacob, H., Rashwan, A. S., & Pinsky, M. R. (2001). Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. *Critical Care Medicine., 29*(6), 1268-1273.

Romalho de Assis, C., Sigulem, D., & de Carvalho, W. B. (2005). Simulation of paediatric basic life support on the Internet. *Medical Education, 39*(5), 522-523.

Rost, D. H., & Schermer, F. J. (1989). Diagnostik des Leistungsangsterlebens. *Diagnostica, 35*, 287-314.

Rouillier, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resource Development Quarterly, 4*, 377-390.

Rowe, R., & Cohen, R. A. (2002). An evaluation of a virtual reality airway simulator. *Anesthesia & Analgesia, 95*(1), 62-66.

Salas, E., & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: applications of psychological research* (pp. 249-280). Washington, D.C.: American Psychological Association.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: a decade of progress. *Annual Review of Psychology, 52*, 471-499.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin, 95*(3), 355-386.

Satava, R. M. (2001). Surgical education and surgical simulation. *World Journal of Surgery, 25*, 1484-1489.

Savoldelli, G. L., Naik, V. N., Hamstra, S. J., & Morgan, P. J. (2005). Barriers to use of simulation-based education. *Canadian Journal of Anaesthesia, 52*(9), 944-950.

Savoldelli, G. L., Naik, V. N., Joo, H. S., Houston, P. L., Graham, M., Yee, B., et al. (2006). Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia residents. *Anesthesiology, 104*, 475-481.

Scardamalia, M., Bereiter, C., & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science, 8*, 173-190.

Schaper, N., Friebe, J., & Sonntag, K.-H. (2003). Merkmale unternehmensbezogener Lernkulturen - eine qualitative Studie mit Experten aus der Unternehmenspraxis und der angewandten Forschung. *Wirtschaftspsychologie, 1*, 80-82.

Schaper, N., Schmitz, A. P., Graf, B., & Grube, C. (2003). Gestaltung und Evaluation von simulatorgestützten Trainings in der Anästhesie. In T. Manser (Ed.), *Komplexes Handeln in der Anästhesie*. Lengerich: Pabst Verlag.

Schaper, N., Schmitz, A. P., Grube, C., Graf, B., & Dieckmann, P. (2002). Design and evaluation of simulator-training in Anaesthesia Crisis Resource Management (ACRM). In D. de Waard, K. A. Brookhuis, J. Moraal & A. Toffetti (Eds.), *Human factors in transportation, communication, health and the workplace*. Maastricht: Shaker Publishing.

Schaper, N., & Sonntag, K.-H. (1998). Aufgabenanalysen und arbeitsplatzbezogene Lernprozesse. *Zeitschrift für Arbeitswissenschaft, 52*(3), 132-143.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, H., Van der Arend, A., Kokx, I., & Boon, L. (1995). Peer versus staff tutoring in problem-based learning. *Instructional Science, 22*, 279-285.

Schmidt, H. G., & Moust, H. C. (2000). Processes that shape small-group tutorial learning: A review of research. In D. H. Evenson & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions*. Hillsdale, NJ: Erlbaum.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schmidt, R. A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors, 39*, 509-525.

Schmidts, M., Beran, Henriette, & Lischka, M. (1998). Patient versus Kunststoffpuppe- profitieren Studenten mit Famulaturerfahrung vom systematischen Traning klinischer Grundfertigkeiten? *Zeitschrift für Hochschuldidaktik, 4*.

Schmitz, A. P. (2002). *Evaluation of an ACRM simulator training for anesthesiologists.* Ruprecht-Karls-Universität, Heidelberg.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Schwab, D., Olian-Gottlieb, J., & Heneman, H. I. (1979). Between subjects expectancy theory research: A statistical review of studies predicting effort and performance. *Psychological Bulletin, 86*, 139-147.

Schwartz, C. E., & Sprangers, M. A. G. (2000). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. In C. E. Schwartz & M. A. G. Sprangers (Eds.), *Adaptation to changing health: Response shift in quality-of-life research* (pp. 81-109). Washington, DC: American Psychological Association.

Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology & Health, 19*(1), 51-69.

Schwarzer, R. (1994). Optimistische Kompetenzerwartung: Zur Erfassung einer personellen Bewältigungsressource. *Diagnostica, 49*(2), 105-123.

Schwid, H. A., & O'Donnell, D. (1990). The anesthesia simulator-recorder: a device to train and evaluate anesthesiologists' responses to critical incidents. *Anesthesiology, 72*, 191-197.

Schwid, H. A., Rooke, G. A., Carline, J., Steadman, R. H., Murray, W. B., Olympio, M., et al. (2002). Evaluation of Anesthesia Residents Using Mannequin-based Simulation: A Multiinstitutional Study. *Anesthesiology, 97*(6), 1434-1444.

Senko, C., & Harackiewicz, J. M. (2005). Regulation of achievement goals: The role of competence feedback. *Journal of Educational Psychology, 97*(3), 320-336.

Seropian, M. A., Brown, K., Gavilanes, J. S., & Driggers, B. (2004). An approach to simulation program development. *Journal of Nursing Education, 43*(4), 170-174.

Shapiro, M. J., Morey, J. C., Small, S. D., Langford, V., Kaylor, C. J., Jagminas, L., et al. (2004). Simulation based teamwork training for emergency department staff: does it improve clinical team performance wehn added to an existing didactic teamwork curriculum? *Quality & Safety in Health Care, 13*, 417-421.

Shimoda, O., & Ikuta, Y. (2005). Mentral strain in medical students during simulator training measured by forhead sweating. *Clinical Autonomic Research, 15*(6), 408-410.

Shuell, T. J. (2001). Learning theories and educational paradigms. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 8613-8620): Elsevier Science Ltd.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge: University Press.

Sinz, E. (2005). Simulation-based education for cardiac, thoraric, and vascular anesthesiology. *Seminars-in-cardiothoracic-and-vascular-anesthesia, 9*(4), 291-307.

Sloan, D. A., Donnelly, M. B., Schwartz, R. W., & Strodel, W. E. (1995). The objective clinical structured examination: The new gold standard for evaluating postgraduate clinical performance. *Annals of Surgery, 222*, 735-742.

Smith-Jentsch, K. A., Payne, S. C., Salas, E., & Jentsch, F. G. (1996). Can pretraining experiences explain individual differences in learning? *Journal of Applied Psychology, 81*(1), 110-116.

Smith-Jentsch, K. A., Salas, E., & Baker, D. P. (1996). Training team performance-related assertiveness. *Personnel Psychology, 49*, 909-936.

Smith-Jentsch, K. A., Salas, E., & Brannick, M. T. (2001). To transfer or not to transfer? Investigating the combined effects of trainee characteristics, team leader support, and team climate. *Journal of Applied Psychology, 86*(2), 279-292.

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Reseacher, 18*(9), 8-14.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco: Jossey-Bass.

