

# Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme

## Projekt Dynamisches Problemlösen<sup>1</sup>

### 1. Einführung

Bis vor wenigen Jahren basierten psychologische Testverfahren zur Kompetenz- und Fähigkeitsmessung fast ausschließlich auf konventionellen Papier- und Bleistift-Methoden. Mit dem Aufkommen von Computern ergaben sich neue und effiziente Möglichkeiten zur Erfassung von Fähigkeiten, die heute in modernen diagnostischen Verfahren wie *Computer Adaptive Testing* (CAT) münden. Neben einer höheren Effizienz haben sich dank der technischen Entwicklung aber auch neuartige Konstrukte entwickelt, die über klassische Formate nicht erfassbar waren. Eines dieser Konstrukte ist *komplexes Problemlösen*, das per se dynamisch und interaktiv ist (vgl. Funke 2003), sodass eine Testung nur computerbasiert möglich ist.

Komplexes Problemlösen (KPL) hat in den vergangenen Jahrzehnten im Hauptinteresse experimentalpsychologischer Forschung gestanden. Demgegenüber vernachlässigt wurde die Individualdiagnostik von KPL, die lediglich vereinzelt Berücksichtigung fand (vgl. z.B. Beckmann/Guthke 1995; Wägener 2001). Zugleich ist ein aufkeimendes Interesse an cross-curricularen Kompetenzen und damit auch an KPL in internationalen Bildungsstudien wie PISA zu beobachten (vgl. Klieme/Leutner/Wirth 2005). Als erste Konsequenz wurden im Rahmen einer Felderprobung sowie nachfolgend in einer nationalen Ergänzungsstudie Deutschlands zu PISA 2000 Hinweise auf die Messbarkeit und das Potential dynamischer Problemlösefähigkeit erbracht (vgl. Klieme u.a. 2001): Über einen semantisch in den Kontext der Raumfahrt eingebetteten finiten Automaten<sup>2</sup> zur Erfassung der Problemlösefähigkeit konnten die beiden Facetten Wissenserwerb und Wissensanwendung erfasst werden. Explorative Faktorenanalysen, lineare Strukturgleichungsmodelle und multidimensionale Skalierungen zeigten empirisch, dass KPL, analytisches Problemlösen, fachspezifische Kompetenzen und Testintelligenz

1 Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: Fu 173/11-2) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

2 Finite Automaten sind Systeme, die eine begrenzte (= finite) Zahl qualitativ unterschiedlicher Zustände annehmen können. Sie beschreiben auf abstrakter Ebene eine Vielzahl alltäglicher Systeme unterschiedlicher Komplexitätsstufen wie Fahrkartenautomaten oder Computerbetriebssysteme und werden in der Forschung häufig zur Messung der Problemlöseleistung herangezogen.

zwar korrelierte, jedoch voneinander abgrenzbare Konstrukte waren; KPL ließ sich dabei am besten separieren (vgl. Leutner u.a. 2005). Begrifflich betrachten wir KPL und dynamisches Problemlösen als identisch, da KPL im Gegensatz zum analytischen Problemlösen in sich stets dynamisch ist.

Auch aus einer praktischen Perspektive finden sich zahlreiche anwendungsbezogene Implikationen von KPL. Eine Vielzahl an Aktivitäten lässt sich formal als komplexe Problemlöseprozesse beschreiben, bspw. medizinische Notfälle, aber auch die Verwaltung der eigenen Finanzen oder die Bedienung von Fahrkartenautomaten am Bahnhof. Gemein sind diesen Aktivitäten die folgenden Merkmale: (a) eine hohe Anzahl an Variablen ist involviert und die Gesamtinformation über das System muss von Problemlösenden adäquat reduziert werden (Komplexität); (b) verschiedene Variablen beeinflussen eines oder mehrere Resultate (Vernetztheit); (c) das zugrunde liegende System ist nicht statisch (Dynamik); (d) die über das System vorliegende Information ist nicht erschöpfend (Intransparenz) und (e) Ziele können einander widersprechen und möglicherweise nicht simultan erreicht werden (Polyelely).

Diese von Dörner (1986) in der Theorie der operativen Intelligenz benannten fünf Eigenschaften komplexer Probleme korrespondieren mit fünf Anforderungen an eine problemlösende Person: (a) die Reduktion überbordender Information auf einen handhabbaren Umfang (Informationsreduktion); (b) die Bildung adäquater Situationsmodelle zum Verständnis der gegebenen Situation (Modellbildung); (c) die Prognose weiterer Entwicklungen aufgrund der gegebenen Situation und im Lichte getroffener Maßnahmen (Prognose); (d) die Beschaffung fehlender, aber für die Problemlösung notwendige Information (Informationssuche und -generierung) und (e) das Treffen von Wertentscheidungen und Prioritätensetzungen, mit denen Ziele gesetzt und Zielkonflikte gelöst werden können (Bewertung).

