

- Tulen, J. H. M., Vokers, A., Stronks, D., Cavelaars, M. & Groenevela, W. (2001). Accelerometry in clinical psychophysiology. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 207–232). Seattle, WA: Hogrefe & Huber.
- Turner, C. F., Villarroel, M. A., Rogers, S. M., Eggleston, E., Ganapathi, L., Roman, A. M. et al. (2005). Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of telephone audio-CASI. *Addiction, 100*, 1432–1444.
- Van den Berg-Emons, R. J., Schasfoort, F. C., De Vos, L. A., Bussmann, J. B. & Stam, H. J. (2006). Impact of chronic pain on everyday physical activity. *European Journal of Pain, 11* (5), 587–593.
- Van't Hooft, M., Díaz, S. & Swan, K. (2004). Examining the potential of handheld computers: Findings from the Ohio pep project. *Journal of Educational Computing Research, 30*, 295–311.
- Vogt, J. & Kastner, M. (2001). Psychophysiological monitoring of air traffic controllers: Exploration, simulation, validation. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 455–476). Seattle, WA: Hogrefe & Huber.
- Vouyioukas, D., Maglogiannis, I. & Pasiadis, V. (2007). Pervasive E-health services using the DVB-RCS communication technology. *Journal of Medical Systems, 31* (4), 237–246.
- Waters, A. J. & Li, Y. (2008). Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology, 97*, 25–35.
- Weiss, H. M., Beal, D. J., Barros, E. & MacDermid, S. M. (2003). *Conceptualizing performance processes: A model to guide research linking quality of life and performance* (Technical report prepared for the Office of Military Community and Family Policy, Department of Defense). West Lafayette, IN: Purdue University, Department of Psychological Sciences.
- West, S. & Hepworth, J. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality and Social Psychology, 59*, 609–662.
- Whalen, C. K., Jamner, L. D., Henker, B. & Delfino, R. J. (2001). Smoking and moods in adolescents with depressive and aggressive dispositions: Evidence from survey and electronic diaries. *Health Psychology, 20*, 99–111.
- Wilhelm, F. H., Roth, W. T. & Sackner, M. A. (2003). The LifeShirt: An advanced system for ambulatory measurement of respiratory and cardiac function. *Behavior Modification, 27* (5), 671–691.
- Wilson, G. (2001). In-flight psychophysiological monitoring. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 435–454). Seattle, WA: Hogrefe & Huber.
- Wittmann, W. W. (1985). *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Berlin: Springer.
- Zeier, H., Häseli, A. & Fischer, J. (2001). Heart rate monitoring in an academic test situation. In J. Fahrenberg & M. Myrtek (Eds.), *Progress in ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 387–398). Seattle, WA: Hogrefe & Huber.

13. Kapitel

Einsatz technischer Mittel
in der psychologischen Diagnostik

Joachim Funke und Bernd Reuschenbach

1 Einleitung

Der Einsatz technischer Mittel innerhalb der Diagnostik ist so alt wie die wissenschaftliche Disziplin selbst. Bekannt dürften apparative Verfahren zur Berufseignung sein, die zu Zeiten der „Industriellen Psychotechnik“ (Giese, 1927; Moede, 1928) Anfang des 20. Jahrhunderts eingesetzt wurden (vgl. Albert & Gundlach, 1997). Schulte (1921) entwickelte etwa eine Apparatur zur Tauglichkeitsprüfung angehender Damenfriseurinnen, mit der die Zugstärke beim Kämmen von Perücken als Indikator für die Sanftheit der Behandlung gemessen werden konnte. Über verschiedene, historisch interessante psychotechnische Prüfverfahren für Straßenbahner und Eisenbahner berichtet im Detail der Artikel von Stapf (1997).

Zur Eingrenzung des Themas ist es wichtig, zunächst den Begriff der Technik zu reflektieren. Technik (altgriechisch: τεχνη) bedeutet übersetzt „Fähigkeit“ oder „Kunstfertigkeit“. Diagnostik (altgriechisch: διαγνωσις) bedeutet übersetzt „Entscheidung“. So gesehen kann auch die Zuweisung von Körperkonstitutionen zu bestimmten Persönlichkeitseigenschaften oder die Entwicklung von Fragebögen als Kunstfertigkeit und damit im weitesten Sinn als „technische Diagnostik“ verstanden werden.

Wenn von *technischen Mitteln* die Rede ist, dann deutet dies auf eine weitere Facette des Technikbegriffs hin, nämlich auf die Nutzung von Mitteln, die mechanischer oder elektrotechnischer Natur sind bzw. sich entsprechendes technologisches Wissen zunutze machen. Die verwendeten Mittel sind das Ergebnis von Entwicklungsarbeiten im Bereich der Elektrotechnik, des Ma-

schinenbaus, der Medizintechnik, der Computertechnik oder der Anlagentechnik. In verschiedenen Bereichen der Enzyklopädie sind solche technischen Mittel schon thematisiert worden, z. B. beim Einsatz von Computern in der Leistungsdiagnostik (vgl. Rammsayer & Stahl, in diesem Band; vgl. Hornke & Franke, 2011), bei der Prüfung der Verkehrstauglichkeit in Fahrsimulatoren, der Nutzung von *Personal Digital Assistants* (PDA) beim ambulanten Assessment (vgl. Reuschenbach & Funke, in diesem Band) oder der Erhebung von biopsychologischen Daten mittels Elektrokardiogramm (EKG), Blutdruckmessgeräten oder aufwändigen Instrumentarien wie Magnetresonanztomografie (MRT), Computertomografie (CT) oder Positronen-Emissions-Tomografie (PET; vgl. Rammsayer & Stahl, in diesem Band).

Während in den dortigen Kapiteln einzelne psychologische Konstrukte (Intelligenz, Angst, Risikobereitschaft etc.) oder bestimmte Einsatzbereiche (Neuropsychologie, Berufseignung, Verkehrstauglichkeit) im Mittelpunkt stehen, sollen hier die technischen Aspekte und die sich daraus ergebenden zukünftigen Möglichkeiten für die Diagnostik dargestellt werden. Der Schwerpunkt liegt dabei auf besonders innovativen Technologien wie virtuellen Welten, komplexen Szenarien oder dem „serious gaming“. Ein Blick in die Geschichte der Psychologie verdeutlicht, dass die Innovationen von damals heute als nicht mehr zeitgemäß und überholt gelten: Beispielhaft sei hier die Sammlung von „automatic testing devices“ von Pressey (1926) erwähnt, die mithilfe mechanischer Vorrichtungen Items darboten und zum Beantworten die Bedienung von Hebeln notwendig machten. Erinnert sei an den Einsatz von Lochkartensystemen in der Diagnostik (Kleinmuntz & Barton, 1962) oder die Verwendung von Schmalfilmen als Vorform moderner Videotests (Lumsdaine, 1960). Selbst derartige Videotests, wie sie in den 1990er Jahren erstmals vorgeschlagen wurden (z. B. Hornke, Dewald & Hausen, 1990; Müller, 1995), sind aufgrund technischer Weiterentwicklungen heute fast schon als veraltet zu bezeichnen.

Die Schnelllebigkeit der technischen Errungenschaften scheint zunächst kaum mit dem Ansinnen einer Enzyklopädie nach einer überdauernden Darstellung des Wissens kompatibel. Wir versuchen dennoch diesen Spagat, indem wir den derzeitigen Stand der technischen Möglichkeiten darstellen und Bewertungen aus benachbarten Bereichen (z. B. den Computerwissenschaften und den Ingenieurwissenschaften) hinzuziehen.

2 Systematisierung des technischen Einsatzes in der Diagnostik

Zur Systematisierung von technischen Mitteln in der Diagnostik bieten sich die drei Facetten (1) Phase der Testentwicklung, (2) psychologisches Konstrukt und (3) verwendete Technik an.

Die erste Facette bezieht sich auf den Einsatz von Technik *in den einzelnen Phasen der Konstruktion und der Anwendung von Tests*. Hier wurde exemplarisch eine globalere Systematisierung des Verlaufs anhand des Phasenschemas von Jäger und Krieger (1994, für eine Systematisierung anhand des Verlaufs) zugrunde gelegt. Technik kann in der Testentwicklung, der Testdurchführung oder in der Auswertung eine Rolle spielen. In den nachfolgenden Abschnitten 3 und 4 liegt der Schwerpunkt im Bereich der Testdurchführung, also der Datenerhebung (vgl. Tab. 1).

Tabelle 1:
Technikeinsatz in einzelnen Phasen des diagnostischen Prozesses

Schritte im diagnostischen Prozess		Technische Mittel
Testkonstruktion	automatisierte Validitätsstudien	– automatisches Scoring (Bennett & Bejar, 1998)
	automatische Itementwicklung und Testkonstruktion	– „GRE – Mathematik Item Generator“ von Bejar et al. (2002) – ITEMGEN1 (2002) – „Test Creation Assistant“ (Singley & Bennett, 2002)
Datenerhebung/ Testdurchführung (vgl. Abschnitt 3.1.5 für weitere Beispiele)	Item-/Stimulusvorgabe	– Computerunterstütztes Testen inkl. Online-Testung (z. B. Martin, 2003; Wirth, 2008) – Videotests (Fricke, 1990) – Virtual Reality (z. B. McGeorge et al., 2001) – Simulationen (Kluwe, 1997) – Itemdarbietung per „interactive voice recording“ (z. B. Cleeland, 2000)
	Unterstützung bei der Dateneingabe	– „self-assessment“ mittels elektronischem Tagebuch (z. B. Roelofs, Peters, Patijn, Schouten & Vlaeyen, 2004) – Eingabehilfen für Interviewer, z. B. PDA-gestützte Leitfadenterviews
	unmittelbare Datenerhebung	– Erhebung von Nutzerdaten mittels RFID-Chips – physiologische Messung (stationär und ambulant) – Augenbewegung/Pupillometrie (Nuthmann, Engbert & Kliegl, 2006) – Erfassung von Bewegungsdaten, z. B. mittels Satellitentracking – Messung von Reaktionszeiten bei der computer-gestützten Testung (Ulrich & Schröter, 2006)

Tabelle 1 (Fortsetzung):
Technikeinsatz in einzelnen Phasen des diagnostischen Prozesses

Schritte im diagnostischen Prozess		Technische Mittel
Auswertung	Analyse von Beobachtungsdaten	<ul style="list-style-type: none"> - vgl. Kahng und Iwata (2000) für einen Überblick über Analysesoftware - INTERACT für Videodaten (Mangold International GmbH, 2006) - KIDTALKER (Analyse von linguistischen und verhaltensbezogenen Eigenheiten) von Kaiser et al. (2000)
	automatisierte Auswertung und Scoring qualitativer Daten	<ul style="list-style-type: none"> - WordMap zur Analyse grammatischer Fehler (Bejar, 1988) - PSYMEDIA (Hageböck, 1990) - „Automated essay scoring“ (z. B. Shermis & Burstein, 2003)
	statistische Analysen	<ul style="list-style-type: none"> - Quantitativ: z. B. SPSS, SAS - Qualitativ: z. B. NVivo (Auerbach & Silverstein, 2003), QCA (Schneider & Wagemann, 2007)
	diagnostische Entscheidungshilfen	<ul style="list-style-type: none"> - DIA-X Expertensystem zur Diagnostik psychischer Störungen (Wittchen, Weigel, Pfister & Perkonig, 1997) - DIASYS1 zur computerunterstützten Diagnose in der schulpyschologischen Beratung (Hageböck, 1994)
	Gutachten-erstellung	<ul style="list-style-type: none"> - computerbasierte Auswertung (z. B. Harris, 1987)
Rückmeldung	Rückmeldung an den Diagnostiker	<ul style="list-style-type: none"> - Hogrefe TestSystem (http://www.hogrefe-testsystem.com/)
	Rückmeldung an die Testperson	<ul style="list-style-type: none"> - „self monitoring“ - „Therapeutic“: direkte Feedbackschleifen zwischen medikamentöser Therapie und Diagnostik
	automatisierte Rückmeldung	<ul style="list-style-type: none"> - „Smart-home“- und „Ambient-intelligent“-Technologie (Beale, 2007)
Datenspeicherung	Computer, Handheld-PC, Smartphones, RFID-Tags, Computernetze, zentrale Datenbanken	

Die *zweite* Facette bezieht sich auf die relevanten *psychischen Dimensionen*, die im Mittelpunkt des diagnostischen Interesses stehen. Ebenso denkbar ist eine Einteilung in Anwendungsfelder. Diese Systematisierung, der auch der erste Band der Enzyklopädie folgt, vernachlässigt jedoch den Umstand, dass psychische Konstrukte (z. B. Leistung) in verschiedenen Bereichen der Psychologie (z. B. in der Arbeitspsychologie, der Pädagogischen Psychologie oder der Rehabilitationspsychologie) eine Rolle spielen. Eine Gliederung auf der Ebene der psychischen Konstrukte scheint daher sinnvoller. Für den klinischen Bereich empfehlen Schandry und Leopold (1996) folgende Dreiteilung: (1) Messung medizinischer Parameter (z. B. Blutdruck, Herzfrequenz), (2) subjektive Symptome (z. B. Schmerzen) und (3) Variablen, die indirekt mit der Erkrankung und Symptomen verbunden sind (z. B. Stimmung).

Die *dritte* Facette bezieht sich auf die *Art der eingesetzten Technik*. Das Spektrum reicht vom Einsatz stationärer EDV-Anlagen inklusive der Online-Testung bis hin zu Mobiltelefonen, Handheld-Geräten (PDA) oder physiologischen Messgeräten (vgl. Kasten 1).

