

Kersting, M. & Süß, H. M. (1995). Kontentvalide Wissensdiagnostik und Problemlösen: Zur Entwicklung, testtheoretischen Begründung und empirischen Bewährung eines problemspezifischen Diagnoseverfahrens. *Zeitschrift für Pädagogische Psychologie*, 9, 83-94.

*Zeitschrift für Pädagogische Psychologie / German Journal of Educational Psychology*, 9 (2), 1995, 83-93

MARTIN KERSTING & HEINZ-MARTIN SÜß

## Kontentvalide Wissensdiagnostik und Problemlösen: Zur Entwicklung, testtheoretischen Begründung und empirischen Bewährung eines problemspezifischen Diagnoseverfahrens

Content-valid diagnosis of knowledge and problem solving: Development, test theoretical justification and empirical validation of a new problem-specific test

*Summary:* The present paper deals with the quantitative assessment of knowledge. The article focusses on the development and validation of a system-specific knowledge-test aimed at the problem-solving scenario «tailor shop». The method of content-valid test construction was chosen. The rationality of this choice is based on test-theoretical criteria as well as on the applicability of this method for modern computer-based problem-solving tasks. Above and beyond the test-theoretical explanations and descriptions of the technical aspects of the test construction process, the article focusses on the validation of the test in a study on 198 high school students and (for a part of the tasks) a replication with about half of the original sample.

*Zusammenfassung:* Die vorliegende Arbeit widmet sich der quantitativen Diagnose von Wissen. In dem Beitrag geht es um die Entwicklung und Erprobung eines systemspezifischen Wissenstests zum Problemlösezenario «Schneiderwerkstatt». Für die Testentwicklung wurde die Methode zur Konstruktion *kontentvalider* Meßinstrumente gewählt. Diese Wahl wird einerseits testtheoretisch begründet, andererseits wird auf die günstigen Voraussetzungen zur Anwendung dieser Methode bei modernen computergestützten Problemlöseaufgaben aufmerksam gemacht. Neben der testtheoretischen Begründung und konstruktionstechnischen Erläuterung handelt der Artikel von der Erprobung des Tests in einer ersten Untersuchung mit 198 Oberschülerinnen und -schülern und (für einen Teil der Aufgaben) in einer mit Abstand von einem Jahr durchgeführten Wiederholungsuntersuchung mit gut der Hälfte derselben Stichprobe.

Sonderdruck aus

Zeitschrift für

# Pädagogische Psychologie

Verlag Hans Huber Bern Stuttgart Toronto

### 1. Zur testtheoretischen Begründbarkeit der Diagnostik systemspezifischen Wissens

Zu den aktuellen Themen der Kognitionsforschung gehört die Frage nach der Bedeutung von Wissen für das Lösen komplexer Probleme. Dabei sind die Ergebnisse zum Zusammenhang von Wissen und Steuerungsleistungen inkonsistent: von einem positiven Zusammenhang berichten z. B. Funke (1985), Putz-Osterloh (1987), Reichert & Dörner (1988), Süß, Kersting & Oberauer (1993), während Morris & Rouse (1985), Putz-Osterloh, Bott & Houben (1988), Leutner (1989) sowie Strohschneider (1991b) keinen kohärenten Zusammenhang auffinden konnten. Bei dieser uneinheitlichen Befundlage liegt es nahe, die Güte der wissensdiagnostischen Instrumente zu hinterfragen. Diesem Aspekt wird in der Problemlöseforschung und der Wissenspsychologie bislang nur wenig Beachtung geschenkt, die traditionsreichen Arbeiten zur Wissensdiagnostik im Bereich der Pädagogischen Psychologie werden nicht ausreichend rezipiert. Fortschritte zeigen sich eher auf der konzeptuellen (z. B. Kluge 1988; Tergan 1989a, b) denn auf der Ebene praktischer Diagnostik. Die empirischen Arbeiten von Funke (1985), Broadbent, Fitzgerald & Broadbent (1986), Berry & Broadbent (1987; 1988), Putz-Osterloh et al. (1988), Hörmann & Thomas (1989) sowie Andresen & Schmidt (1993) erlauben beispielsweise keinen ausreichenden Einblick in die Herleitung und in die Testgüte der ein-

gesetzten wissensdiagnostischen Instrumente; die Stringenz der Ergebnisinterpretation ist dadurch beeinträchtigt. In der Wissenspsychologie wird damit ein – nach Gigerenzer (1981) – weitverbreitetes Mißverhältnis in der Vorgehensweise psychologischer Forschung wiederholt, indem man das Meßproblem vernachlässigt und sich stattdessen um so intensiver mit der Analyse und Interpretation der möglicherweise ungültigen Zahlen beschäftigt