Aus theoretischer Sicht sollte ein Messverfahren Indikatoren zu jeder dieser Anforderungen enthalten; ein derartiges Facettendiagnosticum existiert bislang aber nicht. Ergänzend bestätigen zwar aus empirischer Sicht die oben berichteten PISA-Ergebnisse vorläufig die konvergente und divergente Validität von KPL, der verwendete Raumfahrt-Automat war allerdings ein ad hoc konstruiertes Testverfahren mit unklaren psychometrischen Eigenschaften, unklarem Messbereich und fehlender theoretischer Anbindung.

Wir stellen hier ein neuartiges Diagnosticum, MicroDYN, mit überprüfbar psychometrischen Eigenschaften, theoretischem Bezug und der Möglichkeit, einzelne Facetten der Problemlösefähigkeit zu evaluieren, vor. Im empirischen Teil dieses Artikels konzentrieren wir uns auf Modellbildung als einen der fünf Indikatoren Dörners und leiten für diesen ein vorläufiges Kompetenzmodell ab.

### 2. Der MicroDYN-Ansatz

Ungeachtet des gestiegenen Interesses an der individualdiagnostischen Erfassung von KPL besteht nach wie vor ein grundlegender Mangel an gut eingeführten Testverfahren. Zusätzlich existiert kaum Einvernehmen darüber, wie KPL zu operationalisieren und zu

messen ist. Selbst für bestehende Tests gibt es keine hinreichend gesicherten theoretischen Grundlagen, an denen die Messung ansetzen könnte.

Neben diesen insgesamt unbefriedigenden Aspekten ist gegen die in PISA durchgeführte Form der Messung ein Einwand anzuführen, der im Übrigen alle simulierten Mikrowelten, wie sie erstmals von Dörner in den 1970er Jahren entwickelt wurden (vgl. Funke/Frensch 2007), betrifft: Die gesamte Testung besteht aus einem einzelnen Item, das über eine geraume Zeitspanne bearbeitet wird (*one item testing*). Sämtliche Bearbeitungsschritte hängen dabei von vorherigen Entscheidungen und der durchgängig unveränderten Systemstruktur ab. Im Ergebnis basieren die Aussagen über individuelle Problemlösefähigkeit auf der Leistung in diesem einem Item, was grundlegenden psychometrischen Anforderungen widerspricht. Einige Autor/innen versuchen dieses Problem über (a) die Vorgabe eines Systems, das aus mehreren unabhängigen Teilsystemen besteht und die fälschlicherweise separat ausgewertet werden (vgl. z.B. Müller 1993; Wagener 2001), oder über (b) mehrere Fragen zu einem System (wie im finiten Automaten aus PISA) zu lösen. Dies macht die Items aber nur scheinbar unabhängiger und das grundsätzliche Problem besteht weiterhin: (a) Ein umfangreiches System aus unabhängigen Subsystemen bleibt letztlich auch nur ein einzelnes Item und (b) eine Vielzahl an Items (Fragen) zu einem einzelnen Szenario könnte bestenfalls als lokal abhängiges Item-Bündel verstanden werden.

Vor dem dargestellten Hintergrund stellt sich die Frage, wie dynamisches Problemlösen über psychologische Messverfahren überhaupt getestet werden kann. Wir nehmen an, dass interindividuelle Unterschiede im Kontext linearer Strukturgleichungsmodelle erfassbar werden. Dieser Formalismus (s. Abschnitt 3) wurde verschiedentlich als ökologisch valide bezeichnet und ist bereits häufig in experimentellen Arbeiten als Indikator der Problemlöseleistung verwendet worden (vgl. Funke 2001) – dort allerdings als *one item testing*. Wir wählen nun einen modifizierten Ansatz: Anstatt nur ein einzelnes Item darzubieten, bearbeiten Proband/innen unter strikter Zeitbegrenzung eine ganze Serie *minimal komplexer Systeme*. Wir nennen diesen Zugang MicroDYN, um damit die Orientierung an der kleinsten Einheit der Komplexität zu verdeutlichen und zugleich eine Referenz an den zugrunde liegenden DYNAMIS-Ansatz zu machen (vgl. ebd.).

Der MicroDYN-Ansatz vermag einige Versäumnisse bestehender Messverfahren zu lösen oder zumindest deutlich abzumildern. Er bietet dabei die folgenden Vorteile: (a) durch die Anbindung an Dörners Anforderungen an Problemlösende ist ein gewisser theoretischer Bezug gewährleistet (Theoriebezug); (b) durch die Vorgabe von etwa 15 Items mit kurzer Bearbeitungszeit, aber hinreichender Komplexität wird ein entscheidender Nachteil bisheriger Diagnostik überwunden (Itemindependenz); (c) empirisch nachweisbare Facetten der Problemlöseleistung im Sinne der Dörnerschen Anforderungen können zuverlässig gemessen werden (Facettendiagnostik); (d) Items sind einfach zu entwickeln und hinsichtlich ihrer Schwierigkeit frei variierbar (infiniter Itempool); (e) psychometrische Eigenschaften wie Reliabilität können standardmäßig überprüft werden (psychometrisch orientierte Testentwicklung); und (e) Alltagsaktivitäten können über MicroDYN-Items modelliert werden (ökologische Validität).