Kasten 1:

15 Beispiele für technische Geräte in der Diagnostik

1. Magnetresonanztomografie, Computertomografie, Positronen-Emissions-Tomografie
2. Physiologische Messgeräte für Blutdruck, Puls, Herzrate, Hautleitfähigkeit etc.
3. Computer (mobil + stationär)
4. Computertestsysteme, d. h. besondere Eingabegeräte und zentrale Datenverwaltungsmöglichkeiten, z. B. Wiener Testsystem und Hogrefe TestSystem
5. Virtuelle Welten
6. Ambulante physiologische Messvorrichtungen, z. B. „Vitaport“ und „LifeShirt“
7. Handheld-Computer (Personal Digital Assistants, PDA), inkl. Video- und Sprachaufzeichnungsfunktion
8. Computersimulationen
9. Video und digitale Filmmedien
10. RFID-Chips (Kundenkarten, Zugangskontrollen, Tracking)
11. Eyetracking und Pupillometrie
12. Telefon und Cellular phones (Handys) – inkl. Wireless Application Protocol (WAP), General Packet Radio Service (GPRS), Short Message Service (SMS), Multimedia Messaging Service (MMS), Universal Mobile Telecommunications System (UMTS)
13. Automatisierter Anruf von Probanden mittels „Interactive Voice Recording“ (IVR)
14. Gruppenmessverfahren: „Voting machine“, Tele-Dialog (TED)
15. „Electronic surveillance measures“, „biomechanical activity devices“ inkl. Global Positioning Systems (GPS)

Die Verwendung technischer Hilfsmittel ist natürlich nicht unproblematisch, wenn Psychologie als Verhaltenswissenschaft verstanden wird: Warum sollte

man den Umweg über Hilfsmittel wählen, wenn das Verhalten der eigentliche Gegenstand ist? Baumeister, Vohs und Funder (2007) kritisieren in einem bissigen Beitrag den Verlust der Verhaltensorientierung in der sozialpsychologischen Forschung und beklagen, dass anstelle von Verhalten überwiegend Selbstberichte, hypothetische Szenarios und Fragebogenangaben als abhängige Variablen in entsprechenden Experimenten auftauchten. In der von der *American Psychological Association* ausgerufenen „Decade of Behavior“ (2000 bis 2009) fordern sie eine verstärkte Rückkehr zu beobachtetem Verhalten (das ironischerweise „actual behavior“ genannt wird), das u. a. aus ethischen Gründen immer seltener zum Gegenstand der Datenerhebung gewählt wurde. Nach unserer Meinung steht die Nutzung technischer Hilfsmittel nicht im Widerspruch zur Forderung von Baumeister et al. (2007), sondern könnte – wie zu zeigen sein wird – sogar einen wichtigen Beitrag in die gewünschte Richtung liefern. In den folgenden Abschnitten sollen einige innovative technische Instrumentarien vorgestellt werden, die bereits zum Einsatz kommen und in Zukunft vermutlich noch an Bedeutung gewinnen werden.

3 Beispiele moderner technischer Diagnostika

Wieso kommt es überhaupt zum Einsatz technischer Mittel in der Diagnostik? Ein Blick auf die Entwicklung der Computardiagnostik kann dies beantworten. Die technische Entwicklung fand zunächst außerhalb der Psychologie statt. Erst später wurden diese Möglichkeiten in der Diagnostik genutzt. Anfänglich stand die Adaptation einfacher *Papier-Bleistift*-Verfahren im Vordergrund. Erst allmählich kam es zu eigenständigen Entwicklungen, die sich die technischen Möglichkeiten zunutze machten. An die Stelle klassischer Item- und Antwortformate traten schließlich komplexe „Items“ wie Bilder und Videos (vgl. Abschnitt 3.4), und mit der Entwicklung adaptiver computergestützter Testverfahren nutzte man die Möglichkeit einer sofortigen Bewertung des Probanden und der darauf aufbauenden Itemauswahl. Eine analoge Entwicklung mit der Sequenz „Technikentwicklung → Techniknutzung für etablierte diagnostische Verfahren → Technik zur Weiterentwicklung der Diagnostik“ findet sich auch im Hinblick auf andere technische Geräte wie Pocket-PCs oder die Nutzung der Mobilfunktechnologie (vgl. Reuschenbach & Funke, in diesem Band).

Der verstärkte Einsatz von technischen Geräten in der Diagnostik ist dabei durch folgende fünf Aspekte motiviert:

1. Die Diagnostik kann für die *Testpersonen* einfacher und unauffälliger gestaltet werden. Hier steht das Motiv einer Vereinfachung für die Probandinnen und Probanden im Vordergrund.

2. Der Nutzen für die *Testleitung* besteht in einer erleichterten Datenerhebung und Datenauswertung. Während in früheren Zeiten Papier-Bleistift-Tests mit Schablonen ausgewertet werden mussten und durch Nachschlagen in Normierungstabellen Rohwerte in T-Werte o. Ä. umgerechnet wurden, ist es heute möglich, die Daten mittels Computer, RFID-Tags (RFID = *Radio Frequency Identification*, reiskorngroße Systeme zur automatischen Identifizierung von Gegenständen oder Personen) oder ambulanten physiologischen Messungen direkt zu erheben. Ebenso sind *automatisierte Auswertungen und direkte Rückmeldungen* an die Testleitung sowie an die Testpersonen selbst möglich. Im Konzept des „ambient home“ ist diese automatisierte Rückmeldung in weitreichender Form umgesetzt. „Ambient home“ beschreibt intelligente Hauseinrichtungen, die beispielsweise das Essverhalten der Bewohner erfassen und automatisch entsprechende Anpassungen der Nahrungsvorräte bewirken (vgl. Reuschenbach & Funke, in diesem Band). Derartige technische Mittel erlauben eine unaufdringliche Messung an vielen Personen zu verschiedenen Zeitpunkten und mit vielfältigen Erhebungsformen. Allerdings kann die Ansammlung großer automatisch erfasster Datenmengen für den Diagnostiker auch eine Herausforderung sein. Hier sind Weiterentwicklungen und Kompetenzen im „data mining“ und der fragestellungsgerechten Datenaggregation unerlässlich.
3. Ein weiteres wichtiges Motiv ist das Vorantreiben der differenzialpsychologischen und methodologischen *Grundlagenforschung*. So haben EDV-gestützte Testvorgaben die Entwicklung adaptiver Testverfahren begünstigt. Durch das ambulante Assessment sind neue statistische Methoden entwickelt worden, um den Einfluss von Trait, State, Situation und entsprechende Interaktionen im Hinblick auf die erhobenen Testwerte adäquat zu bestimmen (vgl. Reuschenbach & Funke, in diesem Band).
4. Mittels technischem Equipment ist es möglich, die *Feldwirklichkeit* im Labor realitätsnäher zu gestalten oder gut kontrollierte Messungen im Felde überhaupt erst zu ermöglichen. So können die Vorteile von Labor- und Felddiagnostik hinsichtlich der ökologischen Validität bzw. der internen/externen Validität durch die Annäherung beider Dimensionen genutzt werden. Ökologische Validität wird hier im Sinne Schmucklers (2001) als weitgehende Integrität realer und repräsentativer Anforderungen verstanden, die zu natürlichem, repräsentativen und relevanten Verhalten führen.
5. Technische Geräte ermöglichen auch eine *verhaltensnahe Erfassung relevanter psychologischer Variablen*. Beispielsweise eröffnen Entwicklungen wie Fahr simulatoren oder Akzelerometer eine unmittelbarere Erfassung der Risikobereitschaft bzw. der Bewegungsaktivitäten, was mit Fragebögen nur retrospektiv möglich ist.

Aus dem Fundus der vielen diagnostischen Möglichkeiten, die sich durch die technischen Weiterentwicklungen ergeben, stellen wir im folgenden Compu-

tersimulationen, digitale Lernspiele, virtuelle Welten und Videotests vor. Es sind Verfahren, die im Sinne von Gray (2002) eine hohe „correspondence“ haben, d. h. es besteht eine große Passung zwischen Test und realer Anforderung. Die „correspondence“ bestimmt die Validität, insbesondere die ökologische Validität des Test- und Messverfahrens.

Letztlich stellt bei Items allein schon die Aufforderung an die Testpersonen, sich eine bestimmte Situation vorzustellen und ihr hypothetisches Verhalten mitzuteilen, einen Versuch dar, die externe Validität gegenüber nicht situativen Einstellungsfragebögen zu erhöhen. Mit der Nutzung komplexer Stimuli wie Bildern und Videos wird die „correspondence“ weiter erhöht. Sind schließlich adaptive Eingriffe in die vorgestellte Welt möglich, wie z. B. in virtuellen Realitäten, ist das Diagnostikum dem Ziel der Realitätsabbildung noch ein Stück näher gekommen, da es eine Mensch-Umwelt-Interaktion ermöglicht.

3.1 Computersimulationen

Computersimulationen verfolgen das Ziel, dynamische und vernetzte Aspekte der Umwelt, die das Verhalten, die Einstellung oder Emotionen beeinflussen, realitätsnah abzubilden. An die Stelle von Schilderungen in textlicher Form treten hier wirklichkeitsnahe Stimuli. Sie sind damit ein Kompromiss zwischen Messungen *in situ* und der Kontrollierbarkeit der Stimuli im Labor.

Im Gegensatz zu Arbeitsproben und virtuellen Welten (vgl. Abschnitt 3.3), bei denen neben der realen Anforderung auch die geforderten Handlungen realistisch sind und diese auf die Objekte direkt einwirken, erfolgt bei der Computersimulation die Reaktion meist durch das Anklicken von Buttons oder die Bewegung von Objekten auf dem Bildschirm. Dies führt dann am Bildschirm zu entsprechenden Konsequenzen, die vorher in der Programmierung festgelegt wurden und die mit der Realität und der erwarteten Konsequenz der Testperson übereinstimmen können, aber nicht müssen.

Die Möglichkeit, adaptiv Handlungsfolgen zu beeinflussen, ist ein wesentliches Unterscheidungsmerkmal zwischen tatsächlicher und computerisierter Arbeitsprobe (vgl. Funke, 1993a). Realitätsnähe kann sich also auf die Items (Anforderung), die Eingriffsmöglichkeiten (Tastatur, Joystick, angeschlossene technische Geräte etc.) oder die Handlung-Handlungskonsequenz-Abfolge beziehen. Schiff, Arnone und Cross (1994) unterscheiden „display realism“ und „response-system realism“. *Display realism* bezieht sich auf die reiz- oder itembezogene Ähnlichkeit mit der Umwelt: „We try to simulate the way things look as closely as possible“ (S. 192). *Response-system realism* bezieht sich auf die Art, wie in die Simulation eingegriffen werden kann. So ist es beispielsweise für eine Fahrsimulation nicht

nur wichtig, dass die Straßen, Bäume und Personen mit einer realen Fahrsituation identisch sind (display realism), sondern auch, dass die Bremsvorgänge mit dem tatsächlichen Bremsvorgang ähnlich sind und der Eingriff entsprechende Konsequenzen nach sich zieht, die im Alltag auch zu erwarten sind. Mit zunehmenden Möglichkeiten für adaptive Reaktions-Item-Konsequenzen entfernt sich die Computersimulation von den Anforderungen klassischer Testverfahren und wird von den Testpersonen eher als situierte Aktion denn als Test erlebt.

Vor allem in der Grundlagenforschung zu Denken und Problemlösen, aber auch im Rahmen der Führungskräfteauswahl werden Computersimulationen als Alternative zu klassischen Intelligenztests gesehen (vgl. Kersting, 1999). Omodei und Wearing (1995) beschreiben den Einsatz einer Simulationssoftware „Fire Chief“, mit der Buschbrände simuliert werden können und in der effektive Brandbekämpfung geübt wird. Diese Software wird sowohl anwendungsorientiert zum Training australischer Buschfeuerwehren (Omodei, Wearing & McLennan, 2000) als auch grundlagenorientiert zur Analyse des Problemlösens in sozialen Gruppen eingesetzt (Feuchter & Funke, 2004).

Fünf Merkmale bestimmen die *Eigenschaften von Computersimulationen*, wie sie zur Untersuchung des menschlichen Umgangs mit Ungewissheit und Komplexität eingesetzt werden (vgl. Betsch, Funke & Plessner, 2011, Kap. 12): Komplexität, Vernetztheit, Dynamik, Intransparenz und Polytelie. Alle fünf Merkmale stellen in diagnostischer Hinsicht spezifische Anforderungen dar, die kurz benannt werden sollen.

Komplexität. Komplexität eines Sachverhalts bedeutet, dass man es mit einem Gegenstandsbereich zu tun hat, der aus verschiedenen Variablen besteht, die in ihrer Vielzahl und Vernetzung die Verarbeitungskapazität eines menschlichen Problemlösers überschreiten und daher Maßnahmen zur Komplexitätsreduktion verlangen. Inwiefern ein angemessenes Auflösungslevel einer problematischen Situation gewählt wurde, ist bislang diagnostisch nur durch ein subjektives Urteil zu entscheiden – Messverfahren dazu liegen unseres Wissens nicht vor.