Sonnentag, S., Niessen, C., & Ohly, S. (2004). Learning at work: Training and Development. *International Review of Industrial and Organizational Psychology, 19*.

Sonntag, H.-G. (2003). Heidelberg als Vorreiter. *Deutsches Ärzteblatt, 7*, 380.

Sonntag, K.-H. (Ed.). (1992). *Personalentwicklung in Organisationen*. Göttingen: Hogrefe.

Sonntag, K.-H. (2006a). Personalentwicklung - ein Feld psychologischer Forschung und Gestaltung. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 17-35). Göttingen: Hogrefe.

Sonntag, K.-H. (2006b). Ermittlung Tätigkeitsbezogener Merkmale: Qualifikationsanforderungen und Voraussetzungen menschlicher Aufgabenbewältigung. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 17-35). Göttingen: Hogrefe.

Sonntag, K.-H., Stegmaier, R., & Schaper, N. (2006). Ermittlung organisationaler Merkmale: Organisationsdiagnose und Lernkultur. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 179-205). Göttingen: Hogrefe.

Sonntag, K.-H. (1997). Real-life tasks and authentic contexts in learning as a potential for transfer. *Applied Psychology: An international review, 46*, 344-349.

Sonntag, K.-H., & Schäfer-Rauser, U. (1993). Selbsteinschätzung beruflicher Kompetenzen bei der Evaluation von Bildungsmaßnahmen. *Zeitschrift für Arbeits- und Organisationspsychologie, 37*(4), 163-171.

Sonntag, K.-H., & Schaper, N. (2001). Wissensorientierte Verfahren zur Personalentwicklung. In H. Schuler (Ed.), *Lehrbuch der Personalpsychologie* (pp. 241-263). Göttingen: Hogrefe.

Spillane, L. L., Spencer, M., & Maddow, C. (2004). Is Simulation a Valuable Teaching Tool or a Waste of Time? *Academic Emergency Medicine, 11*(5), 561.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Knowledge representation, content specification, and the development of skills in situation-specific knowledge assembly: Some constructivist issues as they relate to cognitive flexibility theory and hypertext. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 121-135). Hillsdale, NJ: Lawrence Erlbaum.

Spiro, R. J., & Jehng, J.-C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, Education, Multimedia: Exploring ideas in high technology* (pp. 136-205). Hillsdale, NJ: Lawrence Erlbaum.

Spitzer, D. R. (1984). Why training fails. *Performance and Instruction Journal, 9*, 6-11.

Sprangers, M., & Hoogstraten, J. (1987). Response-style effects, response-shift bias and a bogus-pipeline. *Psychological Reports, 61*(2), 579-585.

Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-postest designs. *Journal of Applied Psychology, 74*(2), 265-272.

St.Pierre, M., Hofinger, G., Buerschaper, C., Grapengeter, M., Harms, C., Breuer, G., et al. (2004). Simulatorgestütztes, modulares Human Factors Training in der Anästhesie. *Der Anaesthesist, 53*(2), 144-152.

Stafford, F. (2005). The significance of de-roling and debriefing in training medical students using simulation to train medical students. *Medical Education, 39*, 1083-1085.

Stark, R., Graf, M., Renkl, A., Gruber, H., & Mandl, H. (1995). Förderung von Handlungskompetenz durch geleitetes Problemlösen und multiple Lernkontexte. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, 27*(4), 289-312.

Steadman, R. H., Coates, W. C., Huang, Y. M., Matevosian, R., Larmon, B. R., McCullough, L., et al. (2006). Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Critical Care Medicine, 34*(1), 151-157.

Steele-Johnson, D., Beauregard, R. S., Hoover, P. B., & Schmidt, A. M. (2000). Goal orientation and task demand effects on motivation, affect, and performance. *Journal of Applied Psychology, 85*(5), 724-738.

Steers, R. M., & Porter, L. W. (1975). *Motivation and work behavior*. New York: McGraw-Hill.

Steininger, K. (1995). *Personalentwicklung für komplexe Mensch-Maschine-Systeme*. Weinheim: Beltz.

Stevens, C. K., & Gist, M. E. (1997). Effects of self-efficacy and goal-orientation on negotiation skill maintenance: What are the mechanisms? *Personnel Psychology, 50*, 955-978.

Stone, D. L., Gueutal, H. G., & McIntosh, B. (1984). The effects of feedback sequence and expertise of the rater on perceived feedback accuracy. *Personnel Psychology, 37*, 487-506.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: HarperCollins.

Tan, G. M., Ti, L. K., Suresh, S., Ho, B. S., & Lee, T. L. (2002). Teaching first-year medical students physiology: does the human patient simulator allow for more effective teaching? *Singapore Medical Journal, 43*(5), 238-242.

Tan, J. A., Hall, R. J., & Boyce, C. (2003). The Role of Employee Reactions in Predicting Training Effectiveness. *Human Resource Development Quarterly, 14*(4), 397-411.

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology, 76*(6), 759-769.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*, 399-441.

Tarshis, J., DeSousa, S., Brown, R., Iglesias, S., & Kohlhammer, M. (2005). Creation, implementation, and evaluation of a FPA simulator CE program. *Canadian Journal of Anesthesiology, 52*(suppl_1), A89.

Thierau, H. Wottawa, H., & Stangel-Meseke, M. (2006). Evaluation von Personalentwicklungsmaßnahmen. In K.-H. Sonntag (Ed.), *Personalentwicklung in Organisationen* (pp. 179-205). Göttingen: Hogrefe.

Thorndike, J. E. (1914). *The psychology of learning*. New York: Teachers College.

Thorndike, J. E. (1927). The law of effect. *American Journal of Psychology, 39*, 212-222.

Thorndike, J. E., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*, 247-261.

Thornton, G., & Cleveland, J. (1990). Developing managerial talent through simulation. *The American Psychologist, 43*, 190-199.

Tracey, J. B., Hinkin, T. R., Tannenbaum, S. I., & Mathieu, J. E. (2001). The influence of individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly, 12*(1), 5-23.

Tracey, J. B., Tannenbaum, S. I., & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology, 80*(2), 239-252.

Twitchell, S., Holton, E. F., & Trott, J. R. J. (2001). Technical training evaluation practices in the United States. *Performance Improvement Quarterly, 13*, 84-109.

Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. *Journal of Occupational Psychology, 64*, 167-177.

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review, 1*, 170-182.

Van Emmerik, M. L., & van Rooij, J. C. G. M. (1999, April 13-15). *Efficient simulator training: Beyond fidelity.* Paper presented at the Proceedings of the 10th International Training Equipment Conference, The Hague.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement, 57*(6), 995-1015.

VandeWalle, D., Brown, S. D., Cron, W. L., & Slocum, J. W. (1999). The influence of goal orientation and self-regulation tactics on sales performance: A longitudinal field test. *Journal of Applied Psychology, 84*, 249-259.

VandeWalle, D., Cron, W. L., & Slocum, J. W., Jr. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology, 86*(4), 629-640.

VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation in the feedback seeking process. *Journal of Applied Psychology, 82*, 390-400.

Vozenilek, J., Huff, J. S., Reznek, M., & Gordon, J. A. (2004). See One, Do One, Teach One: Advanced Technology in Medical Education. *Academic Emergency Medicine, 11*, 1149-1154.

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*(2), 347-375.

Watkins, K. E. (1990). Tacit beliefs of human resource developers: Producing unintended consequences. *Human Resource Development Quartely, 1*, 263-275.

Waung, M., & Jones, D. R. (2005). The effect of feedback packaging on ratee reactions. *Journal of Applied Social Psychology, 35*(8), 1630-1655.

Wayne, D. B., Butter, J., Siddall, V., Fudala, M., Linquist, L., Feinglass, J., et al. (2005). Simulation-based training of internal medicine residents in advanced cardiac life support protocols: A randomized trial. *Teaching and Learning in Medicine, 17*(3), 210-216.

Weller, J. M. (2004). Simulation in undergraduate medical education: bridging the gap between theory and practice. *Medical Education, 38*(1), 32-38.

Weller, J., Wilson, L., & Robinson, B. (2003). Survey of change in practice following simulation-based training in crisis management. *Anaesthesia, 58*, 471-473.

Weller, J. M., Bloch, M., Young, S., Maze, M., Oyesola, S., Wyner, J., et al. (2003). Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *British Journal of Anaesthesia., 90*(1), 43-47.

Weller, J. M., Robinson, B. J., Jolly, B., Watterson, L. M., Joseph, M., Bajenov, S., et al. (2005). Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. *Anaesthesia, 60*, 245-250.

Werner, J. (1997). *Lineare Statistik: Das Allgemeine Lineare Modell*. Weinheim: Beltz.

Werner, J. M., O'Leary-Kelly, A. M., Baldwin, T. T., & Wexley, K. N. (1994). Augmenting behavior modeling training: Testing the effects of pre- and posttraining interventions. *Human Resource Development Quartely, 5*, 169-183.

Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal, 29*(3), 503-520.

Wexley, K. N., & Latham, G. P. (1981). *Developing and training human resources in organizations*. Glenview, IL: Scott Foresman.

Wexley, K. N., & Latham, G. P. (1991). *Developing human resources in organizations*. New York: McGraw-Hill.

Wexley, K. N., & Latham, G. P. (2002). *Developing human resources in organizations*. New York: McGraw-Hill.

Wheaton, G., Rose, A. M., Fingerman, P. W., Karotkin, A. L., & Holding, D. H. (1976). *Evaluation of the effectiveness of training devices: Literature review and preliminary model*. Washington, D.C.: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wilkerson, L. (1992, April). *Identification of skills for the problem-based tutor: Student and faculty perspectives.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Winn, W. (1993). A constructivist critique of the assumptions of instructional design. In T. M. Duffy, J. Lowyck, D. H. Jonassen & T. M. Welsh (Eds.), *Designing environments for constructive learning* (pp. 189-212). Heidelberg: Springer.

Winn, W. (1996). Instructional design and situated learning: paradox or partnership? In H. McLellan (Ed.), *Situated learning perspectives* (pp. 189-212). Englewood Cliffs, NJ: Educational Technology Publications.

Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 677-691.

Winston, I., & Szarek, J. L. (2005). Problem-based learning using a human patient simulator. *Medical Education, 39*(5), 526-527.

Wong, A. K. (2004). Full scale computer simulators in anesthesia training and evaluation. *Canadian Journal of Anaesthesia, 51*(5), 455-464.

Woodruff, S. L., & Cashman, J. F. (1993). Task, domain, and general efficacy: A reexamination of the self-efficacy scale. *Psychological Reports, 72*, 423-432.

Wottawa, H., & Thierau, H. (1998). *Lehrbuch Evaluation*. Bern: Huber.

Yamnill, S., & McLean, G. N. (2001). Theories supporting transfer of training. *Human Resource Development Quartely, 12*(2), 195-208.

Yee, B., Naik, V. N., Joo, H. S., Savoldelli, G. L., Chung, D. Y., Houston, P. L., et al. (2005). Nontechnical skills in anesthesia crisis management with repeated exposure to simulation-based education. *Anesthesiology, 103*(2), 241-248.

Yelon, S. L., & Ford, J. K. (1999). Pursuing a multidimensional model of transfer. *Performance Improvement Quarterly, 12*, 58-78.

Youmans, R. J., & Stone, E. R. (2005). To Thy Own Self Be True: Finding the Utility of Cognitive Information Feedback. *Journal of Behavioral Decision Making, 18*(5), 319-341.

Ziv, A., Ben-David, S., & Ziv, M. (2005). Simulation based medical education: An opportunity to learn from errors. *Medical Teacher, 27*(3), 193-199.

Ziv, A., Wolpe, P. R., Small, S. D., & Glick, S. (2003). Simulation-based medical education: An ethical imperative. *Academic Medicine, 78*, 783-788.

Zweig, D., & Webster, J. (2004). Validation of a multidimensional measure of goal orientation. *Canadian Journal of Behavioural Science, 36*(3), 232-243.

Appendix

Contents

Appendix A: Learning contract

Eine erfolgreiche Zusammenarbeit erfordert ...

von den Lehrenden...	*von den Teilnehmenden...*
➢ Bereitstellung verschiedener Arbeitsformen – Präsentation – Kleingruppenarbeit – Simulation ➢ Übernahme verschiedener Rollen – Präsentation theoretischer Inhalte – Unterstützung der Gruppenarbeit	➢ Aktive Erarbeitung von Lösungen und Wissensanwendung ➢ Sich-Einlassen auf die Arbeitsformen: Nutzen der Lernchance – Simulator als Herausforderung nicht als Überforderung betrachten – Übungsinstrument zur Verbesserung der Fähigkeiten – Fehler sind lernförderlich ➢ Einhaltung der Rahmenbedingungen ⇒ Aktive Mitarbeit

von den Lehrenden...

- **Bereitstellung verschiedener Arbeitsformen:**
  - Nach der **Präsentation** der theoretischen Inhalte durch mich im ersten Teil des Seminars, werden im zweiten Teil heute **Kleingruppenarbeit**sphasen folgen, in denen ihr das theoretische Wissen dann auch gleich anwenden sollt.
  - **Simulation** morgen im Simulator: Aktive Anwendung des heute Gelernten.
- **Übernahme verschiedener Rollen:**
  - **Präsentation theoretischer Inhalte**: Lehrender mehr in der Expertenrolle, vermittelt im heutigen vorbereitenden Seminar das theoretische Grundwissen im Bereich Narkoseführung, das morgen und am Freitag beim Simulator HANS dann praktisch angewendet werden soll.
  - **Unterstützung der Gruppenarbeit:** Lehrender mehr in der Rolle des Tutors, der bei den Gruppenarbeitsphasen im Seminar und bei HANS die Gruppe nur bei Bedarf unterstützt, da es in diesen Phasen primär um die Umsetzung und praktische Anwendung des Gelernten gehen wird.

von den Teilnehmenden...