Um Aufschluß über die Gültigkeit von Messungen zu erlangen, stehen in der Psychometrie verschiedene (einander ergänzende) Validierungskonzepte zur Verfügung (siehe Jäger 1986). Cronbach & Meehl (1955) unterscheiden drei Validitätsarten: Konstruktvalidität, Kriteriumsvalidität und Kontentvalidität. Eine *Konstruktvalidierung* mit dem – von Jäger (1986, 203) formulierten – Ziel, «die Beziehungen zwischen Testverhalten und theoretischen Begriffen (den Konstrukten)» zu klären, bringt im Bereich der Wissenspsychologie zur Zeit erhebliche Schwierigkeiten mit sich, da die theoretischen Vorstellungen über den Aufbau und die Funktion von Wissen beim Problemlösen weder hin-

reichend präzise noch konsensfähig entwickelt sind. Die Möglichkeit einer *Kriteriumsvalidierung* hängt von der Gewinnung geeigneter Kriterien ab. Dabei gilt es, die Notwendigkeit der Trennung von Methodenentwicklung und Hypothesenprüfung zu beachten. Eine Testvalidierung an einem Kriterium kommt nur dann in Betracht, wenn der zu erwartende Zusammenhang bereits vorab geklärt ist. Anderenfalls ist es beispielsweise problematisch, einen Wissenstest an den Steuerungsleistungen bei einem komplexen System zu validieren und zugleich den Zusammenhang von Wissen und denselben Steuerungsleistungen zu untersuchen. In diesem Fall kann die Korrelation zu den Steuerungsleistungen nicht oder zumindest nicht als einziges Gütekriterium für den Wissenstest herangezogen werden. Die Validität des Wissenstests muß auch anderweitig geprüft und gesichert werden. Solange aber keine weiteren relevanten und von der Hypothesenprüfung unabhängigen Kriterien vorhanden sind, fällt das Konzept der Kriteriumsvalidierung insgesamt aus. Somit bleibt für die notwendige Validierung zunächst nur das Konzept der *Kontentvalidierung* übrig. Bei der Kontentvalidität geht es um das Verhältnis zweier Mengen zueinander: «In general, content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content» (APA 1985, 10). Kontentvalidität muß bereits bei der Aufgabenkonstruktion angestrebt werden. Voraussetzung für eine kontentvalide Testkonstruktion ist die Definierbarkeit von Itemuniversa. Diese «Definierbarkeit» ist oft angezweifelt worden (Loevinger 1965; Moser 1987; siehe aber Cronbach 1971, 454–455). Tatsächlich ist es dieser Punkt, der die Verbreitung kontentvalider Testverfahren über die oftmals besonders übersichtlichen klassischen Anwendungsbereiche der Pädagogischen Psychologie (z. B. «Grundrechnen») hinaus verhindert hat (siehe auch den Begriff «triviale Validität» bei Lienert 1961, 260). Für computersimulierte Probleme scheint dieses Verfahren jedoch indiziert, da die Grundgesamtheit des Testgegenstandes mit den Simulationsalgorithmen, die diesen Problemen zugrundeliegen, definiert werden kann. Schwierigkeiten ergeben sich insbesondere bei der Transformation der Algorithmen in andere

Darstellungsformen und bei der Definition des Universums möglicher Steuerungseingriffe (siehe unten). Der Anwendung von Testkonstruktionstechniken auf der Basis von Kontentanalysen der Simulationsalgorithmen gebührt dennoch – auch vor dem Hintergrund der zunehmenden Verbreitung tutorieller Systeme – besondere Aufmerksamkeit.

Im folgenden wird – am Beispiel des Problemlösezenarios «Schneiderwerkstatt» – über die Entwicklung und Erprobung eines kontentvaliden Tests zur *quantitativen* Diagnose der Menge korrekten Wissens im Kontext der Identifikation von Systemmerkmalen berichtet. Das Vorgehen der Testkonstruktion orientiert sich an den für den deutschsprachigen Raum grundlegenden Arbeiten von Klauer (1983; 1984a, b). Die Annahme der Kontentvalidität wird durch die empirische Prüfung von aus der Aufgabenanalyse abgeleiteten Hypothesen zusätzlich untermauert<sup>1</sup>.