Vernetztheit. Vernetztheit liegt dort vor, wo mehrere Variablen beteiligt sind. Sie bezieht sich auf die Tatsache, dass in komplexen Sachverhalten nicht einfach nur mehrere Variablen beteiligt, sondern diese Variablen zudem untereinander abhängig sind. Daher ist auf Seiten der handelnden Person Modellbildung erforderlich. Hierfür sind Messverfahren wie z. B. die Kausaldiagramm-Diagnostik entwickelt worden (Blech & Funke, 2006). Auch der MicroDYN-Ansatz (Greiff, 2011; Greiff & Funke, 2010) erlaubt die Messung der Modellbildungskompetenz.

Dynamik. Das Merkmal der Dynamik stellt eine besondere Herausforderung an die problemlösende Person: Während statische Probleme (z. B. eine einzelne

Schachposition) intensives Nachdenken ermöglichen, ohne dass sich die Situation verändert, kommt es bei dynamischen Problemen auch zur besonderen Form der Eigendynamik, d. h. der Systementwicklung aus sich heraus ohne Zutun der handelnden Person. Dynamische Prozesse fordern daher die Fähigkeit zum Abschätzen zeitlicher Entwicklungen. Eine standardisierte Diagnostik dieser Fähigkeit fehlt.

Intransparenz. Entweder kann man aufgrund mangelnder Zeit nicht alle wichtigen Informationen sammeln, die zu einer Entscheidung notwendig wären (zeitlich bedingte Intransparenz), oder man hat zwar viel Zeit, kommt aber aus prinzipiellen Gründen nicht an die benötigten Informationen (strukturell bedingte Intransparenz). Intransparenz verlangt Entscheidungen darüber, welche Informationen noch zu beschaffen sind bzw. wie die vorhandenen Informationen bewertet werden können. Das „Mouselab“-System (Payne, Bettman & Johnson, 1993, Appendix) ist ein Beispiel für ein „process tracing“ des Informationssuch- und beschaffungsverhaltens, das in der Entscheidungsforschung vielfach Anwendung findet (vgl. Funke & Spering, 2006).

Polytelie. Eine polytelische, d. h. vielzielige Problemstellung bringt die problemlösende Person in Wertkonflikte: Welches der möglichen Ziel ist zu erreichen? Wie kann ich das eine Ziel erreichen ohne das andere aufzugeben? Typisch polytelische Situationen zwingen z. B. zum Abwägen zwischen Ökonomie und Ökologie, zwischen Familie und Arbeit. Gefordert ist in diesen Situationen eine Balance zwischen den konfligierenden Zielen, ein Ausgleich zwischen den verschiedenen Interessen in Form einer Kompromissbildung. Instrumente zur Erfassung des Umgangs mit Wertkonflikten stecken noch in den Anfängen (vgl. Blech & Funke, 2010).

3.1.1 Anwendungsbegründung

Computersimulationen ermöglichen die Analyse von Reaktionen und Einstellungen gegenüber *komplexen Anforderungen*. Die Komplexität ergibt sich durch die Vielfalt an visuellen und akustischen Komponenten, die miteinander interagieren (Dynamik) und für die Testpersonen nicht in vollem Umfang einsehbar sind (Intransparenz).

Meist handelt es sich um dynamische Simulationen, bei denen das Verhalten der Testperson den weiteren Testablauf beeinflusst. Gerade das Agieren über die Zeit hinweg und unter sich adaptiv verändernden Anforderungen ist ein wesentlicher Vorteil der Computersimulation gegenüber klassischen Paper-pencil-Verfahren. Sie kommen dann zum Einsatz, wenn es um das Erkennen notwendiger Handlungen ausgehend von komplexen Stimuli geht. Computersimulationen können als Kompromiss zwischen realen Anforderungen des Feldes und kontrollierten

Aktivitäten im Labor gesehen werden (Brehmer & Dörner, 1993). Dies lässt sich an einem Beispiel verdeutlichen: Um die diagnostischen Fähigkeiten von Hautärzten zu testen, reicht es nicht aus, sie mit Fallvignetten zu konfrontieren, die die genaue Symptomatik schriftlich darstellen. Viel näher an der tatsächlichen Anforderung ist die Darstellung von Symptomen durch Bilder, Filme und das in der Computersimulation nachgestellte Frage-Antwort-Verhalten in einer Anamnese (z. B. Ackerman, Evans, Park, Tamassia & Turner, 1999). Bilder sagen nicht nur mehr als Worte, sie erlauben auch vielfältigere Interpretationsmöglichkeiten. Dies ist für die Realitätsnähe bedeutsam, kann jedoch für die Bewertung ein Problem darstellen, da Texte die Aufmerksamkeit klar lenken, während die Wahrnehmung und Interpretation von Bildern oder Videos durch eine Vielzahl verschiedener interpersoneller Faktoren beeinflusst ist, die es zu kontrollieren gilt.

Computersimulationen erlauben eine *Manipulation von Zeit*: Einerseits kann man die Option des Zeitruffers vorsehen (1 Stunde Echtzeit = 12 simulierte Monate), andererseits aber auch die der Zeitlupe (schnell ablaufende Prozesse werden verlangsamt oder angehalten, um sie genau zu beobachten); im Simulator (siehe unten) dagegen ist Echtzeit wichtig.

Reuschenbach (2008) hat in seiner Computersimulation „NursePlan“ den Alltag auf einer Pflegestation multimedial abgebildet, um in diagnostischer Absicht das Agieren von Pflegenden unter Zeitdruck zu erfassen. „NursePlan“ ist eine Computersimulation, die im Zeitraffer typische Planungs- und Problemlöseanforderungen abbildet, wie sie täglich im Stationsalltag vorkommen. Die Aufgabe besteht darin, in der virtuellen Pflegewelt verschiedene pflegerische Probleme zu erkennen und dann geeignete Pflegemaßnahmen umzusetzen. Wie im wahren Stationsalltag müssen viele Patienten und Patientinnen gleichzeitig versorgt werden und die Zeit ist begrenzt. Daher ist es wichtig, die richtigen Prioritäten zu setzen (Was mache ich zuerst?). Weiterhin ist planvolles und vorausschauendes Handeln wichtig (z. B. Was passiert, wenn ich bei der Patientin pflegerische Maßnahmen verzögert umsetze?) Die besondere Innovation der Simulation besteht darin, dass hier patientenübergreifende Anforderungen abgebildet werden. Es müssen also nicht nur einzelne „Fälle“, sondern ganze Patientengruppen versorgt werden (vgl. Abb. 1).

Der inhaltlichen Gestaltung der einzelnen Fälle ging eine qualitative Anforderungsanalyse voraus. Die Lösung solcher alltagsnaher komplexer Probleme ist nicht einfach, da die zeitlichen und personellen Ressourcen begrenzt sind, Wertkonflikte auftreten und unvorhersehbare Handlungsunterbrechungen (z. B. Notfälle) die Umsetzung bekannter Lösungen behindern.

Computersimulationen stellen meist eine *adaptive Testform* dar. Diese Adaptivität kann sich auf die Folge von Testitems beziehen oder aber auch die Gestal-

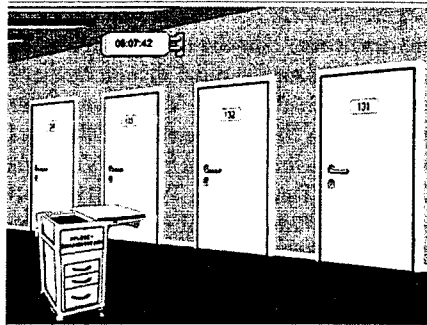


Abbildung 1:

Screenshot von „NursePlan“ (Reuschenbach, 2008): Hinter den Türen geht es zu verschiedenen Patienten, im Rollschrank sind weiterführende Informationen anklickbar (mehr unter <http://www.nurseplan.de/>).

tung der Instruktions- oder Testübungsphase. Für eine adaptive Itemfolge sollten die Tests allerdings nach dem Prinzip der *probabilistischen Testtheorie* aufgebaut sein. Bisher gibt es unseres Wissens nach keine Computersimulation, die wirklich adaptiv im Sinne der probabilistischen Testtheorie ist. Die Ursache dafür ist, dass sich eine Simulation meistens nur schwer in Teile zerlegen lässt, die klein genug sind, um adaptives Testen ökonomisch durchführen zu können. Eine Zergliederung in adaptive Einzelitems widerspricht daher teilweise dem Vorteil von integrierten Handlungsabläufen.

Eine weitere Begründung für Computersimulationen ist der Mehrwert der Datenerhebung. Neben den ausgewählten Antworten fallen nämlich auch noch Daten zur *Reaktionsschnelligkeit* an, die ebenfalls einen diagnostischen Wert haben. Passend bringen es Olson-Buchanan und Drasgow (1999) auf den Punkt:

A pilot selection program would certainly want to avoid an individual who could answer paper-and-pencil spatial skill items very accurately, but was too slow to use a joystick to guide a simulated airplane in real time. Similarly, the veracity of a correct response might be questioned if information critical to the solution of the problem was not presented to the examinee or if the examinee answered an item in just a few millisecond (i. e., without reading the question). (S. 3)

Die bei der Computersimulation anfallenden Reaktionszeiten erlauben somit auch Aussagen über die Validität der Messungen. Weitere Vorteile, die generell mit computergestützter Diagnostik zusammenhängen, sind die leichtere Durchführbarkeit auch als Gruppenverfahren und via Online-Tests sowie die automatische Datenerhebung und -auswertung.

3.1.2 Systematisierung

Gray (2002) systematisiert Computersimulationen anhand der vier Merkmale „fidelity“, „correspondence“, „tractability“ und „engagement“. Die ersten beiden beziehen sich auf Systemeigenschaften, die anderen beiden auf die Versuchsperson bzw. die Interaktion zwischen Versuchsperson und System:

1. *Fidelity* bezeichnet die Passung zwischen Untersuchungsabsicht und den Möglichkeiten des Szenarios. Aber: „The fidelity of even the most complex simulation is relative to the question being asked“ (Gray, 2002, S. 208). Es werden fünf Kategorien unterschieden: (a) „hi-fidelity simulations of complex systems“, z. B. Flugsimulatoren; (b) „hi-fidelity simulations of simple systems“, z. B. der Umgang mit einem GPS-System; (c) „scaled worlds“: Szenarien, bei denen relevante Systemparameter der Umwelt abgebildet werden; (d) „synthetic environments and microworlds“: hier handelt es sich um künstliche Welten, die keine Vorerfahrung notwendig machen. Ein Einfluss von Vorerfahrung aufgrund von Ähnlichkeiten mit bekannten Problemen kann jedoch nie ausgeschlossen werden; (e) „laboratory tasks and simulated task environments“: Simulationen ohne direkten Realitätsbezug.
2. *Correspondence*: Hiermit ist die Passung zwischen Systemparametern des Szenarios und den tatsächlichen Parametern in der Realität gemeint. In der Forschung steht man vor der Frage, ob relevante Systemparameter teilweise abgebildet werden sollen oder ein genaues Abbild der Wirklichkeit erwünscht ist. Die Spanne reicht bei dabei von „one aspect of many systems“ bis zu „many aspects of one system“. Die Korrespondenz bestimmt damit auch die Übertragbarkeit der Befunde auf ähnliche oder gleiche Anforderungen: „(...) the higher the correspondence of a simulated task environment to one system, the less the research on that simulated task environment can be generalized to other systems“ (S. 216). Wird ein Szenario oder ein Problem artifiziell erzeugt, das typische Kennzeichen vieler Probleme abbildet („one aspect of many systems“), dann steigt die Generalisierbarkeit, während gleichzeitig die Korrespondenz sinkt. Umgekehrt steigt mit hoher Korrespondenz die Komplexität, während die Lenkbarkeit sinkt. „For data collection, high correspondence may be the enemy of tractability“ (S. 222).
3. *Tractability*: Die Lenkbarkeit wird durch die Kontrollierbarkeit (von „difficult“ bis „tractable“) und durch die Trainierbarkeit bestimmt: „The more training subjects require before they can use a simulated task environment, the less tractable it is“ (S. 211). Ob ein Problem leicht handhabbar ist, hängt von der Korrespondenz zwischen Szenario und Wirklichkeit und der Komplexität ab. Nach Gray ist Komplexität gerade durch mangelnde Kontrollierbarkeit und Lenkbarkeit gekennzeichnet.

4. *Engagement*: Der Bearbeitung eines Szenarios kann von „boring“ bis „engaging“ eingestuft werden und hängt mit dem Aufforderungscharakter des Szenarios zusammen. Dieser kann aber ebenfalls nur aus dem Verhalten der Testpersonen oder durch deren Befragung erschlossen werden. „A simulated task environment is engaging to the degree to which it involves and occupies the participants; that is, the degree to which they agree to take it seriously“ (S. 217).

Andere Systematisierungsversuche von Simulationen bzw. komplexen Szenarien beziehen sich z. B. auf die Unterscheidung realitätsnaher Szenarios von solchen, die auf der Basis formaler Kriterien erstellt wurden (vgl. Buchner, 1995; Funke, 2003, Kap. 5): Die erstgenannte Kategorie bemüht sich um eine möglichst realistische Abbildung eines bestimmten Gegenstandsbereichs und verwendet dazu alle formalen Möglichkeiten, die überhaupt zur Verfügung stehen (Beispiel „Lohhausen“; Dörner, Kreuzig, Reither & Stäudel, 1983). Dies schafft eine Annäherung an realistische Verhältnisse für den Preis der mangelnden Optimierbarkeit und damit verbundenen Bewertungsproblemen (z. B. aufgrund nicht linearer Funktionen und des deterministischen Chaos, das auftreten kann). Die zuletzt genannte Kategorie verzichtet auf einen Teil der Realitätsnähe zugunsten der Verwendung klarer Formalismen (z. B. lineare Strukturgleichungen, finite Automaten), innerhalb derer z. B. wohldefinierte Kennwerte zur Leistungsmessung zur Verfügung stehen (Beispiel „Sinus“, Funke, 1993b; Beispiel „MicroDYN“ und „MicroFIN“, Greiff & Funke, 2010).