### 3. Die Items

Ein typisches MicroDYN-Item (illustriert in Abbildung 1) besteht aus exogenen und endogenen Variablen (im Beispiel ein 3x3-System). Die exogenen Variablen können im Gegensatz zu den endogenen Variablen aktiv manipuliert werden. Denkbare Verknüpfungen zwischen den Variablen sind Haupteffekte (HE), multiple Effekte (ME), multiple Abhängigkeiten (MA), Eigendynamiken (ED) und Nebeneffekte (NE). *Haupteffekte* beschreiben kausale Relationen einer exogenen auf eine endogene Variable. Wirkt eine exogene Variable auf mehrere endogene, so ist dies ein *multipler Effekt*. Wird umgekehrt eine endogene Variable von mehreren exogenen beeinflusst, wird dies *multiple Abhängigkeit* genannt. Diese drei Effekte können aktiv manipuliert werden und sind nur dann zu beobachten, wenn die exogenen Variablen (durch direkte Eingabe) Werte ungleich 0 aufweisen. Wirkt eine endogene Variable auf andere endogene, ist dies ein *Nebeneffekt*. Wirkt sie hingegen auf sich selbst (mit einem Gewicht ungleich 1), wird dieser Spezialfall eines Nebeneffektes *Eigendynamik* (als Wachstums- oder Schrumpfungsprozess) genannt. NE und ED können nicht aktiv manipuliert werden und hängen grundsätzlich von den Zuständen der endogenen Variablen ab, sind also weitestgehend unabhängig von den aktiven Manipulationen der Proband/innen. Über die Verwendung adäquater Strategien können sie aber eindeutig identifiziert werden.

Ein Item wird von Proband/innen stets in drei Schritten durchlaufen: (a) Explorationsphase, (b) Modellbildungsphase und (c) Steuerphase.

In der *Explorationsphase* (a) können Proband/innen das System eigenständig und vollständig frei explorieren. Ihre Aufgabe besteht darin, sich mit dem System und seiner

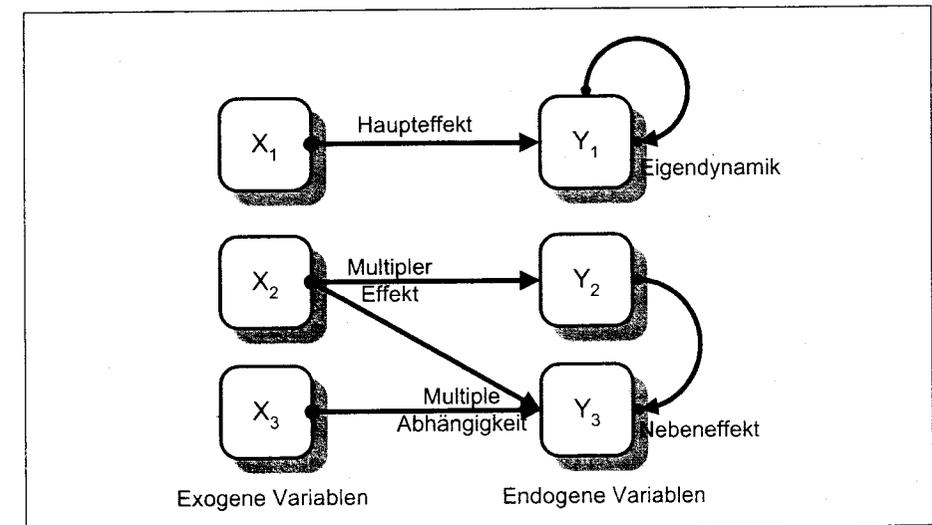


Abb. 1: Struktur eines MicroDYN-Items mit 3 exogenen und 3 endogenen Variablen und den 5 Effektqualitäten

Struktur vertraut zu machen. Die hinterlegten Verknüpfungen sind dabei für die Proband/innen nicht ersichtlich.

Die Systemstruktur soll im Anschluss oder parallel zur ersten Phase in der *Modellbildungsphase* (b) expliziert und aufgezeichnet werden. Phasen (a) und (b) dauern gemeinsam etwa 3 bis 4 Minuten.

In der *Steuerphase* (c) werden Proband/innen mit vorgegebenen Zielwerten in den endogenen Variablen konfrontiert, die sie durch adäquate Manipulation der exogenen Variablen in mehreren Schritten erreichen sollen. Phase (c) dauert 1½ Minuten.

Proband/innen bearbeiten insgesamt 12 bis 15 unabhängige Systeme (Dauer pro Item 4 bis 5 Minuten), was einer angemessenen Itemzahl entspricht und die Testzeit auf ökonomisch vertretbare 60 Minuten begrenzt. Anhand dieser drei Phasen können Aussagen über die Anforderungen der Informationssuche und -generierung (Phase a), der Modellbildung (Phase b) sowie der Prognose (Phase c) getroffen werden. Indikatoren für die beiden Anforderungen der Informationsreduktion und der Bewertung sind über speziell konstruierte Items möglich und derzeit in Vorbereitung.

#### 4. Aktuelle Forschung

Obwohl der beschriebene Itemtyp bereits vielfach als Operationalisierung komplexer Problemlösefähigkeit verwendet wurde, blieb dabei grundsätzlich unklar, ob diese Operationalisierungen reliabel und valide waren. Insbesondere die Vergleichbarkeit zwischen Studien war aufgrund unterschiedlicher Systemstrukturen nicht gewährleistet. Wir glauben, dass ein Blick auf Systemseite und die Aufschlüsselung der Itemschwierigkeit in einzelne systemimmanente Merkmale wichtige Hinweise auf die Anforderungen solcher Systeme, aber auch auf eine mögliche Kompetenzstruktur auf Personenseite geben können. In einer Aufgabenanalyse finden wir sieben Systemdimensionen mit potentiell Einfluss auf die Itemschwierigkeit (Tabelle 1). Diesen Einfluss haben wir in der vorliegenden Untersuchung zu quantifizieren versucht mit dem Ziel, (a) einen Beitrag zur Messbarkeit komplexer Problemlösefähigkeit zu liefern und (b) ein vorläufiges Kompetenzmodell für den Aspekt der Modellbildung abzuleiten. Wir beschränken uns dabei auf diesen Aspekt der Problemlösefähigkeit.

##### 4.1 Design

In einem Messwiederholungsdesign ( $n = 48$ ; 39 weiblich, 9 männlich; Alter  $M = 23.42$ ,  $S = 3.02$ ) bearbeiteten Proband/innen 15 MicroDYN-Items mit einer Gesamtdauer von ca. 60 Minuten. Der Fokus lag dabei auf den ersten drei Dimensionen aus Tabelle 1. Es wurden vornehmlich Haupteffekte untersucht, lediglich eine Interaktion für die als a priori besonders relevant angenommenen Faktoren Effektzahl und Effektivität integrieren wir. Der dreistufige Faktor *Effektivität* (HE, ME, NE; s. Abbildung 1; MA und ED wurden nicht berücksichtigt; Ergebnisse hierzu werden kurz in der Diskussion erwähnt)

Systemmerkmal	mögliche Ausprägungen; Merkmalsklärung
Effektqualität	Haupteffekt, multipler Effekt, multiple Abhängigkeit, Eigendynamik, Nebeneffekt; qualitativ unterschiedliche Verknüpfungen
Effektzahl	frei variierbar; Zahl der Effekte in einem gegebenen System
Variablenzahl	frei variierbar; Zahl der exogenen und endogenen Variablen
relative Effektstärke	frei variierbar; relativer Grad eines Effektes; beinhaltet auch Vorzeichen
Start- und Zielwerte	frei variierbar; Zielwerte nur in der Steuerphase für endogene Variablen
Dispersion	niedrig bis stark; Clusterung bzw. Verteilung der Effekte auf exogene und endogene Variablen
Konfiguration	frei variierbar; Anordnung der Elemente und der Verknüpfungen

Tab. 1: Anhand der Aufgabenanalyse identifizierte Systemmerkmale und ihre Erläuterung

und der zweistufige Faktor *Effektzahl* (2 & 4; s. Abbildung 2) wurden in einem  $2 \times 3$ -Design vollständig gekreuzt. Weiter wurde der dreistufige Faktor *Zahl der Variablen* ( $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$ ; s. Abbildung 3) isoliert variiert. Über Kovarianzanalysen wurde der Einfluss der Itemposition auspartialisiert.

##### 4.2 Hypothesen

In der vorliegenden Untersuchung wurden nur drei der in Tabelle 1 aufgeführten Faktoren getestet: Effektivität, Effektzahl und Variablenzahl. Für *Effektivität* (EQ) nehmen wir die beste Erkennbarkeit bei HE an, gefolgt von ME, während NE schwerlich erkennbar sein sollten. Dies folgt aus der Manipulierbarkeit sowie der Zahl notwendiger Explorationsschritte, um die jeweiligen Effekte zu erkennen. Eine höhere *Effektzahl* (EZ) sollte die Schwierigkeit eines Systems merklich erhöhen. Hinsichtlich der Interaktion  $EQ \times EZ$  vermuten wir keine statistisch bedeutsamen Effekte. Bei steigender *Variablenzahl* sollte ceteris paribus die Leistung der Proband/innen sinken.

##### 4.3 Abhängige Variable

Die Wissensabfrage erfolgte nach der Methode der Kausaldiagramm-Analyse (vgl. Funke 1995), in der Proband/innen die vermutete Systemstruktur auf Papier aufzeichnen und anhand vorgegebener Kategorien die angenommene Stärke eines Pfadgewichtes angeben. Als Indikator wurde die *Güte der Kausaldiagramme* (GdK) in einer

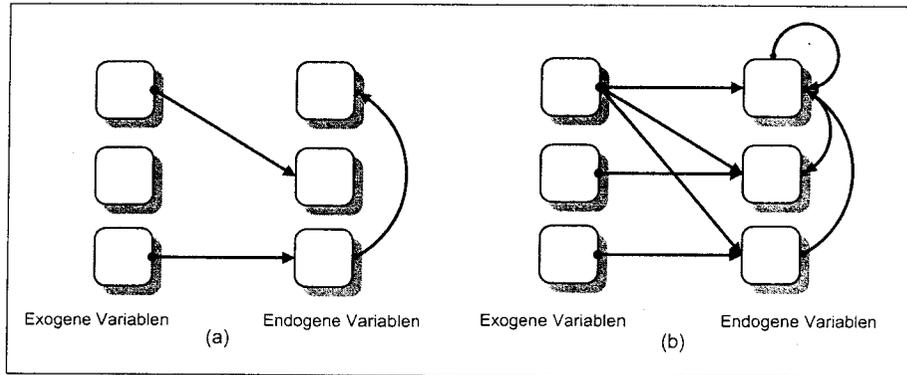


Abb. 2: Zwei Items mit (a) niedriger bzw. (b) hoher Anzahl an Effekten bei konstanter Zahl der Variablen

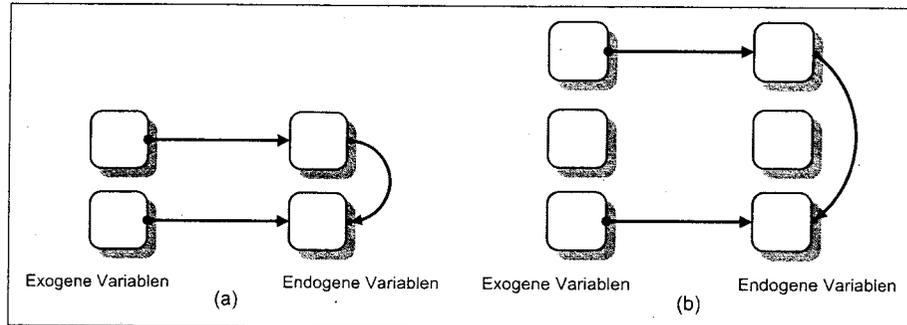


Abb. 3: Zwei Items mit (a) 2 bzw. (b) 3 exogenen und endogenen Variablen bei konstanter Effektzahl

Gewichtung von 0,75 Relations- und 0,25 Stärkewissen verwendet. Richtungswissen (Vorzeichen) wurde nicht ausgewertet. GdK erreicht ein Maximum von 1 (Systemwissen absolut richtig) und ein Minimum von -1 (Systemwissen absolut falsch). GdK hatte sich in einer Simulationsstudie mit insgesamt über 40 möglichen Indikatoren gemessen an einem Kriterium bestehend aus Expertenbeurteilungen fiktiver Kausaldiagramme als überlegener Indikator gezeigt.

#### 4.4 Ergebnisse

Die Voraussetzungen für Kovarianzanalysen waren in hinreichender Weise erfüllt.

Die Effekte der Systemmerkmale auf GdK finden sich in Tabelle 2 und bestätigten generaliter unsere Hypothesen. Die Qualität eines Effektes wirkte sich signifikant ( $p < .001$ ) auf das Kausalwissen aus. HE ( $M = .71$ ;  $SE = .03$ ) und ME ( $M = .70$ ;  $SE = .04$ )

waren leichter als NE ( $M = .60$ ;  $SE = .04$ ;  $p < .001$ ; Kontraste insgesamt nicht dargestellt). HE und ME unterschieden sich dabei nicht ( $p > .10$ ). Dies mag daran liegen, dass Nebeneffekte nur beobachtet, nicht aber aktiv manipuliert werden können. Bei höherer Effektzahl stieg die Schwierigkeit eines Systems deutlich an ( $M_{EZ2} = .75$ ;  $SE_{EZ2} = .04$ ;  $M_{EZ4} = .60$ ;  $SE_{EZ4} = .04$ ;  $p < .001$ ). Dieser Effekt blieb unverändert, wenn die Effektzahl als zufälliger Faktor verstanden wurde (nicht dargestellt). Zwischen der Effektzahl und ihrer Qualität bestand keine Interaktion (Abbildung 4). Eine unterschiedliche Variablenzahl ( $M_{2 \times 2} = .71$ ;  $SE_{2 \times 2} = .04$ ;  $M_{3 \times 3} = .65$ ;  $SE_{3 \times 3} = .05$ ;  $M_{4 \times 4} = .61$ ;  $SE_{4 \times 4} = .06$ ) beeinflusste GdK nicht in der Omnibustestung. Ein geplanter linearer Kontrast allerdings zeigte ein stetiges Ansteigen der Schwierigkeit mit zunehmender Variablenzahl ( $p < .05$ ).

Um erste Aussagen über die psychometrischen Eigenschaften MicroDYNs treffen zu können, verstanden wir die experimentellen Stimuli als eine Testversion mit 15 Items. Cronbachs  $\alpha$  als Reliabilitätsschätzung fiel mit .94 sehr zufriedenstellend aus.

Faktor	Stufen	F	df	p	$\eta^2$	KI $\eta^2$ (.95)
Effektqualität	HE vs ME vs NE	9.95	2/90	<.001**	.18	[.05;.31]
Effektzahl	2 vs 4	37.74	1/45	<.001**	.46	[.23;.60]
IA EQ x EZ		1.61	2/90	>.10	.04	[.00;.12]
Variablenzahl	2 vs 3 vs 4	2.10	2/92	>.10	.04	[.00;.13]

Tab. 2: ANCOVA-Ergebnisse für die getesteten Effekte

Anmerkungen: Kovariate ist Itemposition. KI = Konfidenzintervall; IA = Interaktion; EQ = Effektqualität; EZ = Effektzahl; HE = Haupteffekt; ME = multipler Effekt; NE = Nebeneffekt

#### 5. Implementation

Die Software wurde in enger Kooperation mit dem Deutschen Institut für Internationale Pädagogische Forschung (DIPF, Frankfurt a.M.) und der Firma SOFTCON (München) entwickelt. Die endgültige Version (verfügbar ab Herbst 2009) ist als Autorensystem in die frei zugängliche Plattform TAO (vgl. Latour/Martin 2007) integriert und wird erheblichen Freiraum in Bezug auf graphische Gestaltung, Semantik und Itementwicklung lassen. In Abbildung 5 ist ein Screenshot der derzeitigen Software illustriert.

Neben dem eigentlichen System mit den exogenen Variablen links und den endogenen rechts werden eine Historie sowie ein Zeitbalken dargeboten. Ein Undo- und ein Resetbutton ermöglichten es den Proband/innen in der Explorationsphase, vorherige Schritte zu korrigieren. In der Steuerphase werden zusätzlich extern vorgegebene Zielwerte angezeigt. Der Verlauf wird am unteren Seitenrand dokumentiert.

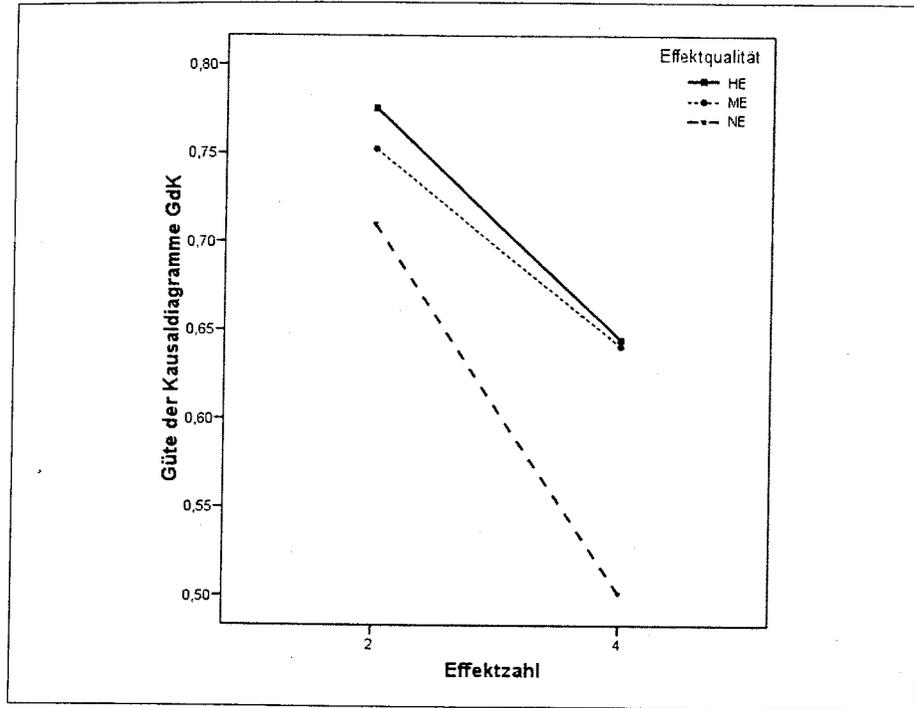


Abb. 4: Mittlere Werte „Güte der Kausaldiagramme“ (GdK) in Abhängigkeit von der Effektzahl (2 & 4) und der Effektqualität (HE = Haupteffekt, ME = multipler Effekt, NE = Nebeneffekt)

## 6. Diskussion

### 6.1 Kompetenzmodell

Auf Grundlage der Ergebnisse entwickelten wir ein vorläufiges Kompetenzniveau-Modell, das in Abbildung 6 dargestellt ist und für den Aspekt der Modellbildung eine mögliche Kompetenzstruktur menschlichen Verhaltens in dynamischen Systemen beschreibt. Hierin integrierten wir die relevanten Merkmale Effektzahl und -qualität. Ergebnisse zu den verbliebenen Itemmerkmalen aus Tabelle 1 wurden hier nicht berichtet, weisen aber darauf hin, dass diese in ihrem Einfluss auf die Schwierigkeit vernachlässigt werden können. Bezogen auf die hier ausschließlich verwendeten 4x4-Systeme bilden gering und stark vernetzte Systeme zwei Kompetenzniveaus, innerhalb derer unterschiedliche Effektqualitäten als verschieden gut erkennbar angenommen werden. HE, ME und MA sind gleich gut erkennbar, gefolgt von ED. Am schwersten zu erschließen sind NE (Ergebnisse zu MA und ED aus nicht dargestelltem Experiment).

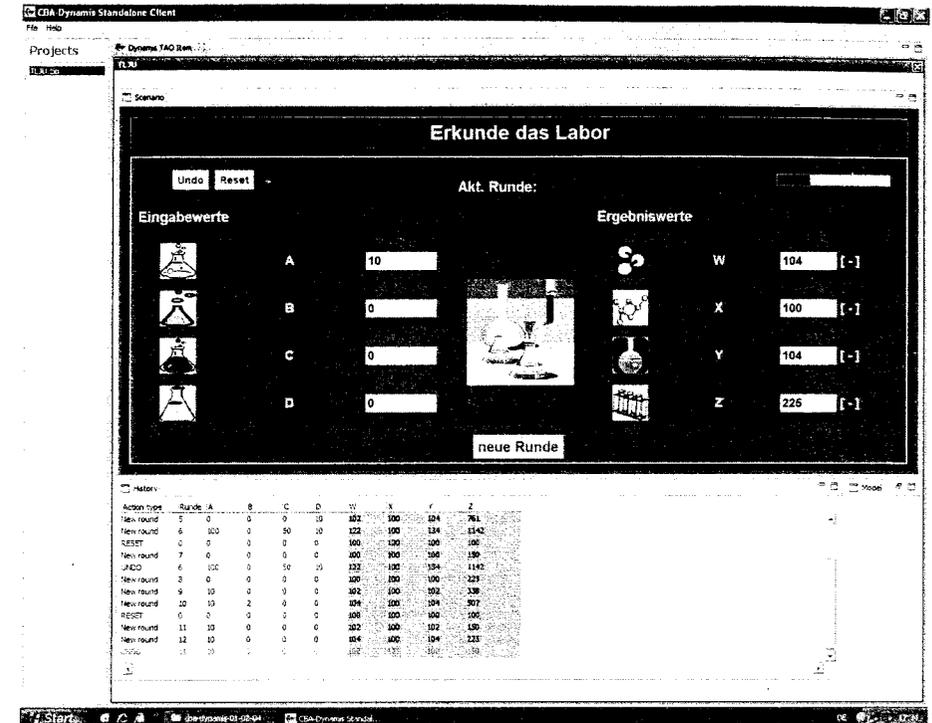


Abb. 5: Screenshot der MicroDYN-Software.

Anmerkung: Im oberen Teil befinden sich links vier exogene und rechts vier endogene Variablen. Im unteren Teil ist die Interventionshistorie angezeigt. Rechts oben ist die Restzeit visualisiert.

### 6.2 Einschränkungen in der vorliegenden Studie

Aus inhaltlicher Sicht stellt sich die grundlegende Frage nach dem Wesen komplexer Problemlösefähigkeit und ob diese überhaupt im Rahmen standardisierter Testverfahren erfassbar ist. Diese Diskussion ist nicht neu, wird aber durch die Kürze und die Vielzahl an Systemen im MicroDYN-Ansatz aktuell. Abstriche in der ökologischen Validität sind u.E. unvermeidbar und im Übrigen bei allen psychologischen Konstrukten (bspw. Intelligenz) gegeben, wenn auch komplexe Probleme hiervon besonders betroffen sein mögen. Systemschwierigkeit auf einzelne Merkmale zurückzuführen ist nicht unumstritten, entsteht die Komplexität doch gerade aus dem Zusammenspiel vieler Einzelvariablen. Die Befunde sprechen allerdings für das gewählte Vorgehen: Systemeigenschaften (als fixe Effekte verstanden) wirken sich deutlich und den Erwartungen entsprechend auf den Umgang mit diesen Systemen aus und unter psychometrischen Aspekten sind unabhängige Items mit kurzer Darbietungsdauer nachgerade unumgänglich. Ein weiterer wesentlicher Einwand tangiert den Realitätsbezug der verwendeten Probleme.

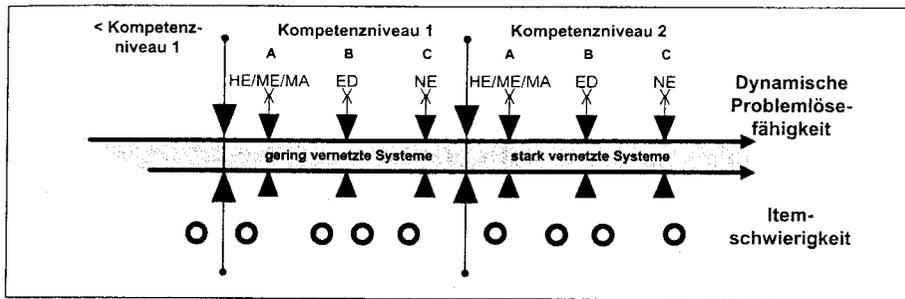


Abb. 6: Angenommenes Kompetenzmodell mit drei qualitativen Kompetenzniveaus und Items variierender Schwierigkeit

Tatsächlich weisen MicroDYN-Systeme an der Oberfläche nur wenige Gemeinsamkeiten mit alltäglichen Situationen auf, können diese aber formal gut modellieren (vgl. Beckmann/Guthke 1995). Im Übrigen erleben Proband/innen die Aufgaben als durchaus komplex und dynamisch. Auch ist es keinesfalls realitätsfern, sich mit alltäglichen Problemen (man denke z.B. an einen Fahrkartenautomat oder die Einstellung eines Thermostats) nur wenige Minuten zu beschäftigen und dann zum nächsten Problem zu wechseln („Wo ist Gleis 21a?“ oder Programmierung des DVD-Rekorders). Modellbildung beschreibt außerdem nur eine der fünf identifizierten Anforderungen an einen Problemlöser. Eine Integration der verbleibenden vier ist in MicroDYN aber möglich und wurde im Abschnitt 3 bereits angedeutet.

### 6.3 Ausblick

Im naturwissenschaftlichen Unterricht sind Schüler/innen häufig mit unbekanntem Systemen konfrontiert, in denen sie selbstständig explorieren und experimentieren müssen. Aber nicht nur dort, sondern auch in vielfältigen anderen Bereichen spielt die Fähigkeit, mit dynamischen Systemen umzugehen, eine wichtige Rolle. Existiert – unabhängig von der semantischen und situativen Einbettung – eine solche breite Kompetenz? In der hier berichteten Erhebung wurde semantisch ein (vorwissenarmes) Chemielabor gewählt. Die Frage nach dem Einfluss des situativen und semantischen Kontexts, die im Rahmen fachübergreifender Kompetenzdiagnostik zentral ist, wird über unterschiedliche Oberflächenstrukturen bei gleich bleibender Systemstruktur derzeit in MicroDYN untersucht. Wie aber sähe ein entsprechend domänenunspezifischer Test aus? Dieser Frage konnte bisher nur unzureichend nachgegangen werden, da mit Ausnahme von experimentellen *ad hoc* Konstruktionen keine adäquaten Erhebungsinstrumente existierten. An diesem Punkt möchten wir anknüpfen und Impulse für die Entwicklung einer fundierten Messung geben, da dynamische Problemlösefähigkeit als cross-curriculare Kompetenz in seiner Relevanz für Bildung und Unterricht keinesfalls zu unterschätzen ist. Sie stellt ein Konstrukt mit inkrementellem Potential dar, das in vielfältigen Situationen von Lernen und Unterricht relevant ist. Wir wünschen uns –

dies sollte deutlich geworden sein – eine weniger an inhaltlichen Konzepten, sondern mehr an testtheoretischer Güte orientierte Messtradition, wie es in anderen Bereichen pädagogischer Bildungsforschung schon seit Jahren Standard ist. Wenn diese Arbeit einen Anstoß in diese Richtung liefern kann, ist viel erreicht.

### Danksagung

Wir danken Ursula Pöll, Britta Veith und Sascha Wüstenberg für ihre Hilfe bei der Datenerhebung.

### Literatur

- Beckmann, J./Guthke, J. (1995): Complex problem solving, intelligence, and learning ability. In: Frensch, P.A./Funke, J. (Hrsg.): Complex problem solving: The European perspective. Hillsdale, NJ: Lawrence Erlbaum, S. 177–200.
- Dörner, D. (1986): Diagnostik der operativen Intelligenz. In: Diagnostica 32, S. 290–308.
- Funke, J. (1995): Experimental research on complex problem solving. In: Frensch, P.A./Funke, J. (Hrsg.): Complex problem solving: The European perspective. Hillsdale, NJ: Lawrence Erlbaum, S. 243–268.
- Funke, J. (2001): Dynamic systems as tools for analysing human judgement. In: Thinking and Reasoning 7, S. 69–89.
- Funke, J. (2003): Problemlösendes Denken. Stuttgart: Kohlhammer.
- Funke, J./Frensch, P.A. (2007): Complex problem solving: The European perspective – 10 years after. In: Jonassen, D.H. (Hrsg.): Learning to solve complex scientific problems. New York: Lawrence Erlbaum, S. 25–47.
- Klieme, E./Funke, J./Leutner, D./Reimann, P./Wirth, J. (2001): Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. In: Zeitschrift für Pädagogik 47, S. 179–200.
- Klieme, E./Leutner, D./Wirth, J. (Hrsg.) (2005): Problemlösekompetenz von Schülerinnen und Schülern. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Latour, T./Martin, R. (2007): TAO, an open and versatile computer-based assessment platform based on semantic web technology. In: ERCIM News 71, S. 32–33.
- Leutner, D./Wirth, J./Klieme, E./Funke, J. (2005): Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000. In: Klieme, E./Leutner, D./Wirth, J. (Hrsg.): Problemlösekompetenz von Schülerinnen und Schülern. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 21–36.
- Müller, H. (1993): Komplexes Problemlösen: Reliabilität und Wissen. Bonn: Holos.
- Wagener, D. (2001): Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation. Lengerich: Pabst Science Publishers.

### Anschrift der Autoren

Dipl.-Psych. Samuel Greiff, Psychologisches Institut der Universität Heidelberg,  
Hauptstr. 47–51, D-69117 Heidelberg  
E-Mail: samuel.greiff@psychologie.uni-heidelberg.de

Prof. Dr. Joachim Funke, Psychologisches Institut der Universität Heidelberg, Hauptstr. 47–51,  
D-69117 Heidelberg  
E-Mail: joachim.funke@psychologie.uni-heidelberg.de