- **Aktive Erarbeitung von Lösungen und Wissensanwendung:** D.h. keine Präsentation von fertigen Lösung, sondern Erarbeitung durch die Gruppe und in der Gruppe. Anschließend soll das Wissen gefestigt werden durch die Anwendung in verschiedenen Übungssituationen im Seminar und vor allem morgen und am Freitag bei HANS.

Den Simulator HANS werde ich Euch gleich in einem Video vorstellen. Vorab aber einige Anmerkungen dazu:

- Wichtig ist: **Sich-Einlassen auf die Arbeitsformen** und **Nutzen der Lernchance** die der Simulator als Anwendungssituation für Euch bereithält.
  - Im Simulator werdet ihr selbständig Narkoseführung an einer realistischen Patientenpuppe üben können. Für die meisten von Euch wahrscheinlich eine neue Erfahrung, bei der sehr viele Dinge bzw. Tätigkeiten gleichzeitig bedacht und ausgeführt werden müssen. **Betrachtet** diese Situation im **Simulator als Herausforderung nicht als Überforderung**
  - Simulator ist ein **Übungsinstrument zur Verbesserung** bzw. Ausbildung eurer **Fähigkeiten** im Bereich Narkoseführung.
  - Hierbei ist es ganz klar, dass ihr Fehler machen werdet. Diese **Fehler sind lernförderlich.** Und hier im Simulator ohne Folgen, weder bzgl. eurer Noten, noch bzgl. der Patientensicherheit ☺
  - Damit ihr von der Simulation optimal profitieren könnt, ist es wichtig, dass ihr auf die **Einhaltung der Rahmenbedingungen** achtet: d.h. pünktlich sein und sich durch die Hausaufgaben auf den Simulatorunterricht vorbereiten.

Fazit: **Aktive Mitarbeit** ist grundlegendes Element der Einheit Anästhesie und Notfallmanagement.

Appendix B: Instructor's guide

PSYCHOLOGISCHES INSTITUT

KLINIK FÜR ANAESTHESIOLOGIE

Ruprecht-Karls-Universität Heidelberg

Seminar „Klinische Anaesthesiologie“ zur Vorbereitung des Simulatortrainings

Dozentenleitfaden

A. P. Schmitz & N. Schaper

Unterrichtseinstieg
Vorbereitende Information

3'

Begrüßung

Material:
/

Inhalte:	**Methoden:**
• Eigene Vorstellung des Dozenten	• Dozentenpräsentation

Notizen:

Unterrichtseinstieg

Vorbereitende Information

8'

Lernzielklärung

Material:

- Lernzielfolie

Inhalte:	Methoden:
Lernzielpräsentation: • Darlegung der Lernziele o für Seminar o für Anwendungssituation HANS Hinweisen auf die Verknüpfung des Seminars mit der Anwendungssituation HANS	• Dozentenpräsentation

Notizen:

Unterrichtseinstieg

Vorbereitende Information

5'

Lernkontrakt / Lernhaltungsbeeinflussung

Material:

- Folie Lernkontrakt

Inhalte:	**Methoden:**
Lernkontrakt • Erwartungen/Anforderungen an die Studierenden • Rolle der Dozenten (jeweils Seminar und HANS) • kurze Übersicht über den Inhalt des Seminars	• Dozentenpräsentation
Beeinflussung der Lernhaltung • Anmerkungen zur Simulationssituation: o Simulator und Arbeitssituation am Simulator nicht mit der Realität gleichzusetzen. o Gerade da liegt aber auch Stärke des Simulators: geschützter Rahmen, üben, ausprobieren. o Begreifen als Art Experimentierraum, in dem Fehler nicht schwer wiegen, keine Noten auf Leistung im Simulator. o Im Gegenteil: gerade durch Dinge, die nicht gleich perfekt laufen, sehr viel zu lernen. o Fokus legen auf eigenen Lernerfolg und nicht darauf ob andere besser sind. o Auch wenn Situation im Simulator sozial eher unangenehm ist, weil ihr euch abwechselnd gegenseitig zuschaut. o D.h. nutzt die Simulation dazu Fragen zu stellen, Unklarheiten offen zu legen, an euren Fehlern zu lernen und die Vorgänge zu verstehen.	• Erläuterung

Notizen:

Unterrichtseinstieg

Problemorientierter Einstieg

5'

Problemfall: HANS braucht Narkose

Material:

- Videokurzsequenz: HANS braucht Narkose

Inhalte:	Methoden:
• HANS Videokurzsequenz mit Anamneseerhebung (Student/Hiwi als Modell) • Verbalisierung von Leitfragen mit der Gruppe: Was muss ich als Anästhesist nun tun? Z.B. welche Effekte, welche Medikamente, welche Reihenfolge, ...	• Videovorführung • Unterrichtsgespräch

Notizen:

Hauptteil

Medikamentenkunde

20'

Säule 1

Material:

- Folien

Inhalte:	Methoden:
• Darstellung der Inhalte (welche Medikamente, Wirkungen, NW,...)	• Erarbeitung mit dem Plenum, • Strukturierung und Visualisierung durch den Dozenten • Dozentenpräsentation

Notizen:

Hauptteil

Medikamentenkunde

20'

Säule 2

Material:

- Folien

Inhalte:	Methoden:
• Darstellung der Inhalte (welche Medikamente, Wirkungen, NW,...)	• Erarbeitung mit dem Plenum, • Strukturierung und Visualisierung durch den Dozenten • Dozentenpräsentation

Notizen:

Hauptteil

Medikamentenkunde

20'

Säule 3

Material:

- Folien

Inhalte:	**Methoden:**
• Darstellung der Inhalte (welche Medikamente, Wirkungen, NW,...)	• Erarbeitung mit dem Plenum, • Strukturierung und Visualisierung durch den Dozenten • Dozentenpräsentation

Notizen:

Genereller Hinweis:

- Vernetzung mit anderen Fachgebieten deutlich machen:
    - Woher kennt ihr das,
    - wo ist das noch wichtig,
    - wo könnt ihr das Wissen noch anwenden?