Die wohl umfassendste Taxonomie zur Klassifikation komplexer Szenarios („TAKS“) stammt von Wagener (2001). Er unterscheidet zwischen den sechs Eigenschaftsbereichen (1) inhaltliche Einkleidung, (2) Bedienungsinterface, (3) formale Systemstruktur, (4) funktionale Systemstruktur, (5) Informationsangebot und (6) Auswertung. Alle sechs Bereiche sind durch vielfältige Einzelmerkmale ausdifferenziert, auf die hier aus Platzgründen nicht eingegangen werden kann. Hilfreich ist diese Systematisierung bei der Einordnung empirischer Befunde, beim Vergleich existierender Systeme und bei der Konstruktion neuer Szenarien.

3.1.3 Bewertung des Ansatzes

Den in der Anwendungsbegründung genannten Argumenten für den Einsatz von Computersimulationen stehen folgende Nachteile gegenüber:

- Die Erstellung von Computersimulationen ist aufwendig und meist auch teuer. Dies gilt insbesondere dann, wenn komplexe Stimuli wie Bilder und Videos eingebunden werden und es sich um ein adaptives Testverfahren

handelt. Hierzu muss eine große Anzahl an Items entworfen werden, die in der Simulation sinnvoll aneinander gereiht werden und dabei testtheoretische Standards erfüllen müssen. Werden in den Simulationen Videos integriert, kann die dadurch eventuell gegebene Abhängigkeit des Filmmaterials zusätzliche Probleme schaffen.

- Steht die Realitätsnähe bei der Computersimulation im Vordergrund, muss eine vorherige Anforderungsanalyse die Passung zwischen Umwelt- und Systemanforderungen sichern (vgl. Funke, 2010).
- Es fehlt an Standardisierung im Hinblick auf die Art und den Umfang der Simulationen. Für verschiedene Bereiche und ähnliche psychologische Konstrukte sind verschiedene Simulationen im Einsatz. Im „Zoo“ der komplexen computersimulierten Problemstellungen befinden sich Exemplare, deren Familienähnlichkeit z. T. schwer zu beurteilen ist – selbst hinter Szenarien mit gleichem Namen („Tailorshop“) verbergen sich unterschiedliche Varianten.
- Da komplexe computerpräsentierte Stimuli meist auch mit komplexen Eingabeoptionen einhergehen, muss der Einfluss von Computervorwissen kontrolliert werden. Hier könnte sich ein massiver Generationeneffekt verbergen.
- Im Falle einer adaptiven Testgestaltung ist die Vergleichbarkeit der Testwerte durch idiosynkratische Bearbeitungsverläufe erschwert. Bereits mit den ersten Eingriffen erzeugt jede einzelne Versuchsperson ihre eigene Umwelt und ist damit nur noch bedingt mit anderen Personen vergleichbar, die das gleiche Szenario bearbeiten.
- Es fehlt meist an einer theoriegeleiteten Auswertung, die es ermöglicht, die gewählten Items und die Schnelligkeit bei der Wahl der Items gleichzeitig bestimmten Faktoren zuzuordnen. Für viele Leistungsvariablen, die bei der Computersimulation erfasst werden können, fehlt die substanzwissenschaftliche und testtheoretische Basis.

Für die meisten der genannten Probleme existieren allerdings Lösungsvorschläge, die bei Funke (2003, Kap. 4 und 5) ausführlicher beschrieben sind. Generell kann man festhalten, dass die hier angeführten kritischen Punkte die Brauchbarkeit dieses diagnostischen Zugangs nicht grundsätzlich in Frage stellen.

3.2 Digitale Lernspiele („serious games“)

Mit dem Begriff der digitalen Lernspiele („serious games“) wird die Nutzung von hochwertiger Spielsoftware zu pädagogischen Zwecken aller Art verstanden. Moderne Rollenspiele wie „Final Fantasy“ oder Ego-Shooter wie „Quake“ bzw. „Counter-Strike“ besitzen einen Softwarekern, der als „Video-Engine“

bezeichnet wird und eine hochwertige 3D-Grafik erzeugt, mittels derer sich eine Spielfigur durch eine komplex gestaltete Computerwelt bewegen kann. Diese Video-Engine kann mithilfe von Zusatzsoftware (*software development kit*, SDK) mit neuen Inhalten gefüttert werden. Solche Modifikationen kommerzieller Produkte werden in der Computerszene „mods“ genannt, der sehr aufwendige Prozess des Neugestaltens „modding“. Während die ersten „mods“ vor allem dazu dienten, neue Umgebungen in anderen Schwierigkeitsstufen für die Spieler zu schaffen (und damit das Interesse an einer bestimmten Spielumgebung länger aufrechtzuerhalten), hat sich in den letzten Jahren eine alternative Szene herausgebildet, die die professionellen Video-Engines für edukative Zwecke heranzieht (Computerspiel-Pädagogik).

Neben der Nachnutzung bestehender Video-Engines werden zunehmend auch eigenständige Programme entwickelt, um digital vermitteltes Lernen zu ermöglichen (Bopp, 2006, 2007). Eine neue Branche der Software-Industrie ist hier entstanden, die für Psychologen Arbeitsplätze mit diagnostischem Potenzial bietet (vgl. <http://seriousgamessource.com/>).

3.2.1 Anwendungsbegründung

„Serious gaming“ macht sich die Tatsache zunutze, dass kommerzielle Computerspiele geradezu suchtbildende Eigenschaften aufweisen. Wenn die Lernbereitschaft, die Menschen zum Erwerb komplexer Wissensstrukturen und aufwendiger Strategien in virtuellen Realitäten aufweisen, auf sinnvollere Lerngegenstände umgelenkt werden kann, ist ein großes Potenzial gewonnen. Folgt man den Ausführungen von Prensky (2000), haben die nach 1974 „digital Geborenen“ (digital natives), die er „twitchspeed generation“ nennt, unter anderem durchschnittlich 10.000 Stunden Erfahrung mit Video- und Computerspielen, etwa ebenso viel Erfahrung mit der Nutzung von Handys, sie haben etwa eine viertel Million E-Mails erhalten und versandt, aber nur 2000 bis 3000 Stunden Leseerfahrung. Mit dem Einsatz von „serious gaming“ kommt man an eine Generation junger Menschen heran, für die der Umgang mit derartigen Umwelten in hohem Maß selbstverständlich ist.

3.2.2 Systematisierung

Eine Systematisierung von „serious games“ kann nach verschiedenen Kriterien erfolgen (z. B. der verwendeten Video-Engine). Hier wird eine Systematisierung nach inhaltlichen Kriterien vorgenommen, die sich auf die Einsatzgebiete beziehen.

3.2.2.1 Schule

Das für amerikanische Mädchen gedachte Spiel „The Adventures of Josie True“ (<http://www.josietrue.com/>) ist ein von der „National Science Foundation“ finanziertes Projekt, mit dem Kinder an Mathematik und Naturwissenschaften herangeführt werden. Die spielende Person wird auf die Suche nach einer verschwundenen Lehrerin geschickt und muss dabei in Art eines Detektivspiels Rätsel lösen und Aufgaben bewältigen, die einschlägiges Fachwissen und Geschichtskenntnisse erfordern.

Neben der Förderung des wissenschaftlichen Interesses gibt es auch ein Anwendungsbeispiel aus dem Bereich der politischen Bildung namens „Darfur is Dying“ (<http://www.darfurisdying.com/>). Hierbei handelt es sich um ein online verfügbares Videospiel, bei dem man in die Rolle des 14-jährigen Elham schlüpft, der in der Sahelregion Darfur Wasser holen und dabei feindlichen Milizen ausweichen muss, wenn er überleben will. In dem Spiel, das vom Musiksender MTV in Zusammenarbeit mit der „Reebok Human Rights Foundation“ und der „International Crisis Group“ erstellt wurde, muss Elham Wasser organisieren, um damit auf einem kleinen Feld die Saat zu bewässern oder um Lehmziegel herzustellen, die für den Wiederaufbau von Hütten dienen, die von den Milizen zerstört wurden. Außerdem müssen knappe Lebensmittel und Medikamente ergattert werden. Und auch im Lager sind die Menschen vor Angriffen der Milizen nicht sicher.

Auch im medizinischen Sektor werden pädagogische Elemente genutzt. Mit den 2008 erscheinenden Videospielen „Nanoswarm: Invasion from inner space“ und „Escape from Diab“ unternimmt die Firma Archimage in Verbindung mit dem „Children’s Nutritional Research Center of Houston’s Baylor College of Medicine“ sowie mit Unterstützung des „National Institute of Diabetes and Digestive and Kidney Diseases“ den Versuch, hochattraktive Spielumgebungen zu schaffen, in denen man den Kampf gegen das Diabetes-Monster führt. Die herkömmlichen Spielelemente (Jump’n’Run, Battle-Fights etc.) werden durch die semantische Einkleidung zu einer Lernumgebung über Diabetes-relevante Fakten, die insbesondere Kinder und Jugendliche ansprechen sollen.

3.2.2.2 Arbeitswelt

Ein riesiges Potenzial für „serious games“ besteht im Bereich der Arbeit. Training, Personalauswahl und -fortbildung setzen bereits seit längerem diese technischen Mittel ein. Ein Beispiel: Der spanische Spiele-Entwickler „Exelweiss“ (<http://www.exelweiss.com/>) hat das Spiel „Worky’s Challenge“ vorgelegt, bei dem die lernende Person sich mit Arbeitsschutzmaßnahmen im Kampf gegen

einen gewissen „Dr. Fatality“ beschäftigt. Auftraggeber dieses Abenteuerspiels war eine Versicherung in Verbindung mit der Bezirksregierung in Valencia, deren Interesse darin bestand, die hohen Unfallzahlen bei niedrig qualifizierten Bauarbeitern zu senken. Die Comicfigur des muskulösen Helden „Worky“ muss verschiedene Herausforderungen bewältigen und schafft damit Lernsituationen, in denen z. B. Vorschriften des Arbeitsschutzes transportiert werden können. Einsatzmöglichkeiten für den diagnostischen Bereich zeigt das von Siemens zusammen mit dem Lehrstuhl für Methodenlehre, Diagnostik und Evaluation der Ruhr-Universität Bochum entwickelte Online-Spiel „Challenge Unlimited“. Testpersonen müssen hierbei als „Cyber-Consultants“ die Stadt Nouvopolis retten. Nach dem Spiel können die Personen die Daten für die Recruiting-Abteilung von Siemens freischalten lassen (Wild, Fontaine & Schaffner, 2001).

3.2.2.3 Militär

Für den militärischen Bereich berichtet Stone (2005) über die Entwicklung des Spiels „America's Army“, das auf der Engine des 1998 erschienenen Ego-Shooters „Unreal“ der Firma Epic Megagames basiert. Es sollte jungen Amerikanern die Möglichkeiten einer militärischen Karriere aufzeigen und wurde kostenlos verteilt. Nach Aussage der US-Army ist es ein kostengünstiges Rekrutierungswerkzeug: Etwa 4,7 Millionen Benutzer sollen die virtuelle Grundausbildung bereits abgeschlossen haben (Stand Juni 2007; Quelle: http://de.wikipedia.org/wiki/Americas_Army).

Die Ähnlichkeit zwischen dieser Software und derjenigen, die im militärischen Alltag zum Einsatz kommt, ist natürlich unbekannt. Allerdings wird befürchtet, dass durch die weite Verbreitung dieses Spiels und seinem hohen Realitätsgehalt die Schwelle zum Töten von Menschen herabgesetzt werden könnte.

3.2.2.4 Umwelt

In Zeiten globalen Wandels und der damit einhergehenden Klimaveränderungen haben sich Lernspiele auch des Themas Umweltschutz angenommen. Vom britischen Nachrichtenkonzern BBC ist ein Klimaspiele namens „Climate Challenge“ online zugänglich (http://www.bbc.co.uk/sn/hottopics/climatechange/climate_challenge/), bei dem man als Spieler in die Rolle eines Staatenlenkers schlüpfen soll, der zehn Perioden zu je 10 Jahren im Zeitraum zwischen 2000 und 2100 steuert. Dabei trifft man auf verschiedenen Ebenen (internationale Ebene, Wirtschaft, Handel, nationale Regeln, Haushaltsebene) Entscheidungen und sieht deren ökonomische und ökologische Auswirkungen. Dieses Feedback wirkt sich auf die Stimmung der Bürger aus, die den Machthaber

sehr schnell seines Amtes entheben (und damit die Spielzeit beenden), wenn keine vernünftigen Pläne gemacht werden. Auch wenn das Spiel sehr lehrreich gestaltet ist, bleibt es diagnostisch hinter dem Wünschbaren zurück: ein differenzierteres Feedback wäre z. B. ein wünschenswerte Weiterentwicklung.

3.2.3 Bewertung des Ansatzes

Der Ansatz digitaler Lernspiele bietet nicht nur medienpädagogisch ein reizvolles Feld (vgl. Bopp, 2006, 2007), sondern ist natürlich auch in diagnostischer Hinsicht von besonderem Interesse und damit vielversprechend. Eine „spielerisch“ erfolgende Diagnostik garantiert hohe Motivation durch die teilnehmenden Personen, bietet ein hohes Maß an Eigenbeteiligung und offenbart damit eine Vielzahl diagnostisch interessanter Parameter: Im spielerischen Handeln kommen die verschiedenen Facetten von Kognition, Emotion und Motivation zum Tragen, die bei entsprechender Spielgestaltung sowohl um die soziale Dimension als auch um die Dimension der Einstellungen und Werte angereichert werden können.