## 2. Konstruktion eines kontentvaliden Wissenstests am Beispiel der «Schneiderwerkstatt»

### 2.1. Aufgabenanalyse

Zentraler Schritt der Testkonstruktion ist eine Aufgabenanalyse im Sinne von Resnick und Ford (1981; siehe auch Schott 1984, 48). Ziel der Aufgabenanalyse ist die Darstellung des zugrundeliegenden Sachverhalts durch eine geeignete Aussagenform. Die resultierende Grundmenge von Aussagen soll den interessierenden Sachverhalt vollständig umfassen. Dies kann entweder durch die Aufzählung aller Elemente der Grundmenge oder durch die Angabe mengenbildender Merkmale erreicht werden. Die erste Darstellung eines computersimulierten Problemszenarios liefert der zugrundeliegenden Simulationsalgorithmus (für die «Schneiderwerkstatt» siehe Funke 1983). Eine vereinfachende Darstellung des Sachverhalts und somit einen weiteren aufgabenanalytischen Schritt leistet eine Vernetzungsgraphik des Systems (für die «Schneiderwerkstatt» siehe z. B. Putz-Osterloh & Lüer 1981, 314). Beide Darstellungsformen wurden hier als Mittel der Aufgabenanalyse auf die «Schneiderwerkstatt» angewandt (siehe Kersting 1991). Die Vernetzungsgraphik der «Schneiderwerkstatt» ist in Abbildung 1 wiedergegeben, da sie zur Veranschaulichung nachstehender Passagen beitragen kann.

Des weiteren können Taxonomien von Systemmerkmalen, wie sie zum Beispiel von Hübner (1989) und Funke (1990) vorgeschlagen wurden, zur Aufgabenanalyse herangezogen werden. Davon handelt der nächste Abschnitt. Eine Diskussion des Für und Wider solcher Taxonomien findet sich bei Strohschneider (1991a) und Funke (1991).

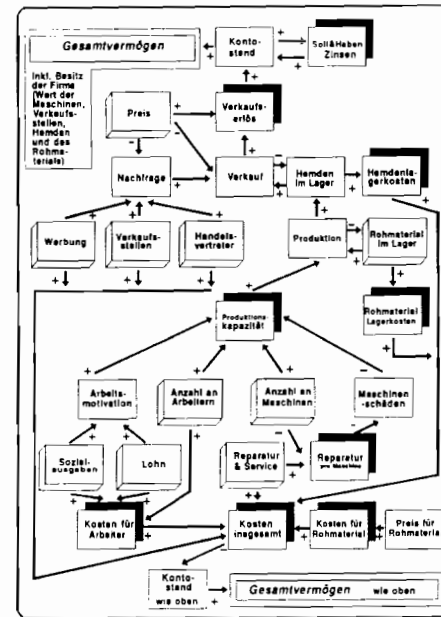


Abbildung 1: Vernetzungsgraphik der Schneiderwerkstatt: dreidimensionale Rahmen: Eingriffsvariablen, einfache Rahmen: Nicht direkt beeinflussbare Variablen. Rahmen mit Schatten: Verdeckte Variablen

### 2.1.1. Systemmerkmale

Sofern nicht alle Elemente des interessierenden Sachverhalts im Test vorgegeben werden können, muß eine *repräsentative* Itemstichprobe gezogen werden. Zur Wahrung der Repräsentativität empfiehlt es sich dabei, aus der Grundmenge von Aussagen Teilmengen zu bilden und dann aus diesen Teilmengen Items zu ziehen (quotierte Stichprobe, s. Cronbach 1984). Um im System «Schneiderwerkstatt» unterscheidbare Teilsachbestände zu entdecken, wurde zunächst die von Funke (1990) entwickelte Systematik von Systemmerkmalen zur Aufgabenanalyse herangezogen und daraus eine speziell für die «Schneiderwerkstatt» ausdifferenzierte Taxonomie entwickelt.

Aus der Taxonomie von Funke wurde die Unterscheidung zwischen der *Menge der Variablen* und der *Menge der Verbindungen zwischen Variablen* (bei Funke «*Wirkungen und Abhängigkeiten*» genannt) übernommen. Jede dieser beiden Mengen wurde dann ergänzt, indem zusätzlich zu den Variablen und Variablenrelationen, die im System implementiert sind, eine Menge mit Elementen festgelegt wurde, die

im System nicht vorkommt. Elemente dieser Mengen waren etwa die «erfundene» Variable «Energiesteuer» oder die «erfundene» Variablenrelation zwischen den Variablen «Werbung» und «Hemdenpreis»; beides Elemente, die im System «Schneiderwerkstatt» nicht vorkommen. Methodisch orientierte sich die Gewinnung zusätzlicher Elemente an dem bei Feger (1984) beschriebenen Verfahren. Die Erzeugung solcher Teilmengen ist in mehrfacher Hinsicht nützlich und sinnvoll: Zunächst ist die explizite Festlegung Garant der Beibehaltung einer rationalen Itemschreibung (Roid & Haladyna 1982). Außerdem wird durch die Berücksichtigung im System nicht implementierter Elemente vermieden, daß die Bearbeitung des Wissenstests zu viele Kenntnisse vermittelt. Weiterhin werden für die jeweils zutreffenden Antwortalternativen Distraktoren gewonnen, die einen ersten Zugang zur Diagnose von «Fehlwissen» vermitteln (siehe Spada & Reimann 1988). Diese Aspekte können hier aus Platzgründen nicht weiter verfolgt werden. Die folgende Darstellung bezieht sich nur auf die Menge der im System tatsächlich implementierten Variablen und Variablenrelationen.

(a) Die *Menge der Relationen* zwischen den Variablen wurde unterteilt in die Mengen der «*direkten*» und «*indirekten*» Verbindungen. Diese Unterscheidung läßt sich anhand der Vernetzungsgraphik (Abbildung 1) erläutern. Während z. B. die Verbindung der Variablen «Lohn» zu «Arbeitsmotivation» eine «*direkte*» Verbindung darstellt, ist die Relation «Lohn»-»Produktionskapazität» eine «*indirekte*» Verbindung, weil sie über die Variable «Arbeitsmotivation» vermittelt wird.

Die Teilmenge der «*direkten*» Verbindungen wurde in die drei Mengen «*Abhängigkeit*», «*Wirkung*» und «*Wechselwirkung*» unterteilt. Der Unterschied zwischen einer «*Abhängigkeits-*» und einer «*Wirkungsrelation*» ergibt sich erst durch unterschiedliche Perspektiven der Systemsteuerungen auf ein und dieselbe Variablenrelation. Es handelt sich nicht um disjunkte Mengen, sondern um Klassifikationsgesichtspunkte. Innerhalb dieser drei Mengen existieren jeweils die beiden Teilmengen «*positive*» und «*negative*» Verbindungen, sowie bei den «*Wechselwirkungen*» (in Abbildung 1 als Doppelpfeil dargestellt) entsprechende Kombinationen.

(b) Bei der *Menge der Variablen* ergab die Aufgabenanalyse fünf Teilmengen, wobei jede Variable Element mehrerer Mengen ist: (1) die *Menge der von Zufallsschwankungen abhängigen Variablen*, (2) die *Menge der von früheren eigenen Ausprägungen abhängigen Variablen* («*Feedback*», «*Eigendynamik*»), (3) die *Menge der auf andere Variablen wirkenden Variablen*, (4) die *Menge der von anderen Variablen abhängigen Variablen* und (5) die *Menge der vernetzten Variablen*. Bei der Menge der vernetzten Variablen wurden die Abhängigkeit und die Wirkungen einer Variablen zu einem Merkmal zusammengefaßt. Die Mengen drei bis fünf wurden aufgrund der *niedrigen, mittleren und hohen Merkmalsausprägung* in jeweils drei weitere Teilmengen untergliedert.

### 2.1.2. Anforderungsmerkmale

Ein System wird erst dann zu einer Aufgabe oder einem Problem, wenn an die Probanden An-

forderungen gestellt werden. Die Aufgabenanalyse muß daher neben Systemmerkmalen auch die erforderlichen psychologischen Verhaltensaspekte umfassen. Die Grundmenge der Aussagen über das System «Schneiderwerkstatt» ergibt sich erst durch eine Kombination der Systemmerkmale mit den Anforderungsmerkmalen.

Funke (1990) unterscheidet zwischen den Anforderungen «Identifikation» und «steuernde Kontrolle». Diese Unterscheidung korrespondiert auf der Wissensseite mit der psychologischen Unterscheidung von Sach- und Handlungswissen (siehe z. B. Kluwe, Misiak & Haider 1991; Strohschneider 1991b). Die hier vorgestellte Aufgabenanalyse orientiert sich an dieser bei Funke (1990) ausführlich erläuterten Einteilung. Bei den Identifikationsanforderungen wurde die von Funke präsentierte Dreiteilung (siehe auch Plötzner, Spada, Stumpf & Opwis 1990; Plötzner & Spada, 1992) zu einem vierstufigen Anforderungskontinuum ausdifferenziert:

- (a) Das bloße Erkennen (= *semi-qualitatives Wissen*): Das Erkennen des «Vorhandenseins» von Variablen oder Relationen zwischen Variablen.
- (b) Richtungswissen (= *qualitatives Wissen*): Das Erkennen der Richtung einer Relation oder der Position einer Variablen in bezug zu einer anderen Variablen.
- (c) Vorzeichenwissen (= *semi-quantitatives Wissen*): Das Erkennen von relativen Stärken von Variablenverbindungen (je mehr von A, desto mehr von B, also Plus- bzw. Minusvorzeichen) bzw. relativen Häufigkeiten (viel/wenig) von Abhängigkeiten, Wirkungen und Vernetzungen einer Variablen.
- (d) Numerisches Wissen (= *quantitatives Wissen*): Das Wissen exakter Gewichtungsfaktoren von Variablenrelationen und exakter Häufigkeiten von Abhängigkeiten, Wirkungen und Vernetzungen einer Variablen.

Bei den Aussagen zur Steuerung von Systemmerkmalen (Handlungswissen) stößt die Konstruktionsidee kontextvalider Tests an ihre Grenzen. Die Möglichkeiten konkreter Eingriffe bzw. die Möglichkeiten zur Formulierung von Steuerungsregeln sind derart mannigfaltig, daß ohne weiteres kein befriedigendes Abbildungsverhältnis erzielt werden kann. Auf die kontextvalide Erfassung von Wissen über die Steuerung von Systemmerkmalen wurde daher verzichtet.

## 2.2. Itemuniversum und Itemstichprobe

Zur Bildung der Aufgabenmenge wurden die Aussagen durch Transformationsregeln mit einer Itemform verknüpft. Alle Aufgaben, welche durch die Transformationsregeln gebildet werden können, bilden gemeinsam die «Grundmenge von Aufgaben», das «Itemuniversum» des zu konstruierenden Tests.

Bezüglich der Aussagen zu den Identifikationsanforderungen wurden für alle Teilmengen des Itemuniversums Samplingvorschriften zur Wahrung der Repräsentativität entwickelt. Konkret wurden diese Itemstichproben jeweils durch eine geordnete Zufallsauswahl gewonnen. Entsprechend fragen die Testitems auf unterschiedlichem «Präzisionsniveau» nach Variablen und Variablenrelationen der «Schneiderwerkstatt». Die Größe der Itemstichproben (und somit die Anzahl der Items pro Skala) variiert in Abhängigkeit von der Größe der Teilmenge des Itemuniversums (so zeigt die Vernetzungsgraphik in Abbildung 1 z. B. 40 direkte Variablenrelationen, aber «nur» 28 Variablen), in Abhängigkeit von spezifischen Repräsentationsvorgaben (z. B. die quotierte Berücksichtigung bestimmter «Bereiche», etwa dem «Bereich» «Produktion») und in Abhängigkeit vom gewählten Antwortformat. Die Samplingvorschriften – ebenso wie die Transformationsregeln und weitere Einzelheiten zur Testkonstruktion sind bei Kersting (1991) dokumentiert.

Ausgespart blieben Fragen auf dem «quantitativen oder numerischen» Niveau der Identifikation. Die Identifikation exakter Gewichtungsfaktoren z. B. von Variablenrelationen ist eine Anforderung, die Problemlöser(innen) nach wenigen Steuerungsdurchgängen der «Schneiderwerkstatt» kaum bewältigen dürften. Tabelle 1 liefert eine Kurzcharakteristik der konstruierten Aufgabentypen. In der Tabelle ist außerdem verzeichnet, ob ein Subtest bereits in der Erstuntersuchung