Hauptteil

Anwendung

15'

Vorgehen bei der Narkose: Unterstützte Anwendung

Material:

- Folien
- Fall 1: Standardpatient jung und gesund

Inhalte:	Methoden:
• Falldarbietung (Ziel: Integration der bisher vermittelten Inhalte) o Ihr kommt in den OP zu HANS und sollt Narkose machen. Wie geht ihr vor?	• Erarbeitung mit dem Plenum, • Strukturierung und Visualisierung durch den Dozenten
• Erarbeitung einer Anleitung zum Narkoseablauf	• Unterrichtsgespräch • Visualisierung der Schritte

Notizen:

Hauptteil

Anwendung

35'

Eigenständige Anwendung

Material:

- Fälle
- Lösungsfolie

Inhalte:	Methoden:
• Bearbeitung der Patientenfälle o anhand von Leitfragen: z.B. welche Form der Einleitung wählt ihr?	• Gruppenarbeit • Ca. 5 Studenten pro Gruppe = 6 Gruppen
• Präsentation der Lösung jeden Falles im Plenum (nach der Gruppenarbeitsphase) • Diskussion im Plenum	• Präsentation der Lösungen durch je eine Gruppe • Unterrichtsgespräch • bei Zeitmangel Lösung auf Folie präsentieren

Notizen:

Abschluss

Transferunterstützung

18'

Problemfall: HANS bekommt Narkose

Material:

- Videokurzsequenz: HANS bekommt Narkose
- Hausaufgabenfälle
- Skript

Inhalte:	Methoden:
• HANS Videokurzsequenz mit Narkoseeinleitung (Student als Modell)	• Videovorführung
• Zusammenfassung des Videos durch Dozenten o Zusammenfassung der Videosequenz: was wird von den Studenten bei HANS gefordert, o Zusammenfassung der zusätzlichen Fertigkeiten wie z.B. Intubation, die im Notfallkurs geübt wurden/werden	• Unterrichtsgespräch
• Was habt ihr für die Simulation heute erarbeitet, • was fehlt euch noch, • gibt es noch Unklarheiten? (Standardvorgehen bei Narkose, Medikamente, ...)	• Unterrichtsgespräch
• Was sollen die Studenten zur Vorbereitung auf HANS am Mittwoch zu Hause tun? • Hausaufgaben: take-home Fälle bearbeiten, die dann als Grundlage für den Unterricht an HANS dienen werden. • Hinweis auf Notfallmanagement Skript	• Unterrichtsgespräch • Hausaufgabenfälle austeilen • Skript austeilen

Notizen:

Appendix C: Student pretest questionnaire

Klinik für Anaesthesiologie

Ruprecht-Karls-Universität Heidelberg

Fragebogen zur HANS-Simulationseinheit in HeiCuMed

Liebe(r) HeiCuMed Teilnehmer(in),

füllen Sie bitte im Rahmen Ihrer Teilnahme am Modul Notfallmedizin und Anästhesiologie den folgenden Fragebogen aus.

Wir versichern Ihnen hiermit, dass die Daten der Fragebögen **vollkommen vertraulich** behandelt und nicht an Dritte weitergegeben werden. Ihre Angaben werden nicht von den unterrichtenden Tutoren eingesehen und haben keinen Einfluss auf Ihre Bewertung. Damit wir die Daten einander zuordnen können, bitten wir Sie, die nachfolgenden Angaben zu machen.

Name: ______________________

Gruppe: ___

Datum: __.__.__

Bearbeitungshinweise:

- Markieren Sie die Antworten, die am ehesten Ihrer Meinung entsprechen bitte mit einem Kreuz und benutzen Sie die jeweils oben angeführten Skalen zu Ihrer Orientierung.
- Bitte seien Sie bei der Angabe Ihrer Übereinstimmung mit den Aussagen aufrichtig. Es gibt keine „richtigen“, „falschen“ oder „erwünschten“ Antworten.
- Bitte bearbeiten Sie alle Aussagen zügig, in der vorgegebenen Reihenfolge und ohne Fragen bzw. Aussagen auszulassen.
- Sollten Sie während der Bearbeitung noch Fragen haben, wenden Sie sich bitte an die anwesenden HeiCuMed-Tutoren.

A. Schmitz, N. Schaper, C. Grube, B. Graf, T. Boeker

Teil A

Beurteilen Sie im Folgenden bitte Ihre eigenen Fähigkeiten im Bereich Anästhesie.

Bitte geben Sie hierbei an, wie Sie Ihre Fähigkeiten **zum jetzigen Zeitpunkt** beurteilen **und** schätzen Sie darüber hinaus ein, in welchem Ausmaß die betreffenden Fähigkeiten **aus heutiger Sicht vor dem Seminar am Dienstag** bei Ihnen vorhanden waren.

Beispiel zur Bearbeitung der Aussagen:

Am Ende jeder aufgeführten Aussage befinden sich zwei Skalen von ① bis ⑥. Markieren Sie auf der ersten Skala bitte den Grad, in dem die betreffende Fähigkeit aus heutiger Sicht **vor dem Seminar** bei Ihnen vorhanden bzw. nicht vorhanden war.
Benutzen Sie die zweite Skala, um anzugeben, in welchem Maß Ihre Fähigkeit **zum jetzigen Zeitpunkt** bei Ihnen vorhanden bzw. nicht vorhanden ist.

Wenn sich an Ihrer Fähigkeitseinschätzung vor dem Seminar im Vergleich zu nach dem Seminar (also jetzt) nichts veränderte, dann markieren Sie bitte die gleiche Ziffer in beiden Skalen.

Orientieren Sie sich für die Fähigkeitseinschätzungen bitte an der folgenden Skala:

Skala:

①	②	③	④	⑤	⑥
Optimal vorhanden	Überwiegend vorhanden	Ausreichend vorhanden	Teilweise vorhanden	Kaum vorhanden	Überwiegend nicht vorhanden

Wie beurteilen Sie Ihre Fähigkeit,...	...vor dem Seminar?						...zum jetzigen Zeitpunkt?					
1. durch Kommunikation mit dem Patienten die relevanten Informationen extrahieren zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
2. eine Standardnarkose selbständig einleiten zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
3. dem Patienten die für eine Standardnarkose erforderlichen Medikamente geben zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
4. eine Maskenbeatmung durchführen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
5. den richtigen Zeitpunkt für eine endotracheale Intubation bestimmen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
6. eine endotracheale Intubation durchführen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥

Fortsetzung Teil A

Orientieren Sie sich auch für diese Fähigkeitseinschätzungen bitte an der folgenden Skala:

Skala:

①	②	③	④	⑤	⑥
Optimal vorhanden	Überwiegend vorhanden	Ausreichend vorhanden	Teilweise vorhanden	Kaum vorhanden	Überwiegend nicht vorhanden

	Wie beurteilen Sie Ihre Fähigkeit,...	**...vor dem Seminar?**						**...zum jetzigen Zeitpunkt?**					
7.	fächerübergreifendes Wissen für die zu erstellenden Diagnosen heranziehen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
8.	nach Anamneseerhebung, klinischer Untersuchung und Auswertung des Monitoring Verdachtsdiagnosen, wie z.B. Herzinfarkt, Asthmaanfall, Anaphylaxie, Lungenembolie, ableiten zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
9.	die obigen Differenzialdiagnosen erstellen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
10.	Therapiepläne für die obigen Differenzialdiagnosen erstellen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
11.	dem Patienten die erforderlichen Medikamente für die obigen Differenzialdiagnosen geben zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
12.	die Dosierungen der erforderlichen Medikamente für die obigen Differenzialdiagnosen bestimmen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
13.	durch aktive Informationssammlung eine mögliche Kontraindikation von Medikamenten abklären zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
14.	anfallende Aufgaben im Team effizient aufteilen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥

Haben Sie bei jeder Aussage 2 Einstufungen vorgenommen?

Bitte umblättern →

Teil B

Bitte geben Sie nun Ihre Erwartungen bezüglich der HANS-Einheit an.
Bei den folgenden Fragen genügt die Einschätzung Ihrer Zustimmung zum jetzigen Zeitpunkt.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme überwiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwiegend nicht zu	Stimme gar nicht zu

	Ich erwarte von der HANS-Simulationseinheit...	Ausmaß Ihrer Zustimmung					
1.	während des Unterrichts selbst aktiv werden zu können.	①	②	③	④	⑤	⑥
2.	Mein theoretisches Wissen praktisch anwenden zu können.	①	②	③	④	⑤	⑥
3.	selbständig arbeiten zu dürfen.	①	②	③	④	⑤	⑥
4.	aus meinen Fehlern lernen zu können.	①	②	③	④	⑤	⑥
5.	im Team Diagnosen erstellen zu können.	①	②	③	④	⑤	⑥
6.	anästhesiologische Aufgaben praktisch ausführen zu können.	①	②	③	④	⑤	⑥
7.	das Monitoring realistisch durchführen zu können.	①	②	③	④	⑤	⑥
8.	meine erworbenen Fähigkeiten für den Bereich Anästhesie erproben zu können.	①	②	③	④	⑤	⑥

Teil C

Bitte geben Sie das Ausmaß Ihrer Zustimmung zu den unten folgenden Aussagen an.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme überwiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwiegend nicht zu	Stimme gar nicht zu

Teil C 1		Ausmaß Ihrer Zustimmung					
1.	Wenn ich in der Simulationseinheit viel lerne, habe ich bessere Chancen in der Prüfung.	①	②	③	④	⑤	⑥
2.	Durch die Anästhesie-Simulation werde ich Fähigkeiten erlernen, die mir in der praktischen klinischen Arbeit helfen werden.	①	②	③	④	⑤	⑥
3.	Für die Abschlussprüfung wäre es sinnvoller, die entsprechende Literatur zu lesen als an einer Anästhesie-Simulation teilzunehmen.	①	②	③	④	⑤	⑥
4.	Ich sehe in der Anästhesie-Simulation Lernmöglichkeiten für mich, die ich in keiner anderen Veranstaltungsform haben werde.	①	②	③	④	⑤	⑥

Fortsetzung Teil C

Bitte geben Sie das Ausmaß Ihrer Zustimmung zu den unten folgenden Aussagen an.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme über-wiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwie-gend nicht zu	Stimme gar nicht zu

Teil C 2		Ausmaß Ihrer Zustimmung
1.	Ich bin überzeugt, dass ich die Lernziele der Anästhesie-Simulation erreichen werde.	① ② ③ ④ ⑤ ⑥
2.	Ich weiß genau, dass ich die in der Anästhesie-Simulation an mich gestellten Anforderungen bewältigen kann.	① ② ③ ④ ⑤ ⑥
3.	Eventuellen Schwierigkeiten beim Simulator-Training sehe ich gelassen entgegen, da ich meinen Fähigkeiten vertrauen kann.	① ② ③ ④ ⑤ ⑥
4.	Ich werde die Inhalte der Anästhesie-Simulation gut umsetzen können.	① ② ③ ④ ⑤ ⑥
5.	Ich bin überzeugt, dass es mir gelingen wird, mein fächerübergreifendes Wissen für die Erstellung der Diagnose heranzuziehen.	① ② ③ ④ ⑤ ⑥
6.	Im Durchschnitt sind andere wahrscheinlich viel fähiger als ich, anästhesiologische Aufgaben auszuführen.	① ② ③ ④ ⑤ ⑥
7.	Ich weiß, dass es mir gelingen wird, eine Standardnarkose selbständig einzuleiten.	① ② ③ ④ ⑤ ⑥
8.	Ich werde lange Zeit brauchen, um Verdachtsdiagnosen wie z.B. Herzinfarkt, Asthmaanfall, Lungenembolie,... ableiten und begründen zu können.	① ② ③ ④ ⑤ ⑥

Teil D

Vor einer Veranstaltung hat man oft verschiedene Gedanken oder Fragen im Kopf. Bitte beurteilen Sie im Folgenden, in welchem Ausmaß Ihnen die aufgeführten Gedanken vor dieser HANS-Einheit durch den Kopf gegangen sind.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme über-wiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwie-gend nicht zu	Stimme gar nicht zu

		Ausmaß Ihrer Zustimmung
1.	Ich frage mich, ob ich in der Veranstaltung etwas lernen werde.	① ② ③ ④ ⑤ ⑥
2.	Ich hoffe, ich werde keine Fehler machen.	① ② ③ ④ ⑤ ⑥
3.	Ich hoffe, wir behandeln keine Inhalte, die ich schon gelernt habe.	① ② ③ ④ ⑤ ⑥
4.	Ich frage mich, wie ich wohl abschneiden werde im Verhältnis zu den anderen Studenten.	① ② ③ ④ ⑤ ⑥
5.	Ich hoffe, wir werden ein paar schwierige Inhalte behandeln, an denen ich richtig arbeiten kann.	① ② ③ ④ ⑤ ⑥
6.	Ich hoffe, ich bin genauso gut wie die anderen Studenten.	① ② ③ ④ ⑤ ⑥
7.	Wird es einen Einfluss auf meine Note haben, wie gut ich heute bin?	① ② ③ ④ ⑤ ⑥
8.	Ich würde gerne die Gelegenheit erhalten, meine Fehler mit anderen zu diskutieren.	① ② ③ ④ ⑤ ⑥
9.	Werde ich kompetent erscheinen?	① ② ③ ④ ⑤ ⑥
10.	Ich kann es kaum erwarten anzufangen und zu versuchen, die Inhalte zu verstehen.	① ② ③ ④ ⑤ ⑥

Teil E

Bitte beantworten Sie zum Abschluss noch einige Fragen zu Ihrer Person.

1.	Ihr Alter:	____ Jahre	
2.	Ihr Geschlecht:	☐ männlich	☐ weiblich
3.	Ich verfüge bereits über praktische klinische Erfahrung (z.B. Rettungsdienst, Praktikum,...)	☐ ja	☐ nein

Falls ja, charakterisieren Sie diese bitte kurz.

____

____

____

Vielen Dank für Ihre Mitarbeit und viel Spaß mit HANS!

Appendix D: Student posttest questionnaire

Klinik für Anaesthesiologie

Ruprecht-Karls-Universität Heidelberg

Fragebogen zur HANS-Simulationseinheit in HeiCuMed

Liebe(r) HeiCuMed Teilnehmer(in),

nachdem Sie an der HANS-Simulation teilgenommen haben, möchten wir Sie bitten, nun noch den folgenden Fragebogen auszufüllen.

Wir versichern Ihnen hiermit, dass die Daten der Fragebögen **vollkommen vertraulich** behandelt und nicht an Dritte weitergegeben werden. Ihre Angaben werden nicht von den unterrichtenden Tutoren eingesehen und haben keinen Einfluss auf Ihre Bewertung. Damit wir die Daten einander zuordnen können, bitten wir Sie, die nachfolgenden Angaben zu machen.

Name: ______________________________

Gruppe: ___

Datum: __.__.__

Bearbeitungshinweise:

- Markieren Sie bitte die Antworten, die am ehesten Ihrer Meinung entsprechen, mit einem Kreuz und benutzen Sie die jeweils oben angeführten Skalen zu Ihrer Orientierung.
- Bitte seien Sie bei der Angabe Ihrer Übereinstimmung mit den Aussagen aufrichtig. Es gibt keine „richtigen“, „falschen“ oder „erwünschten“ Antworten.
- Bitte bearbeiten Sie alle Aussagen zügig, in der vorgegebenen Reihenfolge und ohne Fragen bzw. Aussagen auszulassen.
- Sollten Sie während der Bearbeitung noch Fragen haben, wenden Sie sich bitte an die anwesenden HeiCuMed-Tutoren.

A. Schmitz, N. Schaper, C. Grube, B. Graf, T. Boeker

Teil A

Beurteilen Sie im Folgenden bitte Ihre eigenen Fähigkeiten im Bereich Anästhesie.

Bitte geben Sie hierbei an, wie Sie Ihre Fähigkeiten **zum jetzigen Zeitpunkt** beurteilen **und** schätzen Sie darüber hinaus ein, in welchem Ausmaß die betreffenden Fähigkeiten **aus heutiger Sicht vor dem Training** bei Ihnen vorhanden waren.

Beispiel zur Bearbeitung der Aussagen:

Am Ende jeder aufgeführten Aussage befinden sich zwei Skalen von ① bis ⑥. Markieren Sie auf der ersten Skala bitte den Grad, in dem die betreffende Fähigkeit aus heutiger Sicht **vor dem Training** bei Ihnen vorhanden bzw. nicht vorhanden war.
Benutzen Sie die zweite Skala, um anzugeben, in welchem Maß Ihre Fähigkeit **zum jetzigen Zeitpunkt** (also am Ende des Trainings) bei Ihnen vorhanden bzw. nicht vorhanden ist.

Wenn sich an Ihrer Fähigkeitseinschätzung im Verlauf des Trainings nichts veränderte, dann markieren Sie bitte die gleiche Ziffer in beiden Skalen.

Orientieren Sie sich für die Fähigkeitseinschätzungen bitte an der folgenden Skala:

Skala:	①	②	③	④	⑤	⑥
	Optimal vorhanden	Überwiegend vorhanden	Ausreichend vorhanden	Teilweise vorhanden	Kaum vorhanden	Überwiegend nicht vorhanden

Wie beurteilen Sie Ihre Fähigkeit,...	...vor dem Training?						...nach dem Training (jetzt)?					
15. durch Kommunikation mit dem Patienten die relevanten Informationen extrahieren zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
16. eine Standardnarkose selbständig einleiten zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
17. dem Patienten die für eine Standardnarkose erforderlichen Medikamente geben zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
18. eine Maskenbeatmung durchführen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
19. den richtigen Zeitpunkt für eine endotracheale Intubation bestimmen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
20. eine endotracheale Intubation durchführen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥

Fortsetzung Teil A

Orientieren Sie sich auch für diese Fähigkeitseinschätzungen bitte an der folgenden Skala:

Skala:	① Optimal vorhanden	② Überwiegend vorhanden	③ Ausreichend vorhanden	④ Teilweise vorhanden	⑤ Kaum vorhanden	⑥ Überwiegend nicht vorhanden

Wie beurteilen Sie Ihre Fähigkeit,...	...vor dem Training?						nach dem Training (jetzt)?					
21. fächerübergreifendes Wissen für die zu erstellenden Diagnosen heranziehen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
22. nach Anamneseerhebung, klinischer Untersuchung und Auswertung des Monitoring Verdachtsdiagnosen, wie z.B. Herzinfarkt, Asthmaanfall, Anaphylaxie, Lungenembolie, ableiten zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
23. die obigen Differenzialdiagnosen erstellen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
24. Therapiepläne für die obigen Differenzialdiagnosen erstellen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
25. dem Patienten die erforderlichen Medikamente für die obigen Differenzialdiagnosen geben zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
26. die Dosierungen der erforderlichen Medikamente für die obigen Differenzialdiagnosen bestimmen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
27. durch aktive Informationssammlung eine mögliche Kontraindikation von Medikamenten abklären zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥
28. anfallende Aufgaben im Team effizient aufteilen zu können.	①	②	③	④	⑤	⑥	①	②	③	④	⑤	⑥

Haben Sie bei jeder Aussage 2 Einstufungen vorgenommen?

Bitte umblättern →

Bitte geben Sie nun das Ausmaß Ihrer Zustimmung zu den unten genannten Aussagen an.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme über-wiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme über-wiegend nicht zu	Stimme gar nicht zu

Bei den folgenden Fragen genügt die Einschätzung Ihrer Zustimmung zum jetzigen Zeitpunkt.

Teil B

	Ich konnte in der HANS-Simulationseinheit,...	Ausmaß Ihrer Zustimmung
1.	während des Unterrichts selbst aktiv werden.	① ② ③ ④ ⑤ ⑥
2.	mein theoretisches Wissen praktisch anwenden.	① ② ③ ④ ⑤ ⑥
3.	selbständig arbeiten.	① ② ③ ④ ⑤ ⑥
4.	aus meinen Fehlern lernen.	① ② ③ ④ ⑤ ⑥
5.	im Team Diagnosen erstellen.	① ② ③ ④ ⑤ ⑥
6.	anästhesiologische Aufgaben praktisch ausführen.	① ② ③ ④ ⑤ ⑥
7.	das Monitoring realistisch durchführen.	① ② ③ ④ ⑤ ⑥
8.	meine Fähigkeiten für den Bereich Anästhesie erproben.	① ② ③ ④ ⑤ ⑥

Teil C

Bitte geben Sie das Ausmaß Ihrer Zustimmung zu den unten folgenden Aussagen an.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme überwiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwiegend nicht zu	Stimme gar nicht zu

Teil C 1						
1. Vor der HANS-Einheit wusste ich, welche Effekte die Einheit auf mein Können haben sollte.	①	②	③	④	⑤	⑥
2. Was ich von der HANS-Einheit zu erwarten habe, wusste ich vorher.	①	②	③	④	⑤	⑥
3. Die erwarteten Ergebnisse der HANS-Einheit waren mir von Anfang an klar.	①	②	③	④	⑤	⑥
4. Was in der HANS-Einheit vermittelt wurde, entspricht meiner Einschätzung nach weitgehend den Prüfungsanforderungen.	①	②	③	④	⑤	⑥
5. Normalerweise helfen mir praktische Übungen, meine Leistung zu steigern.	①	②	③	④	⑤	⑥
Teil C 2						
9. Die HANS-Simulation war für mich ein wichtiger Bestandteil von Heicumed.	①	②	③	④	⑤	⑥
10. Den Einsatz von HANS in dieser Form kann ich für folgende Semester empfehlen.	①	②	③	④	⑤	⑥
11. Durch HANS konnte ich relevante Problemstellungen in einer realistischen Situation bearbeiten.	①	②	③	④	⑤	⑥
12. Das Schwierigkeitsniveau der HANS-Simulationen war meinem Kenntnisstand angemessen.	①	②	③	④	⑤	⑥
13. Die eingesetzten Fälle motivierten mich zur weiteren Auseinandersetzung mit den Inhalten der Anästhesie-Einheit.	①	②	③	④	⑤	⑥
14. Die Realitätsnähe von HANS wirkte motivierend auf mich.	①	②	③	④	⑤	⑥
15. Ich konnte mich gut auf die Simulationssituation einlassen.	①	②	③	④	⑤	⑥
16. Ich erachte den Simulator als geeignetes Lernmedium für mich.	①	②	③	④	⑤	⑥
17. Ich konnte unbefangen mit der Simulationssituation umgehen.	①	②	③	④	⑤	⑥
18. Am Simulator konnte ich Fähigkeiten erwerben, die ich mit anderen Lernmedien nicht hätte erwerben können.	①	②	③	④	⑤	⑥
19. Die Realitätsnähe des Simulators fand ich hilfreich für das Erlernen der Inhalte.	①	②	③	④	⑤	⑥
20. Insgesamt hat das HANS-Modul meine Erwartungen erfüllt.	①	②	③	④	⑤	⑥

Bitte blättern Sie um →

Teil D

Bewerten Sie im Folgenden bitte den Tutor, der Sie in der HANS-Einheit unterrichtete. Falls Sie von zwei verschiedenen Tutoren unterrichtet wurden, bewerten Sie bitte den Tutor, der Sie heute unterrichtete.

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme überwiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwiegend nicht zu	Stimme gar nicht zu

		Ausmaß Ihrer Zustimmung					
11.	Der Tutor stellte verständliche Fragen.	①	②	③	④	⑤	⑥
12.	Der Tutor sorgte für einen strukturierten Ablauf.	①	②	③	④	⑤	⑥
13.	Durch das Feedback des Tutors bekam ich konkrete Hinweise, wie ich meine Leistung verbessern kann.	①	②	③	④	⑤	⑥
14.	Der Tutor machte die Lernziele der jeweiligen Einheit deutlich.	①	②	③	④	⑤	⑥
15.	Der Tutor legte Wert darauf, dass die Gruppe die Inhalte verstand.	①	②	③	④	⑤	⑥
16.	Der Tutor passte den Ablauf der Einheiten den Bedürfnissen der Gruppe an.	①	②	③	④	⑤	⑥
17.	Der Tutor motivierte mich zur intensiven Auseinandersetzung mit den Inhalten der Fälle.	①	②	③	④	⑤	⑥
18.	Der Tutor präsentierte die Inhalte klar und verständlich.	①	②	③	④	⑤	⑥
19.	Der Tutor sorgte für ein angenehmes Arbeitsklima.	①	②	③	④	⑤	⑥
20.	Der Tutor passte sich bei Erklärungen an den Vorwissensstand der Gruppe an.	①	②	③	④	⑤	⑥
21.	Der Tutor hinterfragte die Aussagen der Gruppe.	①	②	③	④	⑤	⑥
22.	Der Tutor war kompetent.	①	②	③	④	⑤	⑥
23.	Die praktische Anleitung durch den Tutor war angemessen.	①	②	③	④	⑤	⑥
24.	Der Tutor war motiviert.	①	②	③	④	⑤	⑥

Teil E

Bewerten Sie im Folgenden bitte die Abstimmung der einzelnen Veranstaltungen.

		Ausmaß Ihrer Zustimmung					
1.	Die Abfolge der Veranstaltungen „Seminar Klinische Anästhesie“ – „HANS I“ – „Notfallpraktikum“ – „HANS II“ halte ich für sinnvoll.	①	②	③	④	⑤	⑥
2.	Die Inhalte der obigen Veranstaltungen waren gut aufeinander abgestimmt.	①	②	③	④	⑤	⑥
3.	Die Inhalte des Notfallpraktikums bereiteten mich gut auf „HANS II“ vor.	①	②	③	④	⑤	⑥
4.	Durch die Erfahrungen im Notfallpraktikum profitierte ich mehr von der zweiten Simulationseinheit (HANS II) als von der ersten.	①	②	③	④	⑤	⑥

Vielen Dank für Ihre Mitarbeit und weiterhin viel Erfolg bei HeiCuMed!

Appendix E: Instructor posttest questionnaire

Beurteilung der HANS-Einheit durch den Tutor

Name:

Datum:

Beurteilte Gruppe:

Einschätzung der Gruppe

Skala:

①	②	③	④	⑤	⑥
Stimme voll zu	Stimme über-wiegend zu	Stimme eher zu	Stimme eher nicht zu	Stimme überwie-gend nicht zu	Stimme gar nicht zu

	Die Gruppe	Ausmaß Ihrer Zustimmung
1.	war gut vorbereitet.	① ② ③ ④ ⑤ ⑥
2.	beteiligte sich aktiv.	① ② ③ ④ ⑤ ⑥
3.	war motiviert zu lernen.	① ② ③ ④ ⑤ ⑥
4.	war homogen hinsichtlich ihres Leistungsniveaus.	① ② ③ ④ ⑤ ⑥
5.	konnte auf den Patienten zugehen.	① ② ③ ④ ⑤ ⑥
6.	kooperierte untereinander.	① ② ③ ④ ⑤ ⑥

7. Was ich noch anmerken möchte: ______________________________

____

A. Schmitz, N. Schaper, C. Grube, B. Graf, T. Boeker

Appendix F: Student prior experiences

Table 136. *Type of experience as indicated by participants*

Type of experience	N	Relevancy
Famulatur/Praktikum (nicht näher spezifiziert)	31	No
Famulatur/Praktikum Allgemein-Medizin	1	No
Famulatur/Praktikum Ambulanz	2	No
Famulatur/Praktikum Anästhesie	17	Yes
Famulatur/Praktikum Chirurgie	8	No
Famulatur/Praktikum Gynäkologie	4	No
Famulatur/Praktikum Herzchirurgie	2	No
Famulatur/Praktikum im OP	4	No
Famulatur/Praktikum Innere	11	No
Famulatur/Praktikum Intensivmedizin	3	Yes
Famulatur/Praktikum Notfallmedizin	3	Yes
Famulatur/Praktikum Orthopädie	1	No
Praktikum Pflege	10	No
Notfall-Kurs	3	Yes
Nachtdienst	2	No
Pflege	8	No
Rettungsdienst	51	Yes
Andere (DLRG, Zivi, Doktorarbeit)	15	No