Wie Kauchak (2007) schreibt, ist der Bedarf an „serious games“ auf Seiten des Militärs in Ländern wie Australien, Dänemark, Großbritannien, Kanada, Neuseeland, Spanien und den USA ungebrochen. Die erhoffte Unterstützung bei der Ausbildung von Personal setzt natürlich eine entsprechend ausgestaltete Diagnostik-Komponente voraus. Daran wird auf militärischer Seite offensichtlich mit Hochdruck gearbeitet; es wäre wünschenswert, wenn auch auf Seiten der Pädagogischen Psychologie entsprechend nicht-militärische Forschungen intensiviert werden könnten. Für die Vorselektion im Rahmen der Personalauswahl sichert eine Einkleidung von Diagnostika in solche Spiele eine hohe Teilnahmemotivation, große Teilnehmerzahlen und ist eine gute Möglichkeit zur Imagebildung für das Unternehmen.

3.3 Virtuelle Welten

Ein Versuch, die Anforderungen und die System-Person-Interaktion noch näher an reale Gegebenheiten anzupassen, stellt die Verwendung virtueller Welten dar. Während Computersimulationen, selbst wenn sie 3D-Technologien nutzen, stets auf visuelle Stimuli reduziert sind, gehen virtuelle Welten („virtual reality“, VR) noch einen Schritt weiter: Auch die Lage sowie Bewegung des eigenen Körpers werden in dieser Welt nachgebildet: „VR integrates real-time computer graphics, body tracking sensors, audio/visual display and sensory input devices to immerse a participant in an interactive computer-generated virtual environment (VE) that changes in a natural way with head and

body motion“ (Wiederhold & Rizzo, 2005). Wesentliche Komponenten, die auch für die Diagnostik eine Rolle spielen, sind:

- *Multisensorik*: Neben visuellen können auch akustische und haptische Aspekte der Umwelt abgebildet werden und adaptiv auf das Verhalten angepasst werden.
- *Immersion*: Die Testperson „taucht ein“ in die virtuelle Welt, kann in ihr agieren und verliert dabei zunehmend das Gefühl dafür, dass es sich um eine Simulation handelt. „HANS“, der Heidelberger Anästhesie-Simulator, erlaubt etwa angehenden Ärzten, mit simulierten Zwischenfällen in der Anästhesie sehr realistisch umzugehen (Schaper, Schmitz, Graf & Grube, 2004). Da etwa zwei Drittel aller Zwischenfälle in der Anästhesie auf menschliches Versagen zurückgeführt werden können, ist das Training in der virtuellen Welt von HANS ein wichtiger Ausbildungsbestandteil geworden (vgl. auch Dieckmann, 2005).
- *Verhaltensmessung*: Virtuelle Welten ermöglichen die direkte Erfassung von Verhalten, während in der Computersimulation die von der Testperson gewählten Optionen immer noch verhaltensstellvertretend sind. Bei den VR interagiert die Testperson mit der virtuellen Umwelt. Verhalten verändert adaptiv die virtuelle Umwelt und damit auch die relevanten Anforderungen.

Das Ziel derzeitiger Entwicklungen ist es, die Anwendung zunehmend realistischer zu gestalten. Methodisch bedeutet dies, dass die Leistungen, die eine Person in der VR zeigt, immer klarer auf ihre tatsächliche Leistung und nicht auf die Bedienbarkeit zurückzuführen sind. „The goal is to build (virtual) environments that minimize the learning required to operate within them, but maximize the information yield“ (Stanney, Mourant & Kennedy, 1998). Anforderung an die Bedienung („navigational complexity“) – z. B. die Anforderung, sich von einem Punkt zum anderen zu bewegen – und der Grad der Umweltpassung („degree of sense of presence“) sind die beiden Dimensionen, die bei der Messung von Leistungen in VR beachtet werden müssen, da sie die tatsächliche Lösung von Aufgaben überlagern können. Faktoren, die bei der Erstellung von VR als bedeutsame Anforderungen benannt wurden, sind – eine entsprechende Standardisierung vorausgesetzt – ebenso denkbare diagnostische Zielvariablen, beispielsweise räumliche Orientierung (insbesondere das sog. „dead reckoning“ = Standortbestimmung), visuelle Genauigkeit, geografische Orientierung, Gleichgewichtsirritation/-funktion, Fähigkeiten, im virtuellen Raum Gegenstände zu bewegen und zu verändern, Reaktionsfähigkeit.

Interindividuelle Unterschiede können sich in drei Bereichen ergeben: (1) im Input, also z. B. in der Erkennung von Objekten in der VR, (2) in der Verarbeitung, z. B. in entsprechenden Wahrnehmungsstilen, (3) im Output, d. h. im

Reagieren in der VR. Zur Prüfung solcher Faktoren, die die Leistungsmessung im VR beeinflussen, wurde von Lampton et al. (1994) eine „Virtual Environment Performance Battery“ (VEPAB) entwickelt. Hiermit ist es möglich, die durch das Handling bedingten Einflüsse von den relevanten Leistungen in der VR zu separieren. Weiterhin ist auch das erlebte Präsenzerleben, also das subjektive Eintauchen in die virtuelle Welt, eine diagnostisch relevante Größe (Hofmann, 2002). Hier wird beispielsweise der Grad der Feldabhängigkeit als wichtiger Einflussfaktor diskutiert (Hecht & Reiner, 2007).

3.3.1 Anwendungsbegründung

Die Methode ist sinnvoll, wenn es darum geht, die Wirkung komplexer und alltagsnaher Stimuli/Items zu erfassen und/oder die Interaktion mit der virtuellen Welt zu messen. Riva (1998) nennt als wesentliche Vorteile einer Erhebung in oder mit einer virtuellen Realität folgende vier Aspekte: (1) Personen verhalten sich aufgrund der Immersion realitätsnäher. (2) Die VR ermöglichen ein hohes Maß an Interaktivität. Insbesondere dann, wenn die motorische Interaktion mit den Stimuli und die visuell-haptische Wahrnehmung im Mittelpunkt stehen, sind VR von Vorteil. (3) Relevante Verhaltensmaße können sich auf beinahe alle motorischen Funktionen beziehen, so ist beispielsweise eine Analyse des Gangbildes möglich. (4) Abweichend von der realen Welt können unaufdringlich Hinweisreize gegeben werden, die ein adaptives Testen ermöglichen. Gleichzeitig sind die situationalen Einflussfaktoren systematisiert und metrisch erfassbar. Die Methode bietet sich insbesondere dann an, wenn eine Messung im Feld nicht möglich ist.

3.3.2 Beispiele

In den psychologischen Datenbanken taucht der Begriff „virtual reality“ erstmals in einem Artikel von Tart 1990 auf. Der Autor weist auf die vielfältigen Möglichkeiten der VR hin: „Computer-generated virtual realities offer intriguing possibilities for developing diagnostic, inductive, psychotherapeutic, and training techniques that can extend and supplement current ones“ (S. 222). Heute hat diese Technik in fast alle diagnostische Bereiche Einzug gehalten. In der seit 1998 erscheinende Zeitschrift „CyberPsychology & Behavior“ werden beinahe monatlich neue innovative diagnostische Möglichkeiten mittels VR vorgestellt.

Anfang der 1990er Jahren wurden VR vor allem als therapeutisches Medium genutzt. Beispielhaft ist die Modifikation von Angst bei Phobikern zu nennen (vgl. <http://www.vrphobia.com>) zu nennen. Die angebotenen Bereiche umfas-

sen z. B. Flugangst, Platzangst, Spinnenangst, Höhenangst oder Sozialangst. In allen Bereichen liegt virtuelles Stimulusmaterial vor, das die angstausslösenden Reize unter die Kontrolle des Patienten stellt. Entsprechende Instrumentarien werden schon zur Desensibilisierung bei Phobien verwendet, denn die Möglichkeit, sich in der Welt mit tatsächlichen Körperbewegungen dem Angst auslösenden Objekt zu nähern, ist realistischer als das Heranzoomen eines Bildes auf dem Bildschirm (Côté & Bouchard, 2005).

Im Bereich der Biopsychologie sind parallele Untersuchungen mittels funktionseller Magnetresonanztomografie (fMRI; Wiederhold & Rizzo, 2005) oder physiologischen Messung (Mühlberger, Bühlhoff, Wiedemann & Pauli, 2007) bei weitgehender Kontrolle der Stimuli und einer Standardisierung von Bewegungseinflüssen möglich. Im Bereich der Leistungsdiagnostik macht man sich die Möglichkeit zunutze, dass Personen in der VR tatsächlich Bewegungen im Raum vornehmen können. So kann die Technik zur Erfassung des räumlichen Vorstellungsvermögens oder des räumlichen Gedächtnisses genutzt werden (Gould et al., 2007).

Inzwischen liegt eine standardisierte Testbatterie zur Diagnostik von Lern- und Gedächtnisleistungen vor, der „Virtual Reality Cognitive Performance Assessment Test“ (VRCPAT) von Parsons, Rizzo, Bamatre und Brennan (2007).

Ein weiteres wichtigstes Einsatzgebiet ist die Neuropsychologie. So ist es beispielsweise möglich, Personen nach einem Schlaganfall beim „Gehen“ in der VR mit unterschiedlichen Weggefällen oder virtuellem Schnee zu konfrontieren. Die Reize können im Gegensatz zur Realität adaptiv dosiert werden und gefährden die Person nicht (Nyberg et al., 2006). Auch zur Diagnostik des Neglects werden gehäuft VR eingesetzt (z. B. Broeren, Samuelsson, Stibrant-Sunnerhagen, Blomstrand & Rydmark, 2007). In einer von Japanern entwickelten VR (Baheux, Yoshizawa, Seki & Handa, 2006) können Testpersonen beispielsweise eine virtuelle Sushi-Bar betreten. Die Analyse des Auswahlverhaltens in dieser virtuellen Bar ermöglicht eine Differenzialdiagnostik verschiedener Neglect-Formen.

3.3.3 Bewertung des Ansatzes

Durch die Verbesserung der Computerleistungen und Visualisierungen (z. B. VR-Brillen), aber auch durch wichtige Grundlagenforschung zur Gestaltung von VR hat diese Technologie in den vergangenen Jahren an Bedeutung gewonnen. Dennoch sind bei der Erstellung von VR große Herausforderungen zu meistern. Beispielhaft erwähnt sei die Abbildung von optischen Bewegungsmustern, die Vermittlung eines dreidimensionalen Eindrucks, die realistische

Einbindung von räumlichen Klangerlebnissen und von haptisch-kinästhetisch-olfaktorischen Reizen. Allein für die Darbietung visueller Reize gibt es vielfältige Möglichkeiten, wie beispielsweise „head mounted displays“, Großprojektionen oder Mehrseitenprojektionen (CAVES oder Holodeck genannt). Die Art der verwendeten Darbietungstechnik kann die Leistung beeinflussen (Sharples, Cobb, Moody & Wilson, 2008). Teilweise fehlt es hierfür noch an psychophysischer Grundlagenforschung. Ebenso gibt es zur multimodalen Interaktion der Stimuli, die gerade die „Lebendigkeit“ der VR ausmacht, derzeit kaum Forschung. Weiterhin stellt die Sicherung von standardisierten Bedingungen, ähnlich wie bei anderen Computersimulationen, die Testentwicklung vor große Herausforderungen. Die Auswahl geeigneter Anforderungen und die Bewertung von Handlungen ist auf dem Hintergrund individueller Vorerfahrungen ein bekanntes Problem, das immer dann aufkommt, wenn in Testverfahren Komplexität zugelassen und eine Interaktion zwischen Testperson und Test ermöglicht wird.

Als weitere Probleme der derzeitigen VR sind negative Auswirkungen auf die Testpersonen zu nennen. Stanney et al. (1998) berichten von Phobien, Übelkeit, Ängsten und der sogenannten Cybersickness („simulation sickness“), die die Testung überschatten können. Der Einsatz solcher VR für diagnostische Zwecke sollte in Anbetracht der derzeitigen hohen Investitionskosten daher wohlüberlegt sein.

3.4 Videotests

Videotests nutzen als „Items“ Filmmaterialien. Dies können entweder klassische Videoaufnahmen, digitale Filmaufnahmen oder auch computeranimierte Szenen sein. Durch die zunehmende Annäherung computeranimierter Szenen an „echte“ Filme ist hier keine klare Grenze zu ziehen. Entscheidend ist die Aktivierung interner Schemata und Skripte, die durch das Medium erreicht wird: So kann eine Trickfilm-Szene mit dem „HB-Männchen, das in die Luft geht“, ebenso viel Heiterkeit auslösen wie eine Szene aus einer Luis-de-Funes-Komödie – auch reduzierte Modelle können starke Wirkung haben.

Es kann sich um klassische Filmmaterialien (DVD, CD, Video) handeln, die mit entsprechenden Rekordern getrennt vom Test präsentiert werden, oder um digitale Filme, die auf dem Computer präsentiert werden und hier auch entsprechende Eingaben ermöglichen (z. B. Toobert, Glasgow, Desalvo & Strycker, 1998). Hierbei sind auch adaptive Tests möglich, bei denen das Auswahlverhalten der Testperson die weiteren Videosequenzen bestimmt. Videos können aber nicht nur als Items, sondern auch zur Instruktion genutzt werden.

