

MicroDYN – Ein neuer Ansatz zur Erfassung komplexer Problemlösefähigkeit

Samuel Greiff & Joachim Funke
Psychologisches Institut, Universität Heidelberg

Version 15.06.2009

Dipl.-Psych. Samuel Greiff (Korrespondenzadresse, zugleich Adresse für Korrekturfahnen), beruflich: Psychologisches Institut der Universität Heidelberg, Hauptstr. 47-51, 69117 Heidelberg; Tel. 06221-54-7613, Fax 06221-54-7388, samuel.greiff@psychologie.uni-heidelberg.de; privat: Eppelheimer Strasse 55, 69124 Heidelberg

Prof. Dr. Joachim Funke, Psychologisches Institut der Universität Heidelberg, Hauptstr. 47-51, 69117 Heidelberg; Tel. 06221-54-7305, Fax 06221-54-7388, joachim.funke@psychologie.uni-heidelberg.de; privat: Handschuhsheimer Landstrasse 95, 69121 Heidelberg

Schlagwörter:

Kompetenzdiagnostik, Problemlösen, PISA, Computerbasiertes Testen, Testverfahren, Psychometrie.

Competence diagnostics, problem solving, PISA, computer-based testing, measurement device, psychometrics.

Zeichenzahl (mit Leerzeichen): ca. 25,924

MicroDYN – Ein neuer Ansatz zur Erfassung komplexer Problemlösefähigkeit

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 dann 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 (Funke 2003), so dass eine Testung nur computerbasiert möglich ist.

Komplexes Problemlösen (KPL) in dynamischen Systemen hat in den vergangenen Jahrzehnten im Hauptinteresse experimentalpsychologischer Forschung gestanden. Demgegenüber vernachlässigt wurde die Individualdiagnostik von KPL, die lediglich vereinzelt Berücksichtigung fand (z.B. Beckmann/Guthke 1995; Wagener 2001). Zugleich ist ein aufkeimendes Interesse an cross-curricularen Kompetenzen und damit auch an KPL in internationalen Bildungsstudien wie PISA zu beobachten (Klieme/Leutner/Wirth 2005). Als erste Konsequenz wurden in der nationalen Ergänzungsstudie Deutschlands zu PISA 1999 bereits Hinweise auf die Messbarkeit und das Potential dynamischer Problemlösefähigkeit (wir betrachten KPL und dynamisches Problemlösen als identisch, da KPL im Gegensatz zum analytischen Problemlösen in sich stets dynamisch ist) erbracht (Wirth/Funke 2005): Ein semantisch in den Kontext der Raumfahrt eingebetteter finiter Automat konnte inkrementell zu Testintelligenz Varianz in der Schülerleistung aufklären. Explorative Faktorenanalysen, lineare Strukturgleichungsmodelle und multidimensionale Skalierungen zeigten darüber hinaus, dass KPL, analytisches Problemlösen, fachspezifische Kompetenzen und Testintelligenz zwar korrelierte, jedoch voneinander abgrenzbare Konstrukte waren; KPL ließ sich dabei am besten separieren (Wirth/Leutner/Klieme 2005).

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 vom Problemlöser 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.

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 überprüfbaren 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 Einverständnis 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 Item, was grundlegenden psychometrischen Anforderungen widerspricht. Einige Autoren versuchen dieses Problem über die parallele Vorgabe mehrerer unabhängiger Teilsysteme (z.B. Müller 1993; Wagener 2001) oder über mehrere Fragen zu einem System (wie im finiten Automaten aus PISA) zu lösen. Dies macht die Items aber nur scheinbar unabhängiger und löst das Problem nicht, sondern verlagert es lediglich.

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 (siehe 3) wurde verschiedentlich als ökologisch valide bezeichnet und ist bereits häufig in experimentellen Arbeiten als Indikator der Problemlöseleistung verwendet worden (Funke 1993, 2001) – dort allerdings als *one item testing*. Wir wählen nun einen modifizierten Ansatz: Anstatt nur ein einzelnes Item darzubieten, bearbeiten Probanden 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.

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 einen Problemlöser 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. Wirkt eine endogene Variable auf andere endogene, ist dies ein *Nebeneffekt*. Wirkt sie hingegen auf sich selbst (mit einem Gewicht $\neq 1$), wird dieser Spezialfall eines Nebeneffektes *Eigendynamik* (als Wachstums- oder Schrumpfungsprozess) genannt. NE und ED können nicht aktiv manipuliert, aber über die Verwendung adäquater Strategien entdeckt werden. Zusätzlich können alle Effekte im Hinblick auf ihre Pfadgewichte variieren.

Abbildung 1 etwa hier einfügen

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

In der *Explorationsphase* (a) können Probanden das System eigenständig und vollständig frei explorieren. Ihre Aufgabe besteht darin, sich mit dem System und seiner Struktur vertraut zu machen. Die hinterlegten Verknüpfungen sind dabei für die Probanden 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 Probanden 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.

Probanden 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 Eigenschaften 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.

Tabelle 1 etwa hier einfügen

4.1 Design

In einem Messwiederholungsdesign (n = 48; 39 weiblich, 9 männlich; Alter M = 23.42, S = 3.02) bearbeiteten Probanden 15 MicroDYN-Items mit einer Gesamtdauer

von ca. 70 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 *Effektqualität* integrierten wir. Der 3-stufige Faktor *Effektqualität* (HE, ME, NE; siehe Abbildung 1; MA und ED wurden nicht berücksichtigt; Ergebnisse hierzu werden in der Diskussion erwähnt) und der 2-stufige Faktor *Effektzahl* (2 & 4; siehe Abbildung 2) wurden in einem 2x3-Design vollständig gekreuzt. Weiter wurde der 3-stufige Faktor *Zahl der Variablen* (2x2, 3x3, 4x4; siehe Abbildung 3) isoliert variiert. Über Kovarianzanalysen wurde der Einfluss der Itemposition auspartialisiert.

Abbildung 2 etwa hier einfügen

Abbildung 3 etwa hier einfügen

4.2 Hypothesen

In der vorliegenden Untersuchung wurden nur drei der in Tabelle 1 aufgeführten Faktoren getestet: *Effektqualität*, *Effektzahl* und *Variablenzahl*. Für *Effektqualität* (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 EQxEZ vermuten wir keine statistisch bedeutsamen Effekte. Bei steigender *Variablenzahl* sollte ceterus paribus die Leistung der Probanden sinken.

4.3 Abhängige Variable

Die Wissensabfrage erfolgte nach der Methode der Kausaldiagramm-Analyse (Funke 1995). Als Indikator wurde die *Güte der Kausaldiagramme* (GdK) in einer 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 hatten sich in einer Simulationsstudie mit insgesamt über 40 möglichen Indikatoren 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 ($p < .05$).

Tabelle 2 etwa hier einfügen

Abbildung 4 etwa hier einfügen

Eine weitere Analyse der Daten gibt Hinweise auf folgende Effekte: (a) Die Probanden hatten deutliche Schwierigkeiten, NE zu erkennen und neigten dazu, diese irrtümlich als ME zu interpretieren. (b) Es bestand eine positive, wenngleich schwache manifeste Korrelation zwischen Modellbildung und Prognose von $r = .30$ (hier nicht dargestellt), die suggeriert, dass der Einfluss von Systemmerkmalen auf Prognose anders ausfallen könnte. (c) Auftretende Trainingseffekte waren lediglich moderat. Zwar stieg das Leistungsniveau der Probanden während der Testung leicht an, der Zuwachs betrug aber weniger als eine halbe Standardabweichung.

5. Implementation

Die Software wurde in enger Kooperation mit dem DIPF (Frankfurt am Main) und der Firma SOFTCON (München) entwickelt. Die endgültige Version (verfügbar ab Herbst 2009) ist als Autorensystem in die frei zugängliche Plattform TAO (Reeff/Martin in press) 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 Probanden in der Explorationsphase,

vorherige Schritte zu korrigieren. In der Steuerphase werden zusätzlich extern vorgegebene Zielwerte angezeigt. Der Verlauf wird am unteren Seitenrand dokumentiert.

Abbildung 5 etwa hier einfügen

6. Diskussion

6.1 Kompetenzmodell

Auf Grundlage der Ergebnisse entwickelten wir ein vorläufiges Kompetenzmodell, 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 Kompetenzstufen, innerhalb derer unterschiedliche Effektsqualitäten 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).

Abbildung 6 etwa hier einfügen

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. Tatsächlich weisen MicroDYN-Systeme an der Oberfläche wenige Gemeinsamkeiten mit alltäglichen Situationen auf, können diese aber formal gut modellieren (Beckmann/Guthke 1995). Im Übrigen erleben Probanden 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-Recorders). Modellbildung beschreibt außerdem nur eine der fünf identifizierten Anforderungen an einen Problemlöser. Eine Integration der verbleibenden vier ist in MicroDYN aber möglich und wurde unter 3 bereits angedeutet.

6.3 Ausblick

Im naturwissenschaftlichen Unterricht sind Schüler 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 Eigenschaft? Wie sähe ein entsprechend domänenunspezifischer Test aus? Diesen Fragen 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.

Autorenhinweis

Die hier berichteten Arbeiten wurden ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Az. Fu 173/11-1) im Schwerpunktprogramm

„Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

Literatur

- Beckmann, J./Guthke, J. (1995): Complex problem solving, intelligence, and learning ability. In: Frensch, P. A./Funke, J. (Eds.): Complex problem solving: The European perspective. Hillsdale, NJ: Lawrence Erlbaum, p. 177-200.
- Dörner, D. (1986): Diagnostik der operativen Intelligenz [Diagnosing operational intelligence]. In: Diagnostica, 32, S. 290-308.
- Funke, J. (1993): Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In: Strube, G./Wender, K.-F. (Eds.): The cognitive psychology of knowledge. Amsterdam: Elsevier Science Publishers (p. 313-330).
- Funke, J. (1995): Experimental research on complex problem solving. In: Frensch, P. A./Funke, J. (Eds.): Complex problem solving: The European perspective. Hillsdale, NJ: Lawrence Erlbaum, p. 243-268.
- Funke, J. (2001): Dynamic systems as tools for analysing human judgement. In: Thinking and Reasoning, 7, p. 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. (Ed.): Learning to solve complex scientific problems. New York: Lawrence Erlbaum, p. 25-47.
- Klieme, E./Leutner, D./Wirth, J. (Hrsg.) (2005): Problemlösekompetenz von Schülerinnen und Schülern [Problem solving competencies of students]. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Müller, H. (1993): Komplexes Problemlösen: Reliabilität und Wissen [Complex problem solving: Reliability and knowledge]. Bonn: HoloS.
- Reeff, J.-P./Martin, R. (in press): Use of the internet for the assessment of students' achievement. In Hartig, J./Klieme, E./Leutner, D. (Eds.): Assessment of competencies in educational settings. Göttingen: Hogrefe & Huber.
- Wagener, D. (2001): Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation [Psychological diagnostics with complex scenarios. Taxonomy, development, evaluation]. Lengerich: Pabst Science Publishers.

- Wirth, J./Funke, J. (2005): Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme [Dynamic problem solving: Development and evaluation of a new measuring device to control complex systems]. In: Klieme, E./Leutner, D./Wirth, J. (Hrsg.): Problemlösekompetenz von Schülerinnen und Schülern. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 55-72.
- Wirth, J./Leutner, D./Klieme, E. (2005): Problemlösekompetenz - Ökonomisch und zugleich differenziert erfassbar? [Problem-solving competence – can it be measured economically and differentiated?] In Klieme, E./Leutner, D./Wirth, J. (Hrsg.): Problemlösekompetenz von Schülerinnen und Schülern. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 73-82.

Abbildungen

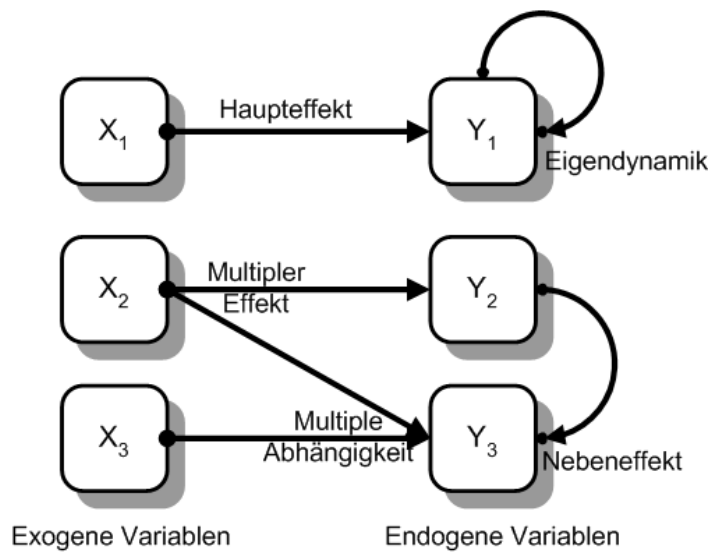


Abbildung 1: Struktur eines MicroDYN-Items mit 3 exogenen und 3 endogenen Variablen und den 5 Effektqualitäten.

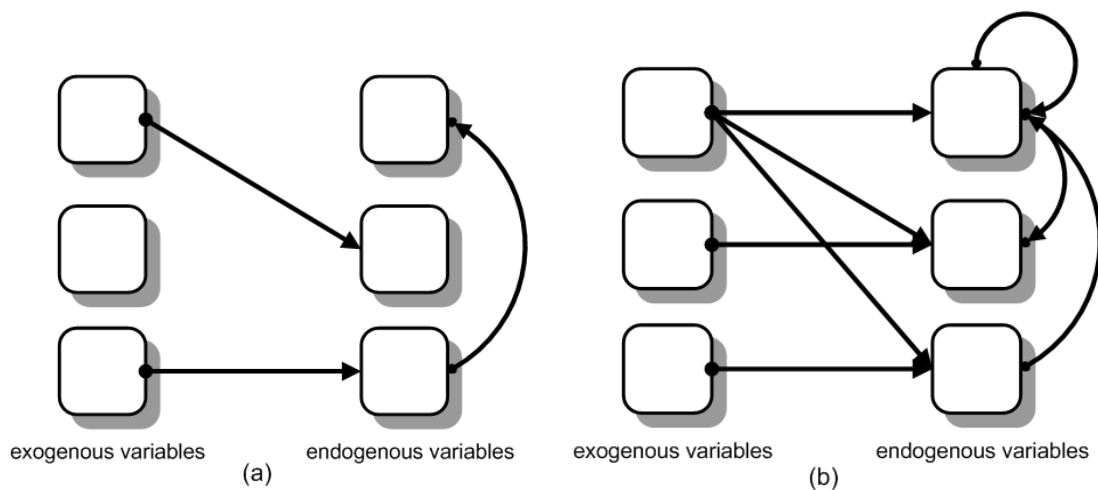


Abbildung 2: Zwei Items mit (a) niedriger bzw. (b) hoher Anzahl an Effekten bei konstant gehaltener Zahl der Variablen.

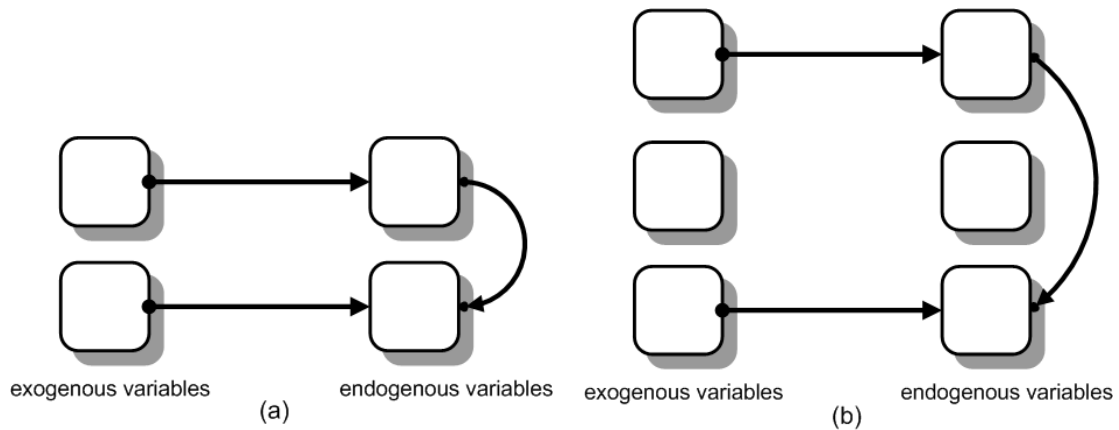


Abbildung 3: Zwei Items mit (a) 2 bzw. (b) 3 exogenen und endogenen Variablen bei konstant gehaltener Effektzahl.

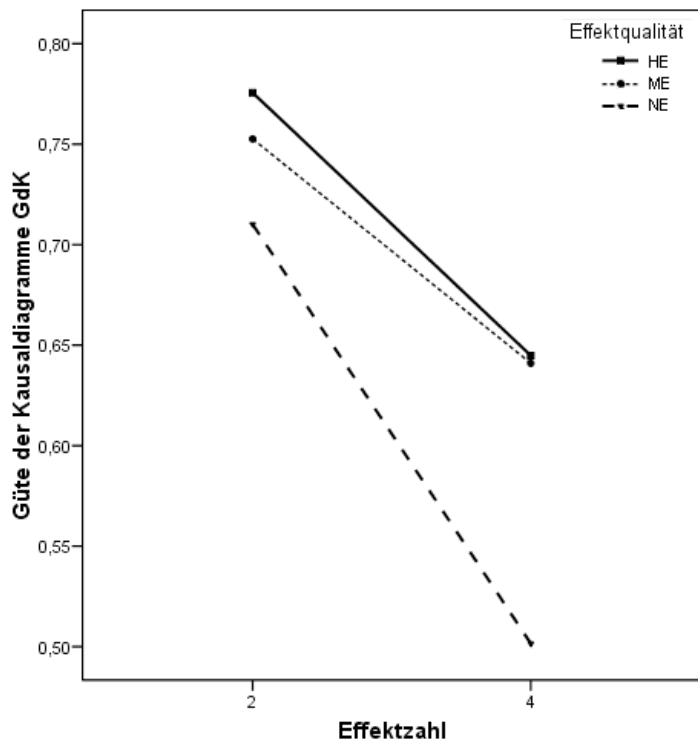


Abbildung 4: Mittlere Werte „Güte der Kausaldiagramme“ (GdK) in Abhängigkeit von der Effektzahl (2 & 4) und der Effektqualität (HE = Haupteffekt, ME = multipler Effekt, NE = Nebeneffekt).

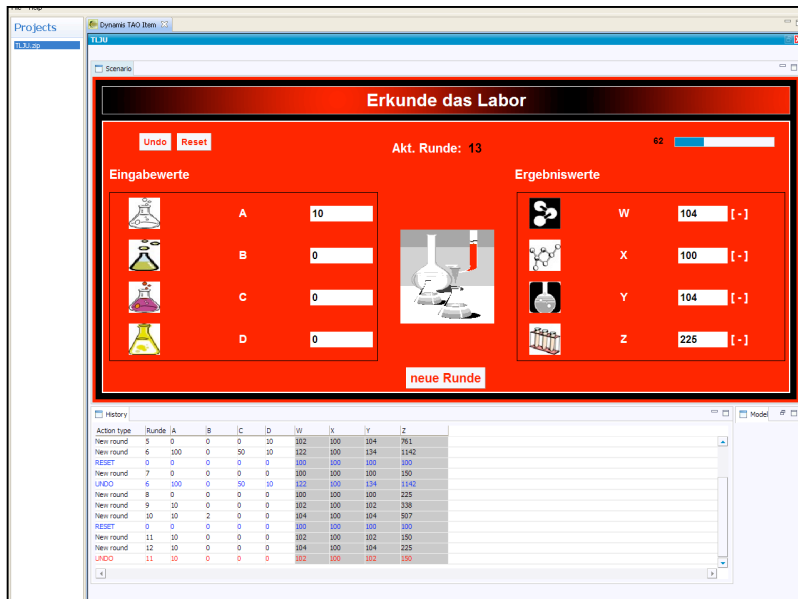


Abbildung 5: Screenshot der MicroDYN-Software. Im oberen Teil befinden sich links vier exogene und rechts vier endogene Variablen. Im unteren Teil ist die Interventionshistorie angezeigt.

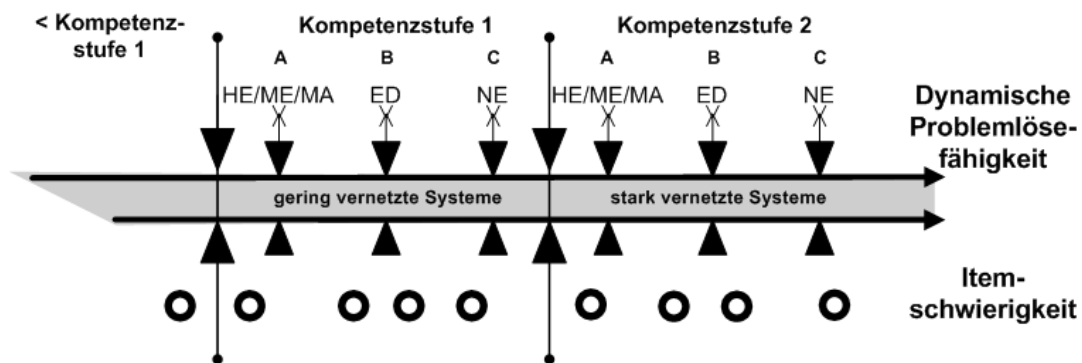


Abbildung 6: Angenommenes Niveaumodell mit qualitativen Kompetenzstufen und Items variierender Schwierigkeit.

Tabellen

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

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

Tabelle 2: ANOVA-Ergebnisse für die getesteten Effekte.

Faktor	Stufen	F	df	p	η^2	CI $_{\eta^2}(.95)$
Effektqualität	HE vs ME vs NE	9.95	2/90	<.001**	.18	[.05;.31]
Effektquantität	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]

CI = Konfidenzintervall; IA = Interaktion; EQ = Effektqualität; EZ = Effektzahl; HE = Haupteffekt; ME = multipler Effekt; NE = Nebeneffekt.