

**Indikatoren der Problemlöseleistung: Sinn und Unsinn verschiedener  
Berechnungsvorschriften**

***Bericht aus dem Psychologischen Institut***

Samuel Greiff & Joachim Funke<sup>1</sup>  
Psychologisches Institut, Universität Heidelberg

Version 31.12.2008

Dipl.-Psych. Samuel Greiff

Psychologisches Institut der Universität Heidelberg  
Hauptstr. 47-51  
69117 Heidelberg  
Tel. 06221-54-7613, Fax 06221-54-7388  
samuel.greiff@psychologie.uni-heidelberg.de

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

---

## Inhaltsverzeichnis

	<b>Seite</b>
<b>Abkürzungen</b>	<b>03</b>
<b>1.) Einführung</b>	<b>04</b>
<b>2.) Kausalanalysen</b>	<b>04</b>
2.1 Abhängige Variablen unter rationalen Gesichtspunkten (KD)	04
2.2 Abhängige Variablen für Kausalanalysen	04
2.3 Simulationsstudie für Kausalanalysen	08
2.3 Faktorenanalytische Befunde für Kausalanalysen	09
2.4 Abschließende Würdigung für Kausalanalysen	10
<b>3.) Steuerwerte</b>	<b>17</b>
3.1 Abhängige Variablen unter rationalen Gesichtspunkten (SW)	17
3.2 Simulationsstudie Steuerwerte	20
3.3 Faktorenanalytische Befunde: Steuerwerte	22
3.4 Abschließende Würdigung für Steuerwerte	22
<b>4.) Zusammenhang zwischen Kausalanalysen und Steuerwerten</b>	<b>28</b>
<b>Literatur</b>	<b>29</b>

## Abkürzungen:

Abhängige Variable	AV
Bezogen auf Relationswissen	rel (indiziert)
Bezogen auf Stärkewissen	stä (indiziert)
Eigendynamik	ED
Euklidische Distanz	EU
Güte der Kausaldiagramme:	GdK
Haupteffekt	HE
Hauptkomponentenanalyse	PCA
Individuelle Index Score	IIS
Item Response Theory	IRT
Kausaldiagramme	KD
Logarithmierte Steuerwerte	SW <sub>log</sub>
Multipler Effekt	ME
Nebeneffekt	NE
Sensitivität	P <sub>r</sub>
Steuerwerte	SW
Versuchsperson	Vp
Versuchspersonen	Vpn
Zufallskorrigierte Abstandswerte	SW <sub>log_cor</sub>

## Indikatoren der Problemlöseleistung: Sinn und Unsinn verschiedener Berechnungsvorschriften

### 1.) Einführung

Mikroszenarien werden, wie verschiedentlich diskutiert wurde (e.g. Gardner & Berry, 1995) als ökologisch valide betrachtet (für eine Einführung in Mikroszenarien siehe etwa Funke, 1990; Funke, 1993; Funke, Fahnenbruck, & Müller, 1986; Leutner, 1988). Demgegenüber wenig diskutiert, aber nicht minder bedeutsam bleibt die Frage nach dem Messbereich und der Wahl adäquater abhängiger Variablen. Gemeinhin werden rational abgeleitete, empirisch aber nicht validierte Maße wie *Güte der Kausaldiagramme* (GdK) von Funke (1992b) oder *Sensitivität* ( $P_r$ ) von Beckmann (1994) für den Modellerwerb sowie *logarithmierte Steuerwerte* ( $SW_{\log}$ ) ebenso von Funke (1992b) oder *zufallskorrigierte Abstandswerte* ( $SW_{\text{cor}}$ ) erstmalig vorgeschlagen von Kluge (2004, 2008) verwendet.

Diese Studie will in einem ersten Abschnitt eine Vielzahl denkbarer Kriteriumsmaße vorstellen und in ihren Eigenschaften beschreiben. In einer Simulationsstudie werden die Maße auf ihre Validität mit Expertenratings als Kriterium überprüft und eine abschließende Hauptkomponentenanalyse (PCA) beschreibt Dimensionen, die von den Variablen abgedeckt werden. Abhängige Variablen (AVs) werden konzeptuell und in der Darstellung nach Kausalanalysen (Kausaldiagramme; KD) und Steuerwerten (SW) getrennt. Abschließend wird eine Empfehlung hinsichtlich geeigneter AVs im Allgemeinen und für die Auswertung der vorliegenden Experimentaldaten (Greiff & Funke, 2008, in press) im Speziellen ausgesprochen.

### 2.) Kausalanalysen

#### 2.1 Abhängige Variablen unter rationalen Gesichtspunkten (KD)

Folgend werden AVs für die Kausalanalysen vorgestellt und ihre jeweiligen Stärken und Schwächen ansatzweise diskutiert.

In der Literatur unterschiedlich gehandhabt wird die Bewertung von Eigendynamiken von 1 (Funke, 1992a). Diese werden von den Vpn in der Regel nicht eingezeichnet, zugleich aber als „Treffer“ gewertet (Funke, 1992b). Die hier als Basis verwendeten Mikroszenarien beinhalten keine ED, was den Vpn mitgeteilt wurde, so dass die betreffenden Pfade gänzlich ausgeblendet wurden. Auf die Bewertung der Richtungsrelation (Funke, 1992b) wird daher verzichtet, da ausschließlich Pfade mit positiven Gewichten verwendet wurden.

## 2.2 Abhängige Variablen für Kausalanalysen

Eine Übersicht über sämtliche Indikatoren mit ergänzenden Informationen findet sich in Tabelle 1.

### 2.2.1 Güte des Systemwissens (GdS)

$$GdS = \frac{R}{R + F} * \frac{R}{R_{Max}}$$

mit R = Trefferzahl  
F = Fehlerzahl  
R<sub>Max</sub> = maximale Trefferzahl

GdS setzt die Zahl richtiger Elemente in Beziehung zu den falschen sowie den maximal richtigen und war eines der ersten Problemlösemaße im Kontext dynamischer Systeme. Probleme, wie etwa ein möglichenfalls undefinierter Nenner, werden ausführlich von Funke (1992b) diskutiert. GdS wird im Rahmen der Simulation sowie der PCA als GdS<sub>Rel</sub> und GdS<sub>Stä</sub> getrennt für Relation und Stärke einbezogen.

### 2.2.2 Güte der Kausaldiagramme (GdK)

$$GdK_{Relation} = (1 - p_{Rel}) * \frac{R_{Rel}}{R_{Max}} - (p_{Rel}) * \frac{F_{Rel}}{F_{Max}}$$

mit p<sub>Rel</sub> = Ratewahrscheinlichkeit  
R<sub>Rel</sub> = Trefferzahl  
F<sub>Rel</sub> = Fehlerzahl  
R<sub>Max</sub> = maximale Trefferzahl  
F<sub>Max</sub> = maximale Fehlerzahl

$$GdK_{Stärke} = (1 - p_{Stä}) * \frac{R_{Stä}}{R_{Max}} - (p_{Stä}) * \frac{F_{Stä}}{F_{Max}}$$

mit p<sub>Stä</sub> = Ratewahrscheinlichkeit  
R<sub>Stä</sub> = Trefferzahl  
F<sub>Stä</sub> = Fehlerzahl  
R<sub>Max</sub> = maximale Trefferzahl  
F<sub>Max</sub> = maximale Fehlerzahl

Dieses in späteren Jahren von Funke (Müller, 1993) entwickeltes Maß hat wohl im Bereich dynamischer Mikrosysteme die weiteste Verbreitung gefunden. Insbesondere das Globalmaß GdK<sub>Glo</sub> wird in experimentellen Paradigmen als Problemlösemaß herangezogen (vgl. Blech & Funke, 2005).

GdK kann getrennt für Relation und Stärke berechnet werden. Ein willkürlich gewichtetes Globalmaß vereint die beiden.

$$GdK_{Global} = Gew1 * GdK_{Rel} + Gew2 * GdK_{Stä}$$

mit Gew1 = Gewicht für GdK<sub>Rel</sub>  
Gew2 = Gewicht für GdK<sub>Stä</sub>

Eine detaillierte Ableitung sowie Begrenzungen des GdK-Maßes finden sich wiederum bei Funke (1992b). Auch die Ratewahrscheinlichkeit  $p$  wird dort näher erläutert, die hier auf 0.50 gesetzt wird.

In unserem empirischen Teil untersuchen wir neben  $GdK_{Rel}$  und  $GdK_{Stä}$  drei verschiedene Gewichtungen von  $GdK_{Global}$  (Gew1=0,75, Gew2=0,25; Gew1=0,50, Gew2=0,50; Gew1=0,25, Gew2=0,75).

### 2.2.3 Sensitivitätsmaß $P_r$ und Bias $B_r$

$$P_r = TR - FR$$

$$\text{wobei } TR = \frac{R + 0,05}{RT + 0,1}$$

$$FR = \frac{F + 0,05}{RT + 0,1}$$

mit TR = Trefferrate

FR = Fehlerrate

RT = Zahl möglicher Pfade

Dieser von Beckmann (1994) entwickelte Index entstammt der Signalentdeckungstheorie (Snodgrass & Corwin, 1988) und wird dort als Weiterentwicklung von GdK aufgefasst. Wir modifizieren  $P_r$  dahingehend, dass wir den Zähler wie oben beschrieben definieren. In der ursprünglichen Form war lediglich die Zahl richtiger Pfade vorgesehen. Unsere Erfahrung mit diesem Maß (Greiff & Funke, in press) sowie rationale Überlegungen weisen allerdings auf Probleme bei einer solchen Definition hin. Ergänzend vermag Beckmanns (1994) Argumentation, aus welchen Gründen er lediglich Relationswissen, nicht aber Stärkewissen berücksichtigt, nicht überzeugen, so dass hier getrennte Sensitivitäten (ebenso Bias) für Relations- und Stärkewissen ( $P_{r\_Rel}$ ,  $B_{r\_Rel}$ ,  $P_{r\_Stä}$ ,  $B_{r\_Stä}$ ) berechnet werden.

$$B_r = \frac{FR}{1 - P_r}$$

mit FR = Fehlerrate

$P_r$  = Sensitivität

Der ebenso der Signalentdeckungstheorie entlehnte Bias beschreibt die generelle Antwortneigung einer Vpn und ist gemeinsam mit der Sensitivität zu sehen. Je stärker die Tendenz einer Vpn, die Existenz von Pfaden anzunehmen, je geringer demnach ihr Reizschwelle, umso höher ist  $B_r$ . Eine genauere Beschreibung findet sich wiederum bei Beckmann (1994).

### 2.2.4 Zahl der Pfade

Die reine Anzahl der in den Kausaldiagrammen (KD) eingezeichneten Pfade stellt einen einfachen, aber möglicherweise hinreichend genauen Indikator der Systemerkennensleistung dar und findet sich in der ein oder anderen Weise in allen verwendeten Maßen. Hier werden sechs Variablen vorgeschlagen, wobei a) bis c) für einen Vergleich zwischen Systemen unterschiedlicher Pfadanzahlen prinzipiell ungeeignet erscheinen.

- a)  $Z_{\text{corr}}$  = Anzahl der richtig erkannten Pfade
- b)  $Z_{\text{fal}}$  = Anzahl der falschen Pfade
- c)  $Z_{\text{diff}}$  = Differenz aus a) und b)
- d)  $Z_{\text{rel}_{\text{cor}}}$  = Anzahl der richtig erkannten Pfade relativiert an Anzahl der vorhandenen Pfade
- e)  $Z_{\text{rel}_{\text{fal}}}$  = Anzahl der falschen Pfade relativiert an Anzahl möglicher falscher Pfade
- f)  $Z_{\text{rel}_{\text{diff}}}$  = Differenz aus d) und e)

Sämtliche Indices werden einmal für Relations- und einmal für Stärkewissen berechnet (siehe Tabelle 1).

### 2.2.5 Ordinal Scoring

Dieser aus der Item Response Theory (IRT) entlehnte Begriff (Bond & Fox, 2001) beschreibt die im Gegensatz zu den bisher beschriebenen Indikatoren ordinale Skalenannahme, wobei lediglich ganzzahlige Scores für ordinal bzw. mitunter lediglich qualitativ differente Abstufungen angenommen werden.

Neben einer globalen Bewertung des Kausaldiagramms erfolgt eine nach Effekten dezidierte sowie relativierte Auswertung.

Das globale ordinale Scoring  $OS_{\text{glo}}$  nimmt Werte zwischen 0 und 2 an. 0 indiziert ein fehlerhaftes Kausaldiagramm hinsichtlich der angegebenen Relationen, während 1 die Richtigkeit aller Relationen ausdrückt. 2 beschreibt, dass ergänzend zu dem relationalen korrektes Stärkewissen besteht.

Single Effects Ordinal Scoring  $OS_{\text{se}}$  bewertet wie im globalen Scoring mit den Kategorien 0, 1 und 2, allerdings nun jeden einzelnen Effekt innerhalb eines Systems. Zusätzlich wird das Fehlen falscher Pfade wiederum äquivalent bewertet. Um Vergleichbarkeit zu gewährleisten erfolgt eine Relativierung an der maximal möglichen Punktzahl. Als Beispiel diene ein Item mit einem Haupteffekt (HE) und einem Nebeneffekt (NE). Eine Vpn, die für HE und NE die richtige Relation, nicht aber die Stärke angibt und ergänzend keine falschen Pfade einzeichnet, erhält je 1 für HE und NE sowie 2 für das Fehlen falscher Pfade. Insgesamt dann einen Wert von  $2/3$  bei einer maximalen Punktzahl von 6.

### 2.2.6 Kritische Würdigung

Eine erste rationale Bewertung der zu überprüfenden Indikatoren lässt einen endgültigen Schluss nicht zu. Insbesondere fraglich erscheint das Abschneiden von Indikatoren, die Relations- und Stärkewissen parallel verbinden gegenüber den separaten Indizes. Wäre in diesem Falle der einzelne und somit ökonomischere Index gegenüber den beiden getrennten und demnach genaueren Variablen zu bevorzugen? Wie verhalten sich die ordinalen Werte mit verhältnismäßig wenig Streuung zu den anderen?

Als Arbeitshypothese wird angenommen, dass die genauesten Ergebnisse bei Berücksichtigung von Relation und Stärke zu erwarten sind. Die Anzahl der Pfade werden sich vermutlich als zu ungenau und wenig trennscharf erweisen. Ergänzend erwähnt sei, dass insbesondere GdK,  $P_r$  und  $B_r$  insofern einen konzeptuellen

Vorsprung genießen, als sie sich bereits in der experimentellen Problemlöseforschung etabliert haben und ihre Verwendung eine Vergleichbarkeit mit bereits bestehenden Studien erlaubt.

Selbstredend ist eine Vielzahl weiterer Indikatoren denkbar, auf deren Berechnung verzichtet wird. Die beschriebenen Indizes decken eine weite Spannbreite ab. Dies gilt im Übrigen auch für die Steuerwerte im folgenden Abschnitt.

## 2.3 Simulationsstudie für Kausalanalysen

Wir führen eine Simulationsstudie durch, in der die Rangfolge der genannten Problemlösemaße mit einer aus Expertenratings gewonnenen „wahren“ Rangreihe verglichen werden. (die Expertenratings können auf Anfrage von den Autoren zur Verfügung gestellt werden) Im Anschluss an eine Darstellung der simulierten Vpn werden die Verteilungen deskriptiv dargestellt. Anschließend folgt die Erörterung der Korrelationsanalysen.

Simulationen erfolgten für ein einfaches mit einem HE und einem multiplen Effekte (ME) sowie für ein schweres Item mit einem HE, einem ME und zwei NE (die genaue Itemstruktur ist auf Anfrage bei den Autoren erhältlich). Besondere Relevanz kommt den simultanen Analysen beiden Items zu. Es ist nicht befriedigend, wenn ein Indikator innerhalb eines Items streng monotone Relationen aufweist. Vielmehr müssen diese auch zwischen Systemen bestehen bleiben.

### 2.3.1 Simulierte Vpn

Folgende Personen wurden simuliert, um eine möglichst hohe Bandbreite des Leistungsspektrums abzubilden:

	Erwarteter Rank
a. Vpn markiert ausschließlich richtige Pfade	1
b. Vpn mit intuitiv überdurchschnittlicher Leistung	2
c. Vpn mit 2/3 richtigen und 1/3 falschen Pfaden	3
d. Vpn mit intuitiv durchschnittlicher Leistung	4
e. Vpn mit 1/3 richtigen und 2/3 falschen Pfaden	5
f. Vpn markiert 50% der möglichen Pfade zufällig	6
g. Vpn mit intuitiv unterdurchschnittlicher Leistung	7
h. Vpn markiert alle Pfade	8,5
i. Vpn markiert keine Pfade	8,5
j. Vpn markiert alle falschen und keine richtigen Pfade	10

Die erwartete Rangreihe laut Expertenratings war identisch für beide Items. Die dezidierten Leistungen der simulierten Versuchspersonen können bei den Autoren erfragt werden.

Die Personen wurden einmal ausschließlich für Relationswissen (keine Angaben hinsichtlich Stärkewissen) und einmal für Stärkewissen (komplett richtige Angaben hinsichtlich Relationswissen) simuliert, so dass insgesamt 20 Personen im fiktiven Datensatz zu finden sind. Auf eine Vermischung zwischen Relations- und Stärkewissen innerhalb einer Person wurde verzichtet, da jede hieraus abgeleitete

erwartete Rangreihe eine implizite Gewichtung von Stärke- vs. Relationswissen beinhalten und die Validität der Expertenratings vermindern würde.

### 2.3.2 Deskriptive Darstellungen (univariat)

Deskriptive Statistiken der Simulation finden sich für die Kausalanalysen in Tabelle 2. Auffällig ist hier – genauso wie für die Steuerwerte -, dass Werte ohne festgelegtes Maximum oder Minimum (mit variierenden Wertebereichen) für Vergleiche zwischen verschiedenen Items problematisch sind, da identisches Verhalten zu unterschiedlichen Indikatorausprägungen führt. Dies gilt insbesondere für die nicht relativierten Pfadzahlen, aber auch für  $P_r$ . Aufgrund der simulierten Personen würde man erwarten, dass die Verteilungen für beide Items nahezu identisch sind. Als ebenso hinderlich ist eine geringe Zahl möglicher Wertausprägungen zu verstehen (was sich nicht unbedingt in einer geringen Streuung ausdrückt), was vornehmlich die ordinal gescorten Indikatoren  $OS_{glo}$  und  $OS_{SE}$  betrifft.

### 2.3.3. Korrelationsanalysen

Für alle Indikatoren wurde die Rangkorrelation zwischen dem erwarteten Rang und dem tatsächlichen Rang berechnet. Für diejenigen Indikatoren, die sachlogisch paarweise zu verstehen sind, wurde zusätzlich eine multiple Regression nach dem Einschlussverfahren berechnet. Diese Berechnungen wurden getrennt für die Items als auch über beide Items hinweg durchgeführt. Die Ergebnisse finden sich in Tabelle 3.

Erwartungsgemäß werden die Rangreihen innerhalb eines Items besser vorhergesagt als zwischen zwei (unterschiedlich schwierigen) Items, wobei von einem validen Indikator zu erwarten ist, dass er auch Leistungen zwischen unterschiedlichen Items gut abzubilden vermag, so dass die letztgenannte Analyse ausschlaggebend ist.

Auffällig ist, dass insbesondere GdS- und GdK-Indikatoren gut abschneiden; insbesondere  $GdK_{Glo,7525}$  vermag die erwartete Rangreihe präzise zu reproduzieren. Insgesamt ist Relationswissen für die Rangreihe bedeutsamer als Stärkewissen, was sich nicht aus der Natur der simulierten Vpn ableiten lässt. Nicht zu empfehlen sind Sensitivität und Bias sowie unrelativierte Pfadzahlen. Insbesondere die Anzahl der falschen Pfade bildet Leistung nicht adäquat ab, sondern ist bestenfalls in Kombination mit richtigen Pfaden zu verwenden. Erstaunlich gut schneidet  $OS_{glo}$  ab, berücksichtigt man, dass es nur drei mögliche Ausprägungen hat.

Durch die Verwendung multipler Korrelationen ändert sich die Rangreihe nur unwesentlich. Auffällig ist, dass einige multiple Korrelationskoeffizienten geringer sind als ihre bivariaten Pendanten, was bei stark ausgeprägter Multikollinearität und dem Einschluss mehrerer Prädiktoren zu erwarten ist (da die sehr instabilen Beta-Gewichte zu geringeren Koeffizienten führen können). Betrachtet man die Eigenwerte der Krossproduktmatrizen sowie andere Kollinearitätsdiagnostica (nicht dargestellt), so liegen in den beobachteten Fällen immer bedenkliche lineare Abhängigkeiten zwischen den Prädiktoren vor. Eine ergänzende Erklärung besteht in den unterschiedlichen Skalenniveaus. Während die bivariaten Korrelationen auf Rangdaten basieren, legt die multiple Regression Intervalldaten zugrunde. Auch hieraus können sich gewisse Abweichungen in den Koeffizienten erklären. Eine

gemeinsame Betrachtung der multikollinearen Prädiktorenpaare erübrigt sich damit weitestgehend.

Abschließend empfehlen wir anhand der Simulationsstudie GdK<sub>Glo\_7525</sub>, da es in der Gesamtheit präzise und ökonomisch ist.

## 2.3 Faktorenanalytische Befunde für Kausalanalysen

Eine Hauptkomponentenanalyse (PCA) mit anschließender Varimax-Rotation (mit Kaiser Normalisierung) soll Aufschluss über orthogonale Komponenten in den Indikatoren geben. Wird nur eine Facette angesprochen, oder beschreiben verschiedene Indikatoren unterschiedliche Aspekte? Die meisten Indikatoren sind mathematisch voneinander abhängig, so dass hohe Ladungen auf einem Faktor erwartet werden dürfen. Diese Abhängigkeiten und die Aufnahme dichotomer Indikatoren verletzen die Annahmen der PCA allerdings nicht, wenn diese rein datenreduzierend verstanden wird.

Einleitend werden deskriptive Statistiken für n=46 Vpn dargestellt, auf deren Daten auch die PCA basiert.

### 2.3.1 Deskriptive Analysen

Mittelwerte, Standardabweichungen, Minimum und Maximum für die Kausalindikatoren finden sich in Tabelle 4. Die höhere Schwierigkeit des Items B spiegelt sich in den Mittelwerten wieder. Lediglich  $P_r$  und  $B_r$  sowie unrelativierte Indikatoren zu Pfadzahlen vermögen diese nicht adäquat abzubilden.

### 2.3.2 Faktorenanalyse

Zwei Faktoren wurden extrahiert (Eigenwert>10; Abfall im Screeplot; Differenz tatsächliche und reproduzierte Korrelationsmatrix gering; Kommunalitäten größtenteils >0,85), die gemeinsam über 80% der Varianz aufklären (Tabelle 5). Die beiden Faktoren klären nach Rotation nahezu identische Varianzanteile auf und sind als (artifizielle) Schwierigkeitsfaktoren zu verstehen. Sämtliche Indikatoren für Item A laden auf Faktor 1, diejenigen für Item B auf Faktor 2. Lediglich für die Indikatoren, die ausschließlich die Anzahl richtiger Pfade berücksichtigen, fallen die Kommunalitäten mit etwa 0,65 nicht ganz befriedigend aus.

Eine alternative Lösung mit vier Faktoren (Eigenwert>2; leichter Abfall im Screeplot; Differenz tatsächliche und reproduzierte Korrelationsmatrix gering; Kommunalitäten größtenteils >0,90) klärt über 90% Varianz auf, führt allerdings zu einer nicht einfach interpretierbaren Faktorenstruktur. Wiederum gibt es zwei Schwierigkeitsfaktoren. Der dritte und vierte Faktor (wiederum getrennt für die Items) beschreiben möglicherweise „richtige Lösungen“ in dem Sinne, dass Indikatoren, die stärkeres Gewicht auf richtige als falsche Aspekte legen, hier laden.

Wir haben uns für die erste Lösung entschieden. Detaillierte faktorenanalytische Befunde können auf Nachfrage von den Autoren zur Verfügung gestellt werden.

## 2.4 Abschließende Würdigung für Kausalanalysen

Rationale Überlegungen, Simulationsstudie, deskriptive Analysen und faktorenanalytische Befunde sprechen in der Zusammenschau für eine Verwendung des bereits etablierten Globalmaßes  $GdK_{Glo}$  in einer Gewichtung von 75% Relations- und 25% Stärkewissen. Die Befunde sprechen für eine ökonomische und zugleich präzise Abbildung tatsächlicher Problemlöseleistung.

Es überrascht, dass andere, durchaus etablierte Indikatoren nicht überzeugen können. Ähnliche Befunde werden bei den Steuerwerten zu beobachten sein.

**Tabelle 1: Überblick über Indikatoren für Kausalanalysen**

Indikator	vollständige Bezeichnung	Absatz	perfekter Score	alle Pfade markiert	keine Pfade markiert	Indikator = 0	Kombination	
GdS <sub>Rel</sub>	Güte des Systemwissens (Relation)	2.2.1	1	0,18	0	keine richtigen Pfade markiert		
GdS <sub>Stä</sub>	Güte des Systemwissens (Stärke)	2.2.1	1	0,18	0			
GdK <sub>Global</sub>	Güte der Kausaldiagramme (Global)	2.2.2	1	0	0	genauso viele Pfade richtig wie auch falsch		
GdK <sub>Relation</sub>	Güte der Kausaldiagramme (Relation)	2.2.2	0,5	0	0			
GdK <sub>Stärke</sub>	Güte der Kausaldiagramme (Stärke)	2.2.2	0,5	0	0			
P <sub>r_Rel</sub>	Sensitivitätsmaß (Relation)	2.2.3	0,99	-4,51	0	genauso viele Pfade richtig wie auch falsch Tendenz, keinerlei Pfade anzunehmen		
B <sub>r_Rel</sub>	Bias (Relation)	2.2.3	0,98	1,29	0,01			
P <sub>r_Stä</sub>	Sensitivitätsmaß (Stärke)	2.2.3	0,99	-4,51	0		genauso viele Pfade richtig wie auch falsch Tendenz, keinerlei Pfade anzunehmen	
B <sub>r_Stä</sub>	Bias (Stärke)	2.2.3	0,98	1,29	0,01			
Z <sub>corr_rel</sub>	Zahl der Pfade (richtig; Relation)	2.2.4	5	5	0	keine richtigen Pfade markiert		
Z <sub>corr_stä</sub>	Zahl der Pfade (richtig; Stärke)	2.2.4	5	5	0			
Z <sub>fal_rel</sub>	Zahl der Pfade (falsch; Relation)	2.2.4	0	23	0	keine falschen Pfade markiert		
Z <sub>fal_stä</sub>	Zahl der Pfade (falsch; Stärke)	2.2.4	0	23	0			
Z <sub>diff_rel</sub>	Differenz richtiger & falscher Pfade (Relation)	2.2.4	5	-18	0	genauso viele richtige wie falsche Pfade markiert		
Z <sub>diff_stä</sub>	Differenz richtiger & falscher Pfade (Stärke)	2.2.4	5	-18	0			
Z <sub>rel_corr_rel</sub>	Zahl der Pfade (richtig; Relation; relativiert)	2.2.4	1	1	0	keine richtigen Pfade markiert		
Z <sub>rel_corr_stä</sub>	Zahl der Pfade (richtig; Stärke; relativiert)	2.2.4	1	1	0			
Z <sub>rel_fal_rel</sub>	Zahl der Pfade (falsch; Relation; relativiert)	2.2.4	0	1	0	keine falschen Pfade markiert		
Z <sub>rel_fal_stä</sub>	Zahl der Pfade (falsch; Stärke; relativiert)	2.2.4	0	1	0			
Z <sub>rel_diff_rel</sub>	Differenz richtiger & falscher Pfade (Relation; relativiert)	2.2.4	1	0	0	genauso viele richtige wie falsche Pfade markiert		
Z <sub>rel_diff_stä</sub>	Differenz richtiger & falscher Pfade (Stärke; relativiert)	2.2.4	1	0	0			
OS <sub>glo</sub>	Ordinal Scoring (global)	2.2.5	2	0	0	fehlerhaftes Kausaldiagramm		
OS <sub>SE</sub>	Ordinal Scoring (single effects)	2.2.5	1	0,75	0,25	nur falsche Pfade markiert		

Anmerkungen: Alle Beispielwerte beziehen sich auf 4x4-System mit 4 Effekten (2x HE, ME & NE). Die Ratewahrscheinlichkeit beträgt für GdK 0,50 und für GdK<sub>Global</sub> ist Relations- und Stärkewissen identisch gewichtet. *Indikator =0*: Bedeutung eines Wertes von 0. *Kombination*: Die gemeinsam markierten Werten werden in der Simulationsstudie sowie den faktorenanalytischen Befunden aus sachlogischen Überlegungen gemeinsam betrachtet und evaluiert.

**Tabelle 2: Deskriptive Statistiken für die simulierten Kausalanalysen separat für die beiden Items**

Indikator	Item A (leicht)				Item B (schwer)			
	M	SD	Min	Max	M	SD	Min	Max
GdS <sub>Rel</sub>	0,66	0,43	0,00	1,00	0,66	0,42	0,00	1,00
GdS <sub>Stä</sub>	0,16	0,30	0,00	1,00	0,16	0,28	0,00	1,00
GdK <sub>Glo 5050</sub>	0,41	0,39	-0,50	1,00	0,39	0,39	-0,50	1,00
GdK <sub>Glo 7525</sub>	0,53	0,45	-0,75	1,00	0,52	0,45	-0,75	1,00
GdK <sub>Glo 2575</sub>	0,28	0,39	-0,50	1,00	0,27	0,38	-0,50	1,00
GdK <sub>Relation</sub>	0,33	0,27	-0,50	0,50	0,32	0,28	-0,50	0,50
GdK <sub>Stärke</sub>	0,08	0,22	-0,50	0,50	0,07	0,22	-0,50	0,50
P <sub>r_Rel</sub>	-0,04	0,29	-0,89	0,11	0,01	0,28	-0,82	0,18
B <sub>r_Rel</sub>	0,08	0,16	0,00	0,50	0,09	0,15	0,00	0,50
P <sub>r_Stä</sub>	-0,10	0,27	-0,89	0,11	-0,08	0,24	-0,82	0,18
B <sub>r_Stä</sub>	0,08	0,16	0,00	0,50	0,09	0,15	0,00	0,50
Z <sub>corr_rel</sub>	2,40	1,05	0,00	3,00	4,00	1,78	0,00	5,00
Z <sub>corr_stä</sub>	0,90	1,25	0,00	3,00	1,50	2,12	0,00	5,00
Z <sub>fal_rel</sub>	3,60	7,82	0,00	25,00	3,65	7,09	0,00	23,00
Z <sub>fal_stä</sub>	3,60	7,82	0,00	25,00	3,65	7,09	0,00	23,00
Z <sub>diff_rel</sub>	-1,20	8,26	-25,00	3,00	0,35	7,93	-23,00	5,00
Z <sub>diff_stä</sub>	-2,70	7,57	-25,00	3,00	-2,15	6,71	-23,00	5,00
Z <sub>rel_corr_rel</sub>	0,80	0,35	0,00	1,00	0,80	0,36	0,00	1,00
Z <sub>rel_corr_stä</sub>	0,30	0,42	0,00	1,00	0,30	0,42	0,00	1,00
Z <sub>rel_fal_rel</sub>	0,14	0,31	0,00	1,00	0,16	0,31	0,00	1,00
Z <sub>rel_fal_stä</sub>	0,14	0,31	0,00	1,00	0,16	0,31	0,00	1,00
Z <sub>rel_diff_rel</sub>	0,66	0,55	-1,00	1,00	0,64	0,55	-1,00	1,00
Z <sub>rel_diff_stä</sub>	0,16	0,45	-1,00	1,00	0,14	0,44	-1,00	1,00
OS <sub>glo</sub>	0,60	0,60	0,00	2,00	0,60	0,60	0,00	2,00
OS <sub>SE</sub>	0,48	0,29	0,00	1,00	0,52	0,30	0,00	1,00

Anmerkungen: M = Mittelwert; SD = Standardabweichung; Min = Minimum; Max = Maximum.

**Tabelle 3: Korrelationswerte für die simulierten Kausalanalysen gemeinsam sowie separat für die beiden Items**

Indikator	Item A (leicht)				Item B (schwer)				beide Items			
	r <sub>Rank</sub>	Rang <sub>r</sub>	MR	Rang <sub>MR</sub>	r <sub>Rank</sub>	Rang <sub>r</sub>	MR	Rang <sub>MR</sub>	r <sub>Rank</sub>	Rang <sub>r</sub>	MR	Rang <sub>MR</sub>
GdS <sub>Rel</sub>	0,90	7	0,96	2	0,90	6	<b>0,97</b>	<b>1</b>	0,90	5	<b>0,97</b>	<b>1,5</b>
GdS <sub>Stä</sub>	0,87	10			0,87	10,5			0,87	9		
GdK <sub>Glo 5050</sub>	0,94	2	<i>0,94</i>	3	0,92	2	<i>0,92</i>	4	0,93	2	<i>0,93</i>	3
GdK <sub>Glo 7525</sub>	<b>0,98</b>	<b>1</b>	<b>0,98</b>	<b>1</b>	<b>0,96</b>	<b>1</b>	<i>0,96</i>	2	<b>0,97</b>	<b>1</b>	<b>0,97</b>	<b>1,5</b>
GdK <sub>Glo 2575</sub>	0,91	4	<i>0,91</i>	6	0,87	10,5	<i>0,87</i>	10	0,89	6,5	<i>0,89</i>	6,5
GdK <sub>Relation</sub>	0,91	4			0,90	6			0,91	3,5		
GdK <sub>Stärke</sub>	0,78	15,5	0,92	4,5	0,62	20,5	0,92	4	0,70	17	0,92	4,5
P <sub>r_Rel</sub>	0,90	7			0,90	6			0,83	11,5		
B <sub>r_Rel</sub>	0,64	20	0,68	12,5	0,64	19	0,75	13	0,59	20	0,71	13
P <sub>r_Stä</sub>	0,21	23,5			0,12	24,5			0,16	25		
B <sub>r_Stä</sub>	0,69	19	0,75	11	0,69	17	0,77	11,5	0,69	19	0,73	11
Z <sub>corr_rel</sub>	0,71	17,5	0,89	8	0,69	17	0,88	8,5	0,55	23	0,80	10
Z <sub>corr_stä</sub>	0,84	12,5			0,83	12,5			0,83	11,5		
Z <sub>fal_rel</sub>	0,78	15,5			0,78	14,5			0,78	14,5		
Z <sub>fal_stä</sub>	0,58	21,5	0,63	14,5	0,58	22,5	0,67	14,5	0,58	21,5	0,65	14,5
Z <sub>diff_rel</sub>	0,90	7			0,90	6			0,83	11,5		
Z <sub>diff_stä</sub>	0,21	23,5	0,68	12,5	0,12	24,5	0,77	11,5	0,17	24	0,72	12
Z <sub>rel_corr_rel</sub>	0,71	17,5			0,69	17			0,70	17		
Z <sub>rel_corr_stä</sub>	0,84	12,5	0,89	8	0,83	12,5	0,88	8,5	0,83	11,5	0,88	8,5
Z <sub>rel_fal_rel</sub>	0,78	15,5			0,78	14,5			0,78	14,5		
Z <sub>rel_fal_stä</sub>	0,58	21,5	0,63	14,5	0,58	22,5	0,67	14,5	0,58	21,5	0,65	14,5
Z <sub>rel_diff_rel</sub>	0,91	4			0,90	6			0,91	3,5		
Z <sub>rel_diff_stä</sub>	0,78	15,5	0,92	4,5	0,62	20,5	0,92	4	0,70	17	0,92	4,5
OS <sub>glo</sub>	0,89	9	<i>0,89</i>	8	0,89	9	<i>0,89</i>	7	0,89	6,5	<i>0,89</i>	6,5
OS <sub>SE</sub>	0,86	11	<i>0,86</i>	10	0,91	3	<i>0,91</i>	6	0,88	8	<i>0,88</i>	8,5

Anmerkungen: Item A mit HE & ME. Item B mit HE, ME & 2x NE. |r<sub>Rank</sub>| = Betrag des Rangkorrelationskoeffizientens nach Spearman. MR = Multiple Regression mit den entsprechenden Prädiktoren zusammengefasst. Rang<sub>r</sub> = Rangplatz der Korrelation nach den bivariaten Analysen. Rang<sub>MR</sub> = Rangplatz der Korrelation nach der multiplen Regression. Kursiv gedruckte Korrelationen in der Spalte MR sind bivariat.

**Tabelle 4: Deskriptive Statistiken für die Kausalanalysen separat für die beiden Items**

Indikator	Item A (leicht)				Item B (schwer)			
	M	SD	Min	Max	M	SD	Min	Max
GdS <sub>Rel</sub>	0,90	0,22	0,17	1,00	0,49	0,33	0,00	1,00
GdS <sub>Stä</sub>	0,85	0,31	0,00	1,00	0,35	0,32	0,00	1,00
GdK <sub>Glo 5050</sub>	0,92	0,20	0,25	1,00	0,45	0,38	-0,31	1,00
GdK <sub>Glo 7525</sub>	0,94	0,16	0,29	1,00	0,50	0,36	-0,30	1,00
GdK <sub>Glo 2575</sub>	0,90	0,25	0,00	1,00	0,39	0,40	-0,37	1,00
GdK <sub>Relation</sub>	0,48	0,07	0,15	0,50	0,28	0,18	-0,15	0,50
GdK <sub>Stärke</sub>	0,44	0,15	-0,14	0,50	0,17	0,22	-0,24	0,50
P <sub>r_Rel</sub>	0,09	0,04	-0,04	0,11	0,01	0,14	-0,28	0,18
B <sub>r_Rel</sub>	0,02	0,04	0,00	0,14	0,10	0,09	0,00	0,29
P <sub>r_Stä</sub>	0,08	0,07	-0,25	0,11	-0,05	0,17	-0,39	0,18
B <sub>r_Stä</sub>	0,02	0,04	0,00	0,20	0,12	0,10	0,00	0,29
Z <sub>corr_rel</sub>	2,91	0,35	1,00	3,00	3,45	1,38	0,00	5,00
Z <sub>corr_stä</sub>	2,72	0,75	0,00	3,00	2,55	1,63	0,00	5,00
Z <sub>fal_rel</sub>	0,41	1,00	0,00	4,00	3,04	3,06	0,00	10,00
Z <sub>fal_stä</sub>	0,61	1,44	0,00	7,00	3,94	3,52	0,00	11,00
Z <sub>diff_rel</sub>	2,50	1,15	-1,00	3,00	0,40	3,91	-8,00	5,00
Z <sub>diff_stä</sub>	2,11	2,08	-7,00	3,00	-1,38	4,76	-11,00	5,00
Z <sub>rel_corr_rel</sub>	0,97	0,12	0,33	1,00	0,69	0,28	0,00	1,00
Z <sub>rel_corr_stä</sub>	0,91	0,25	0,00	1,00	0,51	0,33	0,00	1,00
Z <sub>rel_fal_rel</sub>	0,02	0,04	0,00	0,16	0,13	0,13	0,00	0,43
Z <sub>rel_fal_stä</sub>	0,02	0,06	0,00	0,28	0,17	0,15	0,00	0,48
Z <sub>rel_diff_rel</sub>	0,95	0,13	0,29	1,00	0,56	0,36	-0,30	1,00
Z <sub>rel_diff_stä</sub>	0,88	0,30	-0,28	1,00	0,34	0,44	-0,48	1,00
OS <sub>glo</sub>	1,61	0,77	0,00	2,00	0,26	0,61	0,00	2,00
OS <sub>SE</sub>	0,88	0,24	0,20	1,00	0,45	0,29	0,00	1,00

Anmerkungen: M = Mittelwert; SD = Standardabweichung; Min = Minimum; Max = Maximum.

**Tabelle 5: 2-Faktoren-Lösung für die Kausalanalysen**

	Faktor		
	1	2	kum
Eigenwert	20,75	19,94	40,69
% Varianz	41,49	39,87	81,36
Benennung	Kausalwerte Item A	Kausalwerte Item B	

Anmerkungen: kum = kumuliert über Faktoren.

### 3.) Steuerwerte

#### 3.1 Abhängige Variablen unter rationalen Gesichtspunkten (SW)

Eine Übersicht über sämtliche Indikatoren mit deren Wertebereich und ergänzenden Informationen findet sich in Tabelle 6.

##### 3.1.1 Logarithmierte Abstandswerte

$$SW_{\log\_all} = \frac{1}{k * l} \sum_{i=1}^{n=k} \sum_{i=1}^{n=l} \log(|es_{kl} - vs_k|)$$

mit k = Zahl der end. Variablen  
l = Zahl der Takte in Steuerphase  
es = erreichte Steuerwerte  
vs = vorgegebene Steuerwerte

Dieses von Funke (1992b) entwickelte und wohl am weitesten verbreitete Maß beschreibt den Abstand der erreichten im Verhältnis zu den vorgegebenen Steuerwerten. Die Logarithmierung verhindert eine übermäßige Gewichtung stark abweichender Werte, die gemeinhin – so die Annahme – mit einer geringeren Messgenauigkeit verbunden sind. Verwandt mit  $SW_{\log}$  ist die so genannte City-Block-Metrik (Müller, 1993), die aber bereits als überholt gilt und hier nicht berücksichtigt wird.

SW kann über alle Takte oder lediglich für den letzten Takt berechnet werden, wobei ersteres das validere Maß darstellt, da die Prozessleistung kontinuierlich abgebildet wird. Die Formel für den zweiten Fall lautet:

$$SW_{\log\_last} = \frac{1}{k} \sum_{i=1}^{n=k} \log(|es_k - vs|)$$

mit k = Zahl der end. Variablen  
es = erreichtes Steuerziel  
vs = vorgegebenes Steuerziel

Problematisch an logarithmierten Abstandswerten ist, dass sie in der Regel nicht berücksichtigen, wie viele der endogenen Variablen überhaupt aktiv manipulierbar und damit steuerbar sind. Möglicherweise stehen einzelne unverbunden im System oder werden nur durch endogene Effekte beeinflusst. Diese Kritik gilt im Übrigen für die meisten hier vorgestellten Maße, die sich auf die endogenen Variablen stützen. Einen Ausweg bieten Indikatoren, die den Eingabevektor berücksichtigen. Solche werden in 3.1.6 abgeleitet.

##### 3.1.2 Logarithmierte Abstandswerte – zufallskorrigiert

Kluge (2008) argumentiert, man müsse Steuerleistungen eines Probanden für die Eigenleistung unterschiedlicher Systeme korrigieren, um zufällige und systematische Steueranteile zu trennen. Sie schlägt vor, anhand von Simulationsstudien mit zufälligen Eingaben die Eigenleistung eines dynamischen Systems zu bestimmen

und diese von den individuellen Steuerwerten zu subtrahieren. Die von ihr angegebene Formel für diesen Individuellen Index Score (IIS) lautet:

$$SW_{corr} = IIS = 1 - \left| \frac{\text{Zielwert} - \text{Abweichung}}{\text{Zielwert} - \text{Zufallsaktivität}} \right|$$

Fraglich bleiben die zu wählenden Parameter in den Simulationsstudien wie Mittelwerte und Spannweite. Wir führen zwei Simulationen durch, einerseits mit einer Spannweite von -100 bis +100 (Mittelwert 0; Simulation A), andererseits mit einer Spannweite von 0 bis +100 (Mittelwert 50; Simulation B). Die Werte werden als gleichverteilt angenommen. Eine Alternative wäre, die Verteilungsparameter aus der Stichprobe zu gewinnen. Bei Kluge (2008) finden sich hierzu keinerlei Angaben.

Die IIS wird analog zu Kluge (2008) nur für den letzten Takt berechnet ( $SW_{corr\_last}$ ), da lediglich die gesamte Zufallsleistung des Systems interessiert.

Auffällig in obiger Formel ist weiterhin, dass die von Kluge (2008) angesprochene Logarithmierung nicht expliziert ist und – folgt man Kluges (2008) schriftlichen Ausführungen – über den gesamten Betragsterm geführt wird, was aber im Zweifel zu negativen und damit nicht logarithmierbaren Werten führt. Hier werden zuerst die Zielwerte, die Abweichung sowie die Zufallsaktivität logarithmiert und anschließend der Betrag gebildet. Es ist zu vermuten, dass auch Kluge dieses Vorgehen gewählt hat.

### 3.1.3 Euklidische Distanzen

$$Eu = \frac{1}{l} \sum_{l=1}^n \sqrt{\sum_{i=1}^n (es_{li} - vs_{li})^2}$$

$l$  = Zahl der Takte in Steuerphase

$i$  = Zahl der endogenen Variablen

$es$  = erreichtes Steuerziel

$vs$  = vorgegebenes Steuerziel

Neben logarithmierten Abweichungen gibt es eine Vielzahl anderer Distanzmaße die ebenso gut Verwendung finden könnten. Stellvertretend und ergänzend wird hier die euklidische Distanz (ausschließlich für den letzten Takt) berechnet.

### 3.1.4 Ordinal Scoring

Analog zu den Kausalanalysen kann auch bei den Steuerwerten eine kategorial orientierte Leistungsbewertung angelegt werden. Diese kann verschiedentlich erfolgen.

a) Dichotome Bewertung ( $OS_{SW}(dich)$ )

Wurde das Steuerziel insgesamt (über endogene Variablen hinweg) erreicht? Es erfolgt eine Bewertung mit den Kategorien 0 und 1.

#### b) Dichotome Bewertung ( $OS_{SW\_separat}(dich)$ )

Wurde das Steuerziel für eine einzelne endogene Variable erreicht? Jede endogene Variable wird demnach mit 0 oder 1 gescort; der Gesamtwert ist dann an der Zahl endogener Variablen zu relativieren, damit etwa ein Vergleich von 2x2- und 3x3-Systemen möglich ist.

#### c) Polytome Bewertung ( $OS_{SW}(poly)$ )

Eine leicht dezidiertere Bewertung gegenüber b) wird erreicht, indem drei Kategorien eingeführt werden. Nach wie vor beschreibt die Kodierung 2 eine erreichte Zielvorgabe. Eine 1 wird bei einer Verringerung des Abstandes im Vergleich zum Ausgangswert vergeben, während die 0 eine Vergrößerung, d. h. ein Steuern in die falsche Richtung, beschreibt. In dieser tritomen Weise wird jede Variable kodiert. Die Scores werden aufsummiert und an der maximalen Punktzahl relativiert.

### 3.1.5 Anzahl der Steuertakte

Hier wird lediglich die Zahl der benötigten Takte bis zur Erreichung der Zielwerte verwendet. Das schnelle Erreichen eines Zielzustandes sollte bessere Steuerungsfähigkeiten indizieren.

Sind keine Steuertakte vorhanden, werden diese Werte als fehlend verstanden, da die Aufgabe nicht in Angriff genommen wurde. Weiterhin ist diese Variable für die Simulationsstudie eine Konstante und wird dort nicht einbezogen.

### 3.1.6 Optimalität der Eingabewerte

Optimalität der Eingabewerte lenkt den Fokus von den Werten der endogenen, hin zu den exogenen Variablen. Abweichungen in den Zielwerten hängen insbesondere in ihrem Ausmaß stark von den (willkürlichen) Gewichten innerhalb eines Systems ab und spiegeln mitunter die Leistung eines Probanden nicht adäquat. Beckmann (1994) argumentiert, man müsse vielmehr den Eingabe- und nicht den Ausgabevektor untersuchen. Dies ließe sich im vorliegenden Fall auf nahezu alle berichteten Steuermaße übertragen.

Diesem verhältnismäßig neuen Ansatz wird insofern Rechnung getragen, als Indikatoren, die ihren Fokus auf den Eingabewerten haben, berechnet werden. Es handelt sich dabei um acht unterschiedliche Maße mit variierender Auflösung.

Die Komplexität der verwendeten Steueraufgaben ist gering, so dass bereits im Vorfeld Musterlösungen entwickelt wurden, in denen über drei Takte hinweg je ein Wert manipuliert werden musste, um die Zielvorgaben zu erreichen. Dieser Vektor wird als optimale Lösung verstanden und als Kriterium für einige der Indikatoren verwendet, obgleich geringe Unschärfen bestehen bleiben. Unabhängig hiervon sind – unter gewissen Randbedingungen – zu jedem Steuertakt optimale Lösungen berechenbar, die als individuelle Kriterien benutzbar sind. Eine Software hierzu ist derzeit in Arbeit.

### *Ordinale Optimalitätsmaße:*

#### a) Globale Optimalität $Opt_{glo}$

Jeder richtig manipulierte und richtig nicht manipulierte Wert innerhalb der drei Takte wird als Punkt gewertet. Die Gesamtpunktzahl wird am Maximum (in der Regel 12 bei 4x4-Systemen und 3 Steuertakten) relativiert.

#### b) Manipulation $Opt_{manipu}$ & Nicht-Manipulation $Opt_{non\_manipu}$

Neben einem globalen Optimalitätsindex werden zwei Subwerte analog zum Globalindex, einmal für die zu recht manipulierten und die zu recht nicht manipulierten Werte gebildet.

### *Quantitative Optimalität – logarithmiert ( $Eingabe_{sum}$ & $Eingabe_{sum\_corr}$ )*

Ausgehend von der Musterlösung werden die über alle Takte summierten Eingabewerte verglichen mit der optimalen Lösung. Analog zu den logarithmierten Abstandswerten für die endogenen Variablen werden die Werte wiederum logarithmiert. Der Mittelwert des logarithmierten Abstandes bildet den Indikator. Dies geschieht einmal ohne Korrektur und einmal mit Zufallskorrektur sensu Kluge (2008).

### *Quantitative Optimalität – weitere ( $Eingabe_{opt\_unrel}$ , $Eingabe_{opt\_rel}$ & $Eingabe_{opt\_eu}$ )*

$Eingabe_{opt\_unrel}$ , bildet die Differenz zwischen tatsächlicher und optimaler Eingabe. Dieser Indikator entspricht  $Eingabe_{sum}$ , allerdings ohne Logarithmierung.  $Eingabe_{opt\_eu}$  bildet die entsprechende euklidische Distanz. Interessant ist  $Eingabe_{opt\_rel}$ . Hier lautet die Formel:

$$Eingabe_{opt\_rel} = \frac{|EW - Opt|}{|Opt| + 1}$$

EW = Eingabewert

Opt = Optimaler Eingabewert

Für  $Eingabe_{opt\_unrel}$  &  $Eingabe_{opt\_rel}$  werden jeweils die mittleren Abstände berechnet.

### **3.1.7 Kritische Würdigung**

Analog zu den Kausalanalysen ist eine alleinige Entscheidung anhand rationaler Überlegungen nicht möglich, was die Notwendigkeit einer derartigen Simulationsstudie unterstreicht. Einige Maße sind bereits vielfältig verwendet in der Literatur, was ihnen vergleichbar zu den Kausalanalysen einen konzeptuellen Vorsprung verschafft. Dessen ungeachtet, bleibt es abzuwarten, ob nicht einfachere und eingängigere Maße ähnlich differenziert Steuerleistung abzubilden vermögen. Interessant ist aber insbesondere, die Differenz zwischen endogen und exogen basierten Variablen. Sachlogisch ist es sinnvoller, die Eingabevektoren und nicht die Ausgabevektoren zu berücksichtigen.

### **3.1.8 Strategiemaße und Prozessdaten**

In den hier verwendeten Mikrosystemen liegt es nahe, nicht nur die Leistung einer Person am Ende statisch zu erheben, sondern das Augenmerk auf Prozesse und Strategien zu richten. Erste Ansätze hinsichtlich Strategiemaßen finden sich bei Rollett (2008). Bei den hier berichteten Kausalanalysen handelt es sich um reine Statusdiagnostik. Die Prozesse aus der Explorationsphase werden vernachlässigt. Zu früh scheint es, verschiedene Strategie- und Prozessmaße miteinander zu kontrastieren, stecken diese doch noch in den Kinderschuhen. Lediglich bei den

Steuerdaten vermischen sich in den berichteten Indikatoren Prozess- und Statusaspekte.

## 3.2 Simulationsstudie Steuerwerte

### 3.2.1 Simulierte Vpn

Zur Lösung der Steueraufgaben sind bestimmte exogene Variablen zu manipulieren, andere wiederum sollten nicht manipuliert werden. Die Aufgaben wurden so gewählt, dass ausschließlich positive Werte zu verwenden waren, um die Steuerziele zu erreichen.

Folgende Personen wurden simuliert (die Expertenratings können auf Anfrage von den Autoren zur Verfügung gestellt werden):

	Erwarteter Rank
a. Vpn löst die Aufgabe korrekt	1
b. Vpn manipuliert die korrekten Werte, aber nur 2/3 mit dem richtigen Wert	2
c. Vpn manipuliert die korrekten Werte, aber nur 1/3 mit dem richtigen Wert	3
d. Vpn manipuliert 2/3 der Werte korrekt und belässt die anderen gleich 0	4
e. Vpn manipuliert 1/3 der Werte korrekt und belässt die anderen gleich 0	5
f. Vpn manipuliert 2/3 der Werte korrekt, manipuliert die anderen mit Werten ungleich 0	6
g. Vpn manipuliert 1/3 der Werte korrekt, manipuliert die anderen mit Werten ungleich 0	7
h. Vpn gibt keine Werte ein (lässt alles bei 0)	8
i. Vpn gibt ausschließlich falsche Werte ein (positiv)	9
j. Vpn gibt ausschließlich falsche Werte ein (negativ & positiv)	10
k. Vpn gibt ausschließlich falsche Werte ein (negativ)	11

Die erwartete Rangreihe laut Expertenratings war identisch für beide Items. Die dezidierten Leistungen der simulierten Versuchspersonen können bei den Autoren erfragt werden.

### 3.2.2 Deskriptive Darstellungen (univariat)

Deskriptive Statistiken der Simulation finden sich für die Steuerwerte in Tabelle 7. Betrachtet man die oben für die Kausalanalysen genannten Kriterien, dann ist hier zu vermuten, dass insbesondere die Optimalitätsindikatoren auf Eingabeseite (ohne die beiden zufallskorrigierten Werte) sowie die drei ordinal gescorten Indikatoren gute Reproduktionen der erwarteten Rangreihe liefern sollten. Da (nahezu) identische Vpn simuliert wurden sollten die Mittelwerte der Indikatoren zwischen Items (annähernd) identisch sein. Indikatoren, auf die dies nicht zutrifft, sind prinzipiell nicht geeignet, wie etwa – überraschenderweise –  $Eu$ ,  $SW_{\log\_all}$  und  $SW_{\log\_last}$ .

### 3.2.3 Korrelationsanalysen

Analog zu den Kausalanalysen wurden Rangkorrelationen zwischen dem erwarteten und dem beobachteten Rangplatz für alle Indikatoren für beide Items sowie einzeln für die Items berechnet. Entscheidend ist wiederum die Rangreihe über Items hinweg. Die Koeffizienten finden sich in Tabelle 8.

Rangkorrelationen für Indikatoren, die sich auf die endogenen Variablen beziehen, fallen tendenziell niedriger aus. Offensichtlich ist die Betrachtung von Eingabevektoren angezeigt. Auch eine Zufallskorrektur scheint keine Genauigkeitsvorteile zu bringen. Bei simultaner Betrachtung mehrerer Items führt dies, je nach Verteilungsannahme, sogar dazu, dass Zusammenhänge zwischen der Leistung und dem Indikator mitunter extrem verwischt werden.

Ein Maß, das Steuerleistung adäquat abbildet, sollte demnach Eingabewerte und nicht Ausgabewerte berücksichtigen und diese nach einem Optimalitätskriterium bewerten.  $Opt_{glo}$  scheint hier zuerst einmal geeignet, schneidet aber bei den empirischen Analysen schlecht ab.

## 3.3 Faktorenanalytische Befunde: Steuerwerte

### 3.3.1 Deskriptive Analysen

Mittelwerte, Standardabweichungen, Minimum und Maximum für die Kausalindikatoren finden sich in Tabelle 9. Ein geeigneter Indikator sollte die Schwierigkeitsdifferenz zwischen den Items abbilden.  $Opt_{glo}$  vermag dies nicht zu leisten, da  $V_{pn}$  Eingaben im schwierigen System offensichtlich zurückhaltender tätigen und die richtigerweise nicht manipulierten Werte ( $Opt_{non\_manipu}$ ) besser ausfallen.

### 3.3.2 Faktorenanalyse

Zwei Faktoren wurden extrahiert (Eigenwert  $>7$ ; Abfall im Screeplot; Differenz tatsächliche und reproduzierte Korrelationsmatrix gering; Kommunalitäten größtenteils  $>0,70$ ), die gemeinsam über 60% der Varianz aufklären (Tabelle 10). Die Faktoren klären nach Rotation identische Varianzanteile auf und können analog zur PCA für die Kausalwerte als Schwierigkeitsfaktoren verstanden werden, wenn auch nicht ganz so eindeutig. Für einen der zufallskorrigierten Indikatoren ( $log\_ch\_a$ ) liegt die Ladung erwartungswidrig in negativer Richtung. Mehrere Indikatoren zeigen keinerlei substantielle Ladungen auf einem der Faktoren ( $<0,45$ ). Entsprechend niedrig fallen einige Kommunalitäten aus. Die folgenden Indikatoren lassen sich in der gewählten zweifaktoriellen Lösung scheinbar schlecht einordnen:  $Eu$ ,  $OS_{sw}(dich)$ ,  $Takte$ ,  $Opt_{manipu}$  und  $Opt_{non\_manipu}$ .

Andere Faktorenlösungen waren nicht sinnvoll interpretierbar. Detaillierte faktorenanalytische Befunde können auf Nachfrage von den Autoren zur Verfügung gestellt werden.

### 3.4 Abschließende Würdigung für Steuerwerte

Die Befundlage hinsichtlich der Steuerwerte ist uneindeutiger als für die Kausalanalysen, so dass eine Entscheidung für einen Steuerindikator schwerer fällt. Offensichtlich ist es eine anspruchsvolle Aufgabe, Steuerleistung angemessen abzubilden.

Prinzipiell sind Optimalitätsindikatoren, die auf den Eingabewerten basieren, zu bevorzugen. Zufallskorrekturen sind insgesamt mit zu starken Unwägbarkeiten verbunden. Die ordinal gescorten Maße  $Opt_{glo}$ ,  $Opt_{manipu}$  und  $Opt_{non\_manipu}$  können nicht verwendet werden.  $Opt_{non\_manipu}$  ist bei schweren Items offensichtlich „besser“ ausgeprägt und verwischt Schwierigkeitsunterschiede in  $Opt_{glo}$ .  $Opt_{manipu}$  alleine reicht zur Leistungsabbildung nicht aus.

Sowohl eine Logarithmierung ( $Eingabe_{sum}$ ) als auch die Bildung der euklidischen Distanz ( $Eingabe_{opt\_eu}$ ) ist nicht mit erkennbaren Vorteilen verbunden.

Als (vorläufiges) Mittel der Wahl erscheinen  $Eingabe_{opt\_unrel}$  und  $Eingabe_{opt\_rel}$ . Unter theoretischen Aspekten erscheint die Verwendung von der relativierten Abweichung einer Eingabe von ihrem Optimum ansprechender. Ein Eingabewert von 7 bei einem Optimum von 5 sollte ein größerer Fehler sein, als eine Eingabe von 25 bei einem Optimum von 27. Möglicherweise sind die verwendeten Steueraufgaben nicht geeignet, um diesen Aspekt in den Indikatoren abzubilden (alle Steuerwerte waren verhältnismäßig ähnlich). Diesem theoretischen Vorteil gegenüber steht ein empirisch leicht besseres Abschneiden von  $Eingabe_{opt\_unrel}$ .

Eine endgültige Entscheidung kann immer nur auf dem Hintergrund des konkreten Forschungsgegenstandes getroffen werden. Wir entscheiden uns hier für das relativierte Maß.

**Tabelle 6: Überblick über Indikatoren von Steuerwerten**

Indikator	vollständige Bezeichnung	Absatz	Maximum	Minimum	perfekter Score	Indikator = 0
SW <sub>log_all</sub>	Logarithmierter Abstandswert (alle Takte)	3.1.1	$\infty$	$-\infty$	$-\infty$	"mittlere Abweichung ist 1"
SW <sub>log_last</sub>	Logarithmierter Abstandswert (letzter Takt)	3.1.1	$\infty$	$-\infty$	$-\infty$	
SW <sub>corr_last</sub>	Logarithmierter Abstandswert (zufallskorrigiert; letzter Takt)	3.1.2	$\infty$	$-\infty$	$-\infty$	"mittlere korrigierte Abweichung ist 1"
Eu	Euklidische Distanz	3.1.3	$\infty$	0	0	Steuerziele exakt erreicht
OS <sub>SW(dich)</sub>	Ordinal Scoring (dichotom)	3.1.4	1	0	1	Steuerziele nicht erreicht für keine Variable Steuerziel erreicht für alle Variablen Entfernung vom Steuerziel
OS <sub>SW_separat(dich)</sub>	Ordinal Scoring (variablenweise; dichotom)	3.1.4	1	0	1	
OS <sub>SW(poly)</sub>	Ordinal Scoring (polytom)	3.1.4	1	0	1	
Takte	Anzahl der Steuertakte	3.1.5	max. Taktzahl	1	1	Aufgabe nicht bearbeitet
Opt <sub>glo</sub>	Optimalität der Eingabe (global)	3.1.6	1	0	1	Ausschließlich die falschen Variablen manipuliert
Opt <sub>manipu</sub>	Optimalität der Eingabe (manipulierte Variablen)	3.1.6	1	0	1	Keine richtige Variable manipuliert
Opt <sub>non_manipu</sub>	Optimalität der Eingabe (nicht-manipulierte Variablen)	3.1.6	1	0	1	Alle falschen Variablen manipuliert
Eingabe <sub>sum</sub>	Summierte Eingabewerte (logarithmiert)	3.1.6	$\infty$	$-\infty$	$-\infty$	"mittlere Abweichung ist 1"
Eingabe <sub>sum_corr</sub>	Summierte Eingabewerte (logarithmiert; zufallskorrigiert)	3.1.6	$\infty$	$-\infty$	$-\infty$	"mittlere korrigierte Abweichung ist 1"
Eingabe <sub>opt_unrel</sub>	Optimalität der Eingabe (intervall; unrelativiert)	3.1.7	$\infty$	0	0	tatsächliche & optimale Eingabe identisch
Eingabe <sub>opt_rel</sub>	Optimalität der Eingabe (intervall; relativiert)	3.1.7	$\infty$	0	0	
Eingabe <sub>opt_eu</sub>	Optimalität der Eingabe (intervall; euklidisch)	3.1.7	$\infty$	0	0	

Anmerkungen: *perfekter Score*: Indikatorwert bei perfekter Lösung der Aufgabe. *Indikator = 0*: Bedeutung eines Wertes von 0. *Takte* bezieht sich auf Steueraufgaben mit drei Takten.

**Tabelle 7: Deskriptive Statistiken für die simulierten Steuerwerte separat für die beiden Items**

Indikator	<i>Item A (leicht)</i>				<i>Item B (schwer)</i>			
	<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>	<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
SW <sub>log_all</sub>	-1,11	1,03	-2,56	0,20	0,49	0,65	-0,27	1,45
SW <sub>log_last</sub>	-1,69	1,36	-4,00	0,15	-0,95	1,73	-4,00	1,34
SW <sub>corr_last_A</sub>	2,50	2,01	-0,22	5,94	-1,96	3,59	-8,29	2,79
SW <sub>corr_last_B</sub>	-2,78	2,24	-6,58	0,22	-0,45	0,83	-1,92	0,63
Eu	17,15	21,42	0,00	63,78	14,76	17,40	0,00	55,46
OS <sub>SW(dich)</sub>	0,09	0,30	0,00	1,00	0,09	0,30	0,00	1,00
OS <sub>SW_separat(dich)</sub>	0,52	0,26	0,25	1,00	0,36	0,32	0,00	1,00
OS <sub>SW(poly)</sub>	0,55	0,27	0,25	1,00	0,57	0,28	0,25	1,00
Takte	3,00	0,00	3,00	3,00	3,00	0,00	3,00	3,00
Opt <sub>glo</sub>	0,53	0,33	0,00	1,00	0,52	0,32	0,00	1,00
Opt <sub>manipu</sub>	0,48	0,31	0,00	1,00	0,45	0,31	0,00	1,00
Opt <sub>non_manipu</sub>	0,58	0,49	0,00	1,00	0,58	0,49	0,00	1,00
Eingabe <sub>sum</sub>	-0,94	2,23	-4,00	2,10	-1,03	2,19	-4,00	2,12
Eingabe <sub>sum_corr_A</sub>	-5,47	12,92	-23,21	12,19	-1,08	2,30	-4,21	2,23
Eingabe <sub>sum_corr_B</sub>	-0,44	1,03	-1,86	0,98	-0,48	1,02	-1,87	0,99
Eingabe <sub>opt_unrel</sub>	32,84	46,01	0,00	361,25	34,66	45,79	0,00	133,75
Eingabe <sub>opt_rel</sub>	18,04	25,58	0,00	198,39	18,04	24,89	0,00	64,38
Eingabe <sub>opt_eu</sub>	71,75	91,51	0,00	789,30	79,72	90,34	0,00	269,49

Anmerkungen: M = Mittelwert; SD = Standardabweichung; Min = Minimum; Max = Maximum.

**Tabelle 8: Korrelationswerte für die simulierten Steuerwerte gemeinsam sowie separat für die beiden Items**

Indikator	Item A (leicht)		Item B (schwer)		beide Items	
	r <sub>Rank</sub>	Rang <sub>r</sub>	r <sub>Rank</sub>	Rang <sub>r</sub>	r <sub>Rank</sub>	Rang <sub>r</sub>
SW <sub>log_all</sub>	0,72	16	0,80	10	0,55	15
SW <sub>log_last</sub>	0,86	11	0,79	12	0,79	11
SW <sub>corr_last_A</sub>	0,86	11	0,79	12	0,03	17
SW <sub>corr_last_B</sub>	0,86	11	0,79	12	0,70	14
Eu	<b>0,95</b>	<b>1,5</b>	0,82	8,5	0,87	7
OS <sub>SW(dich)</sub>	0,50	17	0,50	14	0,50	16
OS <sub>SW_separat(dich)</sub>	0,79	15	0,74	17	0,72	13
OS <sub>SW(poly)</sub>	0,87	9	0,77	14	0,81	10
Takte	-	-	-	-	-	-
Opt <sub>glo</sub>	<b>0,95</b>	<b>1,5</b>	<b>0,94</b>	<b>3,5</b>	<b>0,94</b>	<b>1</b>
Opt <sub>manipu</sub>	0,82	13,5	0,70	16	0,76	12
Opt <sub>non_manipu</sub>	0,82	13,5	0,82	8,5	0,82	9
Eingabe <sub>sum</sub>	0,94	5	<b>0,94</b>	<b>3,5</b>	0,93	3
Eingabe <sub>sum_corr_A</sub>	0,94	5	<b>0,94</b>	<b>3,5</b>	0,84	8
Eingabe <sub>sum_corr_B</sub>	0,94	5	<b>0,94</b>	<b>3,5</b>	0,93	3
Eingabe <sub>opt_unrel</sub>	0,94	5	<b>0,94</b>	<b>3,5</b>	0,93	3
Eingabe <sub>opt_rel</sub>	0,88	8	0,88	7	0,88	6
Eingabe <sub>opt_eu</sub>	0,94	5	<b>0,94</b>	<b>3,5</b>	0,92	5

Anmerkungen: Item A mit HE & ME. Item B mit HE, ME & 2x NE. |r<sub>Rank</sub>| = Betrag des Rangkorrelationskoeffizientens nach Spearman. Rang<sub>r</sub> = Rangplatz der Korrelation nach den bivariaten Analysen.

**Tabelle 9: Deskriptive Statistiken für die Steuerwerte separat für die beiden Items**

Indikator	Item A (leicht)				Item B (schwer)			
	M	SD	Min	Max	M	SD	Min	Max
SW <sub>log_all</sub>	-2,90	1,17	-4,00	-0,30	-0,40	1,08	-1,71	1,65
SW <sub>log_last</sub>	-3,25	1,28	-4,00	-0,42	-1,26	1,51	-4,00	1,71
SW <sub>corr_last_A</sub>	4,82	1,90	0,63	5,94	-2,61	3,14	-8,29	3,54
SW <sub>corr_last_B</sub>	-5,35	2,11	-6,58	-0,72	-0,60	0,73	-1,92	0,84
Eu	1,50	3,52	0,00	14,28	31,54	48,88	0,00	192,94
OS <sub>SW(dich)</sub>	0,70	0,47	0,00	1,00	0,02	0,15	0,00	1,00
OS <sub>SW_separat(dich)</sub>	0,83	0,29	0,25	1,00	0,43	0,29	0,00	1,00
OS <sub>SW(poly)</sub>	0,88	0,22	0,25	1,00	0,67	0,21	0,25	1,00
Takte	1,96	0,92	1,00	3,00	2,93	0,33	1,00	3,00
Opt <sub>glo</sub>	0,39	0,16	0,11	0,72	0,42	0,18	0,00	0,94
Opt <sub>manipu</sub>	0,30	0,18	0,00	1,00	0,14	0,29	0,00	1,00
Opt <sub>non_manipu</sub>	0,49	0,29	0,11	1,00	0,70	0,33	0,00	1,00
Eingabe <sub>sum</sub>	-2,96	1,19	-4,00	0,15	-1,25	1,57	-4,00	2,12
Eingabe <sub>sum_corr_A</sub>	-17,19	6,90	-23,21	0,87	-1,32	1,65	-4,21	2,23
Eingabe <sub>sum_corr_B</sub>	-1,37	0,55	-1,86	0,07	-0,59	0,73	-1,87	0,99
Eingabe <sub>opt_unrel</sub>	1,82	3,21	0,00	13,75	51,45	2,50	0,00	632,50
Eingabe <sub>opt_rel</sub>	0,47	0,90	0,00	3,06	28,32	0,23	0,00	451,25
Eingabe <sub>opt_eu</sub>	5,98	10,52	0,00	50,25	157,40	10,00	0,00	1874,90

Anmerkungen: M = Mittelwert; SD = Standardabweichung; Min = Minimum; Max = Maximum.

**Tabelle 10: 2-Faktoren-Lösung für die Kausalanalysen**

	Faktor		kum
	1	2	
Eigenwert	9,71	9,47	19,18
% Varianz	32,36	31,55	63,91
Benennung	Steuerwerte Item A	Steuerwerte Item B	

Anmerkungen: kum = kumuliert über Faktoren.

#### 4.) Zusammenhang zwischen Kausalanalysen und Steuerwerte

Kontrovers diskutiert wird der Zusammenhang zwischen Kausalanalysen und Steuerleistung. Die Korrelationen zwischen den Kausalindikatoren und den Steuerindikatoren schwanken zwischen 0,05 und 0,80.

Für die ausgewählten Steuerwerte  $\text{Eingabe}_{\text{opt\_rel}}$  und  $\text{Eingabe}_{\text{opt\_unrel}}$  liegen die Korrelationen mit dem Kausalmaß  $\text{GdK}_{\text{Glo\_7525}}$  bei -0,36 und -0,39, respektive. Tendenziell geht eine bessere Steuerleistung also mit genaueren Kausalanalysen einher, dessen ungeachtet scheinen unterschiedliche Kompetenzen angesprochen zu werden. Der Zusammenhang fällt deutlicher für das leichte Item aus.

## Literatur

- Beckmann, J. F. (1994). *Lernen und komplexes Problemlösen. Ein Beitrag zur Konstruktvalidierung von Lerntests*. Bonn: Holos.
- Blech, C., & Funke, J. (2005). Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology. Heidelberg University.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. *Sprache & Kognition*, 9, 143-153.
- Funke, J. (1992a). Dealing with dynamic systems: Research strategy, diagnostic approach and experimental results. *German Journal of Psychology*, 16, 24-43.
- Funke, J. (1992b). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Berlin: Springer.
- Funke, J. (1993). 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 Science Publishers.
- Funke, J., Fahnenbruck, G., & Müller, H. (1986). DYNAMIS - Ein Computerprogramm zur Simulation dynamischer Systeme. *Berichte aus dem Psychologischen Institut der Universität Bonn*, 12(3).
- Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology*, 9(Special Issue), S55-S79.
- Greiff, S., & Funke, J. (2008). What makes a problem complex? Factors determining difficulty in dynamic situations and implications for diagnosing complex problem solving competence. In J. Zumbach, N. Schwartz, T. Seufert & L. Kester (Eds.), *Beyond knowledge: the legacy of competence* (pp. 199-200). Wien: Springer.
- Greiff, S., & Funke, J. (in press). Measuring Complex Problem Solving - The MicroDYN approach. In F. Scheuermann (Ed.), *The Transition to Computer-Based Assessment - Lessons learned from large-scale surveys and implications for testing*. Luxembourg: Office for Official Publications of the European Communities.
- Kluge, A. (2004). *Wissenserwerb für das Steuern komplexer Systeme*. Lengerich: Pabst Science Publishers.
- Kluge, A. (2008). Performance assessment with microworlds and their difficulty. *Applied Psychological Measurement*, 32(2), 156-180.
- Leutner, D. (1988). Computersimulierte dynamische Systeme: Wissenserwerb unter verschiedenen Lehrmethoden und Sozialformen des Unterrichts. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 20, 338-355.
- Müller, H. (1993). *Komplexes Problemlösen: Reliabilität und Wissen*. Bonn: Holos.
- Rollett, W. (2008). *Strategieinsatz, erzeugte Information und Informationsnutzung bei der Exploration und Steuerung komplexer dynamischer Systeme*. Münster: LIT.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.