3.4.1 Anwendungsbegründung

Anwendungen finden sich insbesondere dort, wo komplexe Situationen zu analysieren und entsprechende Entscheidungen oder Einschätzungen („situational judgement“) zu treffen sind. Weiterhin dort, wo soziale Situationen richtig bewertet und entsprechend gehandelt werden soll. Eine Schilderung der Szenen in textlicher Form würde eine Fokussierung auf die verbal relevanten Aspekte bedeuten und für die Wahrnehmung und Interpretation der Szenen wenig Raum lassen. Videotests sind daher auch für die Testung von Personen mit geringer Lesekompetenz geeignet. Chan und Schmitt (1997) konnten durch den Vergleich eines Papier-Bleistift-Tests mit einem Videotest zur Situationseinschätzung zeigen, dass der Einfluss des Leseverständnisses einen Teil der Varianz im Papier-Bleistift-Test aufklärt. Außerdem wird die Augenscheinvalidität von Videotests von den Testpersonen höher eingeschätzt.

3.4.2 Beispiele

Schon in den 90er Jahren wurden erste Versuche unternommen, durch die filmische Darstellung realistischer Problemfälle aus dem betrieblichen Alltag die *situative Einschätzungsfähigkeit* von Führungskräften zu bestimmen. Weiterhin stammen aus dieser Zeit multimediale Postkorb-Verfahren mit Videoeinspielungen, die beispielsweise die Planungsfähigkeit von Bediensteten am Flugschalter testen (vgl. Drasgow, Olson-Buchanan & Moberg, 1999). Zur Messung *diagnostischer Fertigkeiten bei medizinischen Notfällen* kommen Videotests häufig zur Anwendung (Lievens, Buyse & Sackett, 2005). Hierbei geht es um die Messung der Risikoeinschätzung und der diagnostischen Fähigkeiten. Computer-Video-Szenarien findet man insbesondere im Bereich der *Verkehrspsychologie*, z. B. zur Bestimmung der Reaktionszeit, der präferierten Geschwindigkeit oder der Detektion kritischer Verkehrssituationen (Schiff et al., 1994). Gerade zur Personalauswahl in sozialen Berufen ist die Videosimulation gut geeignet, um interpersonelle Skills aufzudecken (Lievens et al., 2005; Schoech, 2001). Ebenso sind adaptive Videotests möglich, wie z. B. der „Interaktive Videotest“ (IAV) zur Messung der Konfliktlösefähigkeit (Drasgow et al., 1999).

Der von Funke und Grube-Unglaub (1993) beschriebene „Skript-Monitoring-Test“ (SMT) kann als ein Beispiel für die Erfassung des Planungsvermögens und der Handlungsregulation mittels Video im Bereich der Neuropsychologie bei frontallhirngeschädigten Patienten angesehen werden. Filmisch vorgeführt werden alltägliche Routinehandlungen in Form kleiner Videoclips, die vom Patienten zu überwachen und auf das Vorliegen eventueller Fehler wie auch auf

die mögliche Fortführung einer angefangenen Handlungssequenz zu prüfen sind. Die gemessenen Dimensionen „Planüberwachung“, „Fehlerdiagnostik“ und „Abfolgen erkennen“ erwiesen sich in der genannten Pilotstudie als trennscharfe Indikatoren, die herkömmlichen Tests (wie z. B. „Turm von Hanoi“) überlegen waren. Gerade im Hinblick auf berufliche Wiedereingliederung von Reha-Patienten ist das realistische Videomaterial von hohem diagnostischen Wert. Wie Grube-Unglaub und Funke (1995) betonen, ist die hohe ökologische Validität ein zentrales Ziel der Testentwicklung gewesen.

In jüngster Zeit werden Videotests zur Erfassung von Kommunikationsfähigkeit im klinischen Kontext eingesetzt. Die angehenden Therapeutinnen und Therapeuten oder Psychologiestudierenden sind hierbei mit Videosequenzen von Klientinnen bzw. Klienten konfrontiert, die entsprechend dem therapeutischen Grundverständnis von der Testperson zu beantworten sind. Rosengren, Baer, Hartzler, Dunn und Wells (2005) beschreiben beispielsweise einen solchen Tests zur Messung der „motivational interviewing (MI) skills“ bei Abhängigen. Kuntze, van der Mole und Born (2007) entwickelten den „Communication Skills Progress Test“ (CSPT), bei dem die Testpersonen mit 42 Kurzvideos von Klienten konfrontiert sind und aus dem Frageverhalten im Anschluss an die Videopräsentation beispielsweise die Fähigkeit zur Reflexion und Paraphrasierung strukturiert erfasst werden kann.

Die Wege zur Erstellung des Filmmaterials für den Videotest systematisieren Drasgow et al. (1999) in drei Stufen:

1. Umsetzung existierender *Handlungsskripte* aus simulierten Rollenspielen.
2. Auswahl der relevanten Szenen nach einer vorherigen *Anforderungsanalyse*; z. B. können Interviews im Sinne der „Critical Incident Technique“ (Flanagan, 1954) geführt werden. Diese führen zur genauen verhaltensnahen Darstellung der Situation, der optimalen Lösung und zu einer Bewertung alternativer Lösungsmöglichkeiten. Diese müssen dann in ein Drehbuch und schließlich in einen Film überführt werden. Die Verhaltensmöglichkeiten der Testperson müssen aufgrund von Experteninterviews eingeschätzt und ein entsprechendes Scoring entwickelt werden. Die angebotenen Alternativen können entweder in schriftlicher Form oder ebenfalls durch Filmsequenzen dargeboten werden.
3. Im Sinne einer *theoriegeleiteten Itemauswahl* können Beispielinstanzen entwickelt werden. Abschließend ist jedoch ebenfalls ein Expertenrating der Filmausschnitte notwendig.

Häufig finden sich Mischmodelle, z. B. die Erstellung der „Items“ aufgrund von Expertenratings und die anschließende theoretische Einordnung.

Verbesserte Komprimierungsmöglichkeiten und die gleichzeitig wachsenden Speicherplatzkapazitäten erleichtern die Speicherung von Filmen auch in digitaler Form. Damit wird die Bedeutung von Videotests via Smartphones oder Pocket-PCs sicherlich an Bedeutung gewinnen. Durch immer schnellere Übertragungstechniken werden sicherlich auch eine Online-Einspielung von digitalen Filmen und eine begleitende Online-Messung möglich sein.

3.4.3 Bewertung des Ansatzes

Natürlich ist die Art des Filmmaterials für die gewünschte Realitätsnähe von hoher Bedeutung. Hier ist vorab zu klären, ob reale Fälle gefilmt werden (z. B. zur Einschätzung von Gefahrenpotenzialen oder der besonderen Kommunikationssituation), ob eine Schlüsselszene von Laien oder Schauspielern nachgestellt wird oder ob animierte Videos verwendet werden. Für die Erstellung entsprechender Testverfahren ist die Herstellung von Filmen und die Bewertung der Antwortoptionen die schwierigste Aufgabe. Beides wird durch die Komplexität der dargestellten Szenen beeinflusst. Diese Komplexität ist schwierig darstellbar und Antwortalternativen sind schwierig zu bewerten. Realitätsnähe bedeutet auf der anderen Seite eben auch, dass das Antwortverhalten von Unterschieden in der Situationswahrnehmung oder von erfahrungsbasierten Skripten der Testpersonen überlagert sein kann. Antwortbedingte Szenenfortführungen (verzweigende Videotests) verkomplizieren die Erstellung des Testmaterials und schaffen zudem unterschiedliche Bedingungen für verschiedene Testpersonen.

3.5 Diagnostikbedarf durch technische Entwicklungen

Technik bietet nicht nur Chancen für die Diagnostik, sondern macht auch Diagnostik notwendig. Eine wichtige Aufgabe ist die Messung von Einstellungen gegenüber technischen Geräten, dem Wissen um die Bedienung oder dem Nutzungsverhalten. Erst die Entwicklung des Computers führte dazu, dass sich die Pädagogische Psychologie mit der Messung der „computer literacy“ und der computerbezogenen Selbstwirksamkeit und die Arbeitspsychologie um Aspekte der Softwareergonomie und der „Usability“ kümmern. Liegen entsprechende Messverfahren für den Umgang mit technischen Mitteln vor, dann können diese Methoden auch als beeinflussende Faktoren bei der Anwendung technikhorientierter Diagnostik identifiziert und entsprechend von den relevanten Daten separiert werden.

Als weitere Beispiele für die Messung von Verhalten und Einstellung gegenüber technischen Instrumenten, die beim Einsatz der Technik als Diagnostikum wichtig sind, können genannt werden:

1. Die *Erfassung von Software-Qualitäten* spielt eine zunehmend wichtigere Rolle im Designprozess. Hatte sich Donald Norman (1988) seinerzeit mit den Kriterien rationaler Designphilosophie auseinandergesetzt, wird in seinem neuen Buch (Norman, 2004) die emotionale Seite des Designs beleuchtet. Mit dem „User Experience Questionnaire“ (Laugwitz, Schrepp & Held, 2006) liegt z. B. ein Fragebogen vor, der mit 26 Items eine schnelle Messung verschiedener Kriterien der Softwarequalität erlaubt, und zwar sowohl „harte“ Usability- als auch „weiche“ User-Experience-Kriterien. Die Relevanz der Kriterien für die Beurteilung wurde durch eine empirische Selektion sichergestellt. Die Items lassen sich den sechs Faktoren Attraktivität, Durchschaubarkeit, Effizienz, Vorhersagbarkeit, Stimulation und Originalität zuordnen lassen. Erste Validierungsuntersuchungen deuten auf eine zufriedenstellende Konstruktvalidität hin.
2. Für die *Diagnostik mittels virtueller Welten* (vgl. Abschnitt 3.3) ermöglicht die „Virtual Environment Performance Battery“ (VEPAB) die Messung des Einflusses grundlegender kognitiver und psychomotorischer Variablen auf die Leistungen der Testperson (Lampton et al., 1994). Zur Erfassung der häufigen Nebenwirkungen wie Übelkeit oder Orientierungsstörungen bei der Nutzung von virtuellen Welten liegen inzwischen verschiedene Messverfahren vor, z. B. der „Simulator Sickness Questionnaire“ (Bouchard, Robillard & Renaud, 2007).
3. Für die *Computerdiagnostik* ist es bedeutsam, den Einfluss von computerbezogenen Einstellungen und praktischem Computerwissen zu kennen. Die Messung entsprechender Konstrukte ist mit dem „Inventar zur Computerbildung“ (INCOBI) von Naumann, Richter und Groeben (2001) möglich. Die Teilskalen SUCA (Sicherheit im Umgang mit Computern und Computeranwendungen) und VECA (Vertrautheit mit verschiedenen Computeranwendungen) ermöglichen die Erfassung von Einstellung, Erfahrungen und Verhaltensweisen gegenüber dem Computer. Weitere, häufig verwendete Skalen sind die „Computer Anxiety Scale“ (CAS) und das „Computer Thoughts Survey“ (Gordon, Killely, Shelvin, McIlroy & Tierney, 2003).
4. „*Technophobie*“ ist ein eigenständiges Konstrukt, das im Hinblick auf die Einstellung zu technischen Geräten eingeführt wurde. Es ist im Hinblick auf die Compliance bei technischer Diagnostik wichtig. Insbesondere im Fall von Computerangst ist die Störung wegen der alltäglichen Präsenz von Computern so schwerwiegend, dass spezielle Therapieprogramme dafür entwickelt wurden (Brosnan & Thorpe, 2006). Technophobie scheint zum einen durch die Erfahrung, die Personen mit den Geräten gemacht haben, und durch Persönlichkeitsmaße beeinflusst zu sein. Technophobie korreliert positiv mit Neurotizismus und negativ mit dem Faktor „Openness“. Alter und Geschlecht scheinen keinen Einfluss auf dieses Phänomen zu haben (Anthony, Clarke & Anderson, 2000).

5. In den Anfängen befindet sich die Messung der *Leistungsfähigkeit in der Bedienung* von PDA-Geräten oder Handys (Ziefle & Bay, 2005), z. B. im Hinblick die Schnelligkeit bei der Eingabe von „Short-Message“-Nachrichten (Faulkner & Culwin, 2005).
6. Ebenso ist für den klinischen Bereich auch die *Messung pathologischer Nutzung* der Technologie bedeutsam. Beispielhaft sei hier die „Mobile Phone Problem Use Scale“ von Bianchi und Phillips (2005) erwähnt.

Scheuer, Mühlenbrock und Melis (2007) berichten über automatisierte Auswertungshilfen bei der Verarbeitung von Log-Daten, die aus der Interaktion von webbasierten Lernumgebungen resultieren. Gerade in pädagogischen Kontexten kommen hier interessante Anwendungen in Frage, wie z. B. anhand von „student tracking“ in der Lernumgebung „Moodle“ demonstriert wird (Mazza & Botturi, 2007).

4 Chancen und Risiken zukünftiger Entwicklungen

Die technische Entwicklung nimmt erheblichen Einfluss auf die Gestaltung psychologischer Diagnostik. Insbesondere die modernen Informationstechnologien eröffnen ungeahnte Chancen zur Datengewinnung. Bei aller Begeisterung für technische Weiterentwicklungen im Bereich der Diagnostik bleiben aber auch einige Problemfelder zu bedenken, für die Lösungen gesucht werden müssen. Doch zunächst zu den Chancen.

4.1 Chancen

In der psychologischen Diagnostik ergeben sich durch technische Weiterentwicklungen:

1. Chancen durch *größere Realitätsnähe* der diagnostischen Situation: Die Grundannahme hierbei ist, dass eine stärkere Verhaltensnähe die Validität der Messungen erhöht.
2. Chancen durch *höhere Motivation* der Teilnehmer: Gerade die jüngere Generation, die bereits mit einer stärker von Technik durchdrungenen Welt aufgewachsen ist, wird womöglich mit größerem Spaß und höherer Motivation an diagnostische Situationen herangehen, die spielerisch angelehnt sind an Umgebungen, die z. B. aus Computerspielen bekannt sind.
3. Chancen durch *höhere Akzeptanz* diagnostischer Verfahren durch Testanden und Auftraggeber (soziale Validität): Durch die höhere Realitätsnähe

wird die Rechtfertigung der diagnostischen Situation in nachvollziehbarer Weise geliefert.

4. *Bessere Datenqualität und umfassendere Datenquellen*: Durch die elektronischen Erhebungsmöglichkeiten besteht die Chance zu fehlerfreien Datensätzen, die zu vielen Zeitpunkten erhoben werden können und damit die Möglichkeit zu Zeitreihenanalysen bieten (Singer & Willett, 2003; Stroekunold & Werner, 2007; Werner, 2005).
5. Gewinne durch *adaptive Verfahren*: Je intelligenter Diagnostik wird, umso weniger müssen überflüssige Informationen eingeholt werden. Adaptive Verfahren können sich im Leistungsbereich an der aussagefähigen mittleren Schwierigkeit bewegen, im Persönlichkeitsbereich auf kritische Items hin Vertiefungen in bestimmte Richtungen vornehmen.
6. Chancen durch *neue Konstrukte*: Genannt seien hier beispielhaft allgemeines Suchverhalten, Entscheidungsverhalten, Handlungsplanung, Planungshorizont, Informationssuche im Internet, Medienkompetenz: All diese Merkmale sollten durch entsprechende Messverfahren abgebildet werden können. Zu einigen der genannten Konstrukte sind bestehende diagnostische Instrumente genannt worden, an anderer Stelle sind Verfahren weiterzuentwickeln.

4.2 Risiken

Probleme des Aufzeichnens von Daten. Langheinrich, Coroama, Bohn und Matern (2005) systematisieren die Durchdringung von einstmalig „messfreien“ Räumen anhand folgender Dimensionen: natürliche Grenzen werden überwunden, z. B. Mauern, Türen, Kleider; soziale Grenzen werden überwunden, z. B. findet eine Erhebung auch von anderen Personen über Personen statt, z. B. bei der Analyse von Arztbesuchen etc.; zeitliche und örtliche Grenzen fallen weg, d. h. Verhalten aus früheren Zeiten beeinflusst die Bewertung heutiger Daten und das Verhalten an einem Ort beeinflusst die Datenerhebung an einem anderen Ort. Durch umfassende Erhebungs- und Speichermöglichkeiten („data logging“) fallen die Grenzen der beschränkten Nachhaltigkeit („ephemeral or transitory effects“) weg. Während bei klassischen Testverfahren die Testanden entscheiden können, wie und wann sie sich darstellen, verschwinden beim Agieren in virtuellen Welten oder bei automatisierten Messung weitgehend solche Möglichkeiten. Was für den Diagnostiker als Erhöhung der Validität gelobt wird, ist für die Testpersonen Ausdruck mangelnder Einflussnahmen.

Gefahren durch Zugangselektion. Eine weitere Gefahr technischer Neuerungen stellt die eingeschränkte Zugänglichkeit („amenability“) für einzelne Personen-

gruppen (Testleiter und Testpersonen) dar. So ist zu hinterfragen, ob die Möglichkeit einer Online-Testung für die Personalauswahl als Teil des Recruitments nicht Personen ohne Internetzugang benachteiligt. Es ist wichtig, dass alle gesellschaftlichen Gruppen Zugang zu den neusten diagnostischen Verfahren haben und auch bei der Testung nicht benachteiligt werden.

Mangelnde theoretische Fundierung. Eine Herausforderung für Psychologen wird die adäquate Nutzung auf der Grundlage einer theoretischen Fundierung sein. Hierbei ist der Grundsatz „Erst der Inhalt, dann die Methode“ eine wichtige Leitlinie. Im Hinblick auf die derzeit noch hohen Entwicklungs- und Anschaffungskosten muss geprüft werden, ob der Einsatz der Technik auch eine Verbesserung im Sinne testtheoretischer Erwägungen mit sich bringt. Andererseits ist davor zu warnen, dass die nicht-technische Diagnostik als veraltet und damit als nicht förderungswürdig gilt, nur weil der Einsatz technischer Mittel en vogue ist.

Ökologische Folgekosten. Da die meisten Geräte elektrisch betrieben werden und auch der Herstellungsaufwand groß ist, sollten neben den Kosten auch die ökologischen Aspekte bei der Anwendung technischer Mittel beachtet werden. Ein Blick in manches Gerätelager von Hochschulen zeugt von der Begeisterung für technische Geräte, aber auch von der geringen Nutzungsdauer der teuren Anschaffungen. Zurück bleibt neben der Ernüchterung, dass die Anschaffungskosten und die Mühen bei der Einarbeitung in die Technologie im Vergleich zum Nutzen zu hoch waren, auch ein großer Entsorgungsfall.

5 Schlussfolgerungen

Psychologie und Technik – in weiten Teilen der Bevölkerung wird man hier wohl kaum eine Verbindung vermuten. Tatsächlich ist diese Verbindung doch sehr stark gewachsen: Technische Entwicklungen haben auch in der Psychologie erhebliche Spuren hinterlassen. Die Psychologie kann vor diesen Veränderungen nicht die Augen verschließen und muss hier aktiv gestaltend einwirken. Dies in zweierlei Hinsicht: (1) Für viele technische Geräte mangelt es heute noch an diagnostischen Verfahren, um die Einstellung, die Bedienung und mögliche Nebenwirkungen zu erfassen. Hier muss die Psychologie vermehrt geeignete Testinstrumente entwickeln. (2) Für den Einsatz technischer Mittel in der Diagnostik ist es wichtig, dass deren Inhalte die Wahl der technischen Geräte mitbestimmen. Es darf nicht darum gehen, ein trendiges Diagnostikum zu schaffen, sondern Technik ist zu nutzen, um den diagnostischen Prozess zu erleichtern, um messtheoretische Fortschritte zu machen oder aber um neue psychologische Konstrukte zu erschließen, die mit bisherigen Verfahren nicht (geignert) erfasst werden konnten.

Literatur

- Ackerman, T. A., Evans, J. A., Park, K., Tamassia, C. & Turner, R. (1999). Computer assessment using visual stimuli: A test of dermatological skin disorders. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 137–150). Mahwah, NJ: Erlbaum.
- Albert, D. & Gundlach, H. (Hrsg.). (1997). *Apparative Psychologie: Geschichtliche Entwicklung und gegenwärtige Bedeutung*. Lengerich: Pabst.
- Anthony, L. M., Clarke, M. C. & Anderson, S. J. (2000). Technophobia and personality subtypes in a sample of South African university students. *Computers in Human Behavior*, 16, 31–44.
- Auerbach, C. F. & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York: New York University Press.
- Baheux, K., Yoshizawa, M., Seki, K. & Handa, Y. (2006). Virtual reality pencil and paper tests for neglect: A protocol. *CyberPsychology & Behavior*, 9, 192–195.
- Baumeister, R. F., Vohs, K. D. & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2, 396–403.
- Beale, R. (2007). Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *International Journal of Human-Computer Studies*, 65, 421–433.
- Bejar, I. I. (1988). A sentence-based automated approach to the assessment of writing: A feasibility study. *Machine-Mediated Learning*, 2, 321–332.
- Bejar, I. I., Lawless, R. R., Morley, M. W., Wagner, M. E., Bennett, R. E. & Revuelta, J. (2002). *A feasibility study of on-the-fly item generation in adaptive testing*. Graduate Record Examination Proposal 98–12. Chicago, IL: Educational Testing Service.
- Bennett, R. E. & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. *Educational Measurement: Issues and Practice*, 17, 9–17.
- Betsch, T., Funke, J. & Plessner, H. (2011). *Denken – Urteilen, Entscheiden, Problemlösen*. Heidelberg: Springer.
- Bianchi, A. & Phillips, J. G. (2005). Psychological predictors of problem mobile phone use. *CyberPsychology & Behavior*, 8, 39–51.
- Blech, C. & Funke, J. (2006). Zur Reaktivität von Kausaldiagramm-Analysen beim komplexen Problemlösen. *Zeitschrift für Psychologie*, 117, 185–195.
- Blech, C. & Funke, J. (2010). You cannot have your cake and eat it, too: How induced goal conflicts affect complex problem solving. *Open Psychology Journal*, 3, 42–53.
- Bopp, M. (2006). Didactic analysis of digital games and game-based learning. In M. Pivec (Ed.), *Affective and emotional aspects of human-computer interaction. Game-based and innovative learning approaches* (pp. 8–37). Amsterdam: IOS Press.
- Bopp, M. (2007). Storytelling as a motivational tool in digital learning games. In T. Hug & M. Lindner (Eds.), *Didactics of microlearning* (pp. 261–279). Münster: Waxmann.
- Bouchard, S., Robillard, G. & Renaud, P. (2007). Revising the factor structure of the Simulator Sickness Questionnaire. *Annual Review of CyberTherapy and Telemedicine*, 5, 117–122.

- Brehmer, B. & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9, 171–184.
- Broeren, J., Samuelsson, H., Stibrant-Sunnerhagen, K., Blomstrand, C. & Rydmark, M. (2007). Neglect assessment as an application of virtual reality. *Acta Neurologica Scandinavica*, 116, 157–163.
- Brosnan, M.J. & Thorpe, S.J. (2006). An evaluation of two clinically-derived treatments for technophobia. *Computers in Human Behavior*, 22 (6), 1080–1095.
- Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27–63). Hillsdale, NJ: Erlbaum.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143–159.
- Cleeland, C.S. (2000). Cancer-related symptoms. *Seminars in Radiation Oncology*, 10, 175–190.
- Côté, S. & Bouchard, S. (2005). Documenting the efficacy of virtual reality exposure with psychophysiological and information processing measures. *Applied Psychophysiology and Biofeedback*, 30, 217–232.
- Dieckmann, P. (2005). „Ein bisschen wirkliche Echtheit simulieren“: Über Simulatorsettings in der Anästhesiologie (Dissertation). Oldenburg: Universität, Fachbereich 5 Philosophie/Psychologie/Sportwissenschaft.
- Dörner, D., Kreuzig, H.W., Reither, F. & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Drasgow, F., Olson-Buchanan, J.B. & Moberg, P.J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow & J. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 199–196). Mahwah, NJ: Erlbaum.
- Faulkner, C. & Culwin, F. (2005). When fingers do the talking: A study of text messaging. *Interacting with Computers*, 17, 167–185.
- Feuchter, A. & Funke, J. (2004). Positive Effekte sozialen Faulenzens beim Lösen komplexer Probleme. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 56, 304–325.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.
- Franke, J. & (Hornke, L. F. 2011). Eine Konzeption zum systematischen Aufbau von Eignungsuntersuchungen – nach 40 Jahren noch aktuell. In L.F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder der psychologischen Diagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 1, S. 243–279) Göttingen: Hogrefe.
- Fricke, R. (1990). Videotests: „True-to-life“-Testsituationen durch interaktives Video. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 463–466). Göttingen: Hogrefe.
- Funke, J. (1993a). Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 119–129.
- Funke, J. (1993b). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube & K.-F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 313–330). Amsterdam: Elsevier.

- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11, 133–142.
- Funke, J. & Grube-Unglaub, S. (1993). Skriptgeleitete Diagnostik von Planungskompetenz im neuropsychologischen Kontext: Erste Hinweise auf die Brauchbarkeit des „Skript-Monitoring-Tests“ (SMT). *Zeitschrift für Neuropsychologie*, 4, 75–91.
- Funke, J. & Spering, M. (2006). Methoden der Denk- und Problemlöseforschung. In J. Funke (Hrsg.), *Denken und Problemlösen* (Enzyklopädie der Psychologie, Serie Kognition, Bd. 8, S. 647–744). Göttingen: Hogrefe.
- Giese, F. (1927). *Methoden der Wirtschaftspsychologie. Handbuch der biologischen Arbeitsmethoden*. Berlin: Urban & Schwarzenberg.
- Gordon, M., Killey, M., Shelvin, M., McIlroy, D. & Tierney, K. (2003). The factor structure of the Computer Anxiety Rating Scale and the Computer Thoughts Survey. *Computers in Human Behavior*, 19, 291–298.
- Gould, N.F., Holmes, M.K., Fantie, B.D., Luckenbaugh, D.A., Pine, D.S., Gould, T.D. et al. (2007). Performance on a virtual reality spatial memory navigation task in depressed patients. *American Journal of Psychiatry*, 164, 516–519.
- Gray, W.D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognition research. *Cognitive Science Quarterly*, 2, 205–227.
- Greiff, S. (2011). *Individualdiagnostik der komplexen Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S. & Funke, J. (2011). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. *Zeitschrift für Pädagogische Psychologie*, 56 (Beiheft), 216–227.
- Grube-Unglaub, S. & Funke, J. (1995). Der „Skript-Monitoring-Test“ als Diagnostikum für den neuropsychologischen Einsatz. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (S. 141–152). Bonn: Deutscher Psychologen Verlag.
- Hageböck, J. (1990). *Programmsystem für die psychometrische Einzelfalldiagnostik (PSYMEDIA)*. Göttingen: Hogrefe.
- Hageböck, J. (1994). *Computerunterstütztes Diagnosesystem für die Einzelfallhilfe in der schulp-psychologischen Beratungsstelle (DIASYS 1)*. Göttingen: Hogrefe.
- Harris, W.G. (1987). Computer-based test interpretations: Some development and application issues. *Applied Psychology*, 36 (3–4), 237–247.
- Hecht, D. & Reiner, M. (2007). Field dependency and the sense of object-presence in haptic virtual environments. *CyberPsychology & Behavior*, 10, 243–251.
- Hofmann, J. (2002). *Raumwahrnehmung in virtuellen Umgebungen: Der Einfluss des Präsenzzempfindens in Virtual Reality-Anwendungen für den industriellen Einsatz*. Wiesbaden: Deutscher Universitätsverlag.
- Hornke, L.F., Dewald, D. & Hausen, C. (1990). *Videogestütztes Diagnoseverfahren der sozialen Interaktion im Straßenverkehr*. Bergisch Gladbach: Bundesanstalt für Straßenwesen.
- ITEMGEN1. (2002). *Item generator for non-verbal intelligence test items*. Lawrence, KS: Psychological Data Corporation.

- Jäger, R. S. & Krieger, W. (1994). Zukunftsperspektiven der computergestützten Diagnostik dargestellt am Beispiel der treatmentorientierten Diagnostik. *Diagnostica*, 40, 217–243.
- Kahng, S. W. & Iwata, B. A. (2000). Computer systems for collecting real-time observational data. In T. Thompson, D. Felce & F. J. Symons (Eds.), *Behavioral observation: Technology and applications in development disabilities* (pp. 35–45). Baltimore, MD: Brookes Publishing.
- Kaiser, A. P., Tapp, J., Solomon, N. A., Delaney, E. M., Ezell, S. S., Hester, P. P. et al. (2000). Observing complex adult-child interactions: Computer-supported coding, analysis and graphing. In T. Thompson, D. Felce & F. J. Symons (Eds.), *Behavioral observation: Technology and applications in development disabilities* (pp. 177–192). Baltimore, MD: Brookes Publishing.
- Kauchak, M. (2007). Forecast for gamers. *Military Training Technology Online Archives*, 12 (6). Retrieved July 31, 2008, from <http://www.military-training-technology.com/article.cfm?DocID=2266>
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlösenszenarien? Zur Kriteriumsvalidität von Problemlösenszenarien und Intelligenztests*. Göttingen: Hogrefe.
- Kleinmuntz, B. & Barton, A. L. (1962). Computer program for the Meehl-Dahlstrom MMPI profile rules. *Educational and Psychological Measurement*, 22, 193–199.
- Kluwe, R. H. (1997). Simulation in der empirisch-psychologischen Forschung. In D. Albert & H. Gundlach (Hrsg.), *Apparative Psychologie: Geschichtliche Entwicklung und gegenwärtige Bedeutung* (S. 203–224). Lengerich: Pabst.
- Kuntze, J., van der Molen, H. T. & Born, M. P. (2007). Progress in mastery of counseling communication skills: Development and evaluation of a new instrument for the assessment of counseling communication skills. *European Psychologist*, 12, 301–313.
- Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J. & Blau, B. (1994). The virtual environment performance assessment battery (VEPAB): Development and evaluation. *Presence: Teleoperators and Virtual Environments*, 3, 145–157.
- Langheinrich, M., Coroama, V., Bohn, J. & Mattern, F. (2005). Living in a smart environment – Implications for the coming ubiquitous information society. *Telecommunications Review*, 15 (1), 132–143.
- Laugwitz, B., Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In A. M. Heinecke & H. Paul (Hrsg.), *Mensch & Computer 2006 – Mensch und Computer im Strukturwandel* (S. 125–134). München: Oldenbourg.
- Lievens, F., Buyse, T. & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442–452.
- Lumsdaine, A. A. (1960). Teaching machines: An introductory overview. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book* (pp. 5–22). Washington, DC: National Education Association of the United States.
- Mangold International GmbH (2006). *INTERACT. Software Paket*. Arnsdorf: Mangold.
- Martin, R. (2003). Le testing adaptatif par ordinateur dans la mesure en education: Potentialities et limites. *Psychologie et Psychometrie*, 2, 83–110.
- Mazza, R. & Botturi, L. (2007). Monitoring an online course with the GISMO tool: A case study. *Journal of Interactive Learning Research*, 18 (2), 251–265.
- McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., Milne, A. B. et al. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence*, 10, 375–383.
- Moede, W. (1928). 10 Jahre Institut für industrielle Psychotechnik T. H. Berlin. *Werkstattstechnik*, 20, 587–592.
- Mühlberger, A., Bühlhoff, H. H., Wiedemann, G. & Pauli, P. (2007). Virtual reality for the psychophysiological assessment of phobic fear: Responses during virtual tunnel driving. *Psychological Assessment*, 19, 340–346.
- Müller, A. (1995). *Das Situative Video Verfahren. Die Entwicklung eines neuartigen Verfahrens der videogestützten Diagnostik von polizeilichen Kompetenzen*. Unveröffentlichte Diplomarbeit im Fach Psychologie, Philosophische Fakultät der Universität Bonn, Bonn.
- Naumann, J., Richter, T. & Groeben, N. (2001). Validierung des Inventars zur Computerbildung (INCOBI) anhand eines Vergleichs von Anwendungsexperten und Anwendungsneovizen. *Pädagogische Psychologie*, 15, 219–232.
- Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.
- Norman, D. A. (2004). *Emotional design. Why we love (or hate) everyday things*. New York: Basic Books.
- Nuthmann, A., Engbert, R. & Kliegl, R. (2006). Messung von Blickbewegungen. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 705–711). Göttingen: Hogrefe.
- Nyberg, L., Lundin-Olsson, L., Sondell, B., Backman, A., Holmlund, K., Eriksson, S. et al. (2006). Using a virtual reality system to study balance and walking in a virtual outdoor environment: A pilot study. *CyberPsychology & Behavior*, 9, 388–395.
- Olson-Buchanan, J. B. & Drasgow, F. (1999). Beyond bells and whistles. In F. Drasgow & J. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 1–6). Mahwah, NJ: Erlbaum.
- Omodei, M. M. & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments & Computers*, 27, 303–316.
- Omodei, M. M., Wearing, A. J. & McLennan, J. (2000). Relative efficacy of an open versus a restricted communication structure for command and control decision making: An experimental study. In C. McCann & R. Pigeau (Eds.), *The human in command: Exploring the modern military experience* (pp. 369–386). New York: Plenum.
- Parsons, T. D., Rizzo, A. A., Bamattre, J. & Brennan, J. (2007). Virtual Reality Cognitive Performance Assessment Test. *Annual Review of CyberTherapy and Telemedicine*, 5, 143–149.
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Prensky, M. (2000). *Digital game-based learning*. New York: McGraw-Hill.
- Pressey, S. L. (1926). A simple apparatus which gives tests – and teaches. *School and Society*, 23, 373–376.

- Reuschenbach, B. (2008). *Einfluss von Expertise auf Problemlösen und Planen im komplexen Handlungsfeld Pflege*. Berlin: Logos.
- Riva, G. (1998). Virtual reality in psychological assessment: The Body Image Virtual Reality Scale. *CyberPsychology & Behavior*, 1 (1), 37–44.
- Roelofs, J., Peters, M. L., Patijn, J., Schouten, E. G. W. & Vlaeyen, J. W. S. (2004). Electronic diary assessment of pain-related fear, attention to pain, and pain intensity in chronic low back pain patients. *Pain*, 112, 335–342.
- Rosengren, D. B., Baer, J. S., Hartzler, B., Dunn, C. W. & Wells, E. A. (2005). The video assessment of simulated encounters (VASE): Development and validation of a group-administered method for evaluating clinician skills in motivational interviewing. *Drug and Alcohol Dependence*, 79, 321–330.
- Schandry, R. & Leopold, C. (1996). Ambulatory assessment of self-monitored subjective and objective symptoms of diabetic patients. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 393–402). Seattle, WA: Hogrefe & Huber.
- Schaper, N., Schmitz, A. P., Graf, B. & Grube, C. (2004). Gestaltung und Evaluation von simulatorgestützten Trainings in der Anästhesie. In T. Manser (Hrsg.), *Komplexes Handeln in der Anästhesie* (S. 229–260). Lengerich: Pabst.
- Scheuer, O., Mühlenbrock, M. & Melis, E. (2007). Results from action analysis in an interactive learning environment. *Journal of Interactive Learning Research*, 18 (2), 185–205.
- Schiff, W., Arnone, W. & Cross, S. (1994). Driving assessment with computer-video scenarios: More is sometimes better. *Behavior Research Methods, Instruments & Computers*, 26, 192–194.
- Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2, 419–436.
- Schneider, C. Q. & Wagemann, C. (2007). *Qualitative comparative analysis (QCA) und Fuzzy Sets. Ein Lehrbuch für Anwender und jene, die es werden wollen*. Opladen: Budrich.
- Schoech, D. (2001). Using video clips as test questions: The development and use of a multimedia exam. *Journal of Technology in Human Services*, 18 (3–4), 117–131.
- Schulte, R. W. (1921). *Die Berufseignung des Damenfriseurs. Methoden und Ergebnisse eines psychotechnischen Prüfungssystems auf der Grundlage einer Funktionsanalyse* (Schriften zur Psychologie der Berufseignung und des Wirtschaftslebens, Bd. 17). Leipzig: Johann Ambrosius Barth.
- Sharples, S., Cobb, S., Moody, A. & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, 29 (2), 58–69.
- Shermis, M. D. & Burstein, J. (Eds.). (2003). *Automated essay scoring: A cross disciplinary perspective*. Mahwah, NJ: Erlbaum.
- Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis. Modeling change and event occurrence*. New York: Oxford University Press.
- Singley, M. K. & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 361–384). Mahwah, NJ: Erlbaum.

- Stanney, K. M., Mourant, R. R. & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, 7, 327–351.
- Stapf, K.-H. (1997). Apparative Diagnostik in der Psychotechnik. In D. Albert & H. Gundlach (Hrsg.), *Apparative Psychologie: Geschichtliche Entwicklung und gegenwärtige Bedeutung* (S. 179–202). Lengerich: Pabst.
- Stone, R. J. (2005). Serious gaming. *Defence Management Journal* (Issue 31), 142–144.
- Stroe-Kunold, E. & Werner, J. (2007). Sind psychologische Prozesse kointegriert? Standortbestimmung und Perspektiven der Kointegrationsmethodologie in der psychologischen Forschung. *Psychologische Rundschau*, 58, 225–237.
- Tart, C. T. (1990). Multiple personality, altered states and virtual reality: The world simulation process approach. *Dissociation: Progress in the Dissociative Disorders*, 3, 222–233.
- Toobert, D. J., Glasgow, R. E., Desalvo, M. A. & Strycker, L. A. (1998). Computerized touchscreen video vs. paper-and-pencil assessment of dietary behavior. *CyberPsychology & Behavior*, 1, 257–261.
- Ulrich, R. & Schröter, H. (2006). Mentale Chronometrie. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 685–690). Göttingen: Hogrefe.
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst.
- Werner, J. (Hrsg.). (2005). *Zeitreihenanalysen mit Beispielen aus der Psychologie*. Berlin: Logos.
- Wiederhold, B. K. & Rizzo, A. (2005). Virtual reality and applied psychophysiology. *Applied Psychophysiology and Biofeedback*, 30, 183–185.
- Wild, B., Fontaine, A. de la & Schafsteller, C. (2001). Fishing for Talents: Internet-Recruiting auf neuen Wegen. Mit dem Online-Spiel „Challenge Unlimited“ erprobte die Siemens AG neue Wege des Recruitings im Internet. *Personalführung*, 34 (1), 66–70.
- Wirth, J. (2008). Computer-based tests: Alternatives for test and item design. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 235–252). Göttingen: Hogrefe.
- Wittchen, H.-U., Weigel, A., Pfister, H. & Perkonig, A. (1997). *DIA-X: Expertensystem zur Diagnostik Psychischer Störungen*. Frankfurt: Swets.
- Ziefle, M. & Bay, S. (2005). How older adults meet complexity: Aging effects on the usability of different mobile phones. *Behaviour & Information Technology*, 24, 375–389.

ENZYKLOPÄDIE DER PSYCHOLOGIE

In Verbindung mit der
Deutschen Gesellschaft für Psychologie

herausgegeben von

Prof. Dr. Niels Birbaumer, Tübingen
Prof. Dr. Dieter Frey, München
Prof. Dr. Julius Kuhl, Osnabrück
Prof. Dr. Wolfgang Schneider, Würzburg
Prof. Dr. Ralf Schwarzer, Berlin

Themenbereich B

Methodologie und Methoden

Serie II

Psychologische Diagnostik

Band 3

Leistungs-, Intelligenz-
und Verhaltensdiagnostik



Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle

Leistungs-, Intelligenz- und Verhaltensdiagnostik

herausgegeben von

Prof. em. Dr. Lutz F. Hornke, Aachen
Prof. em. Dr. Manfred Amelang, Heidelberg
Prof. Dr. Martin Kersting, Münster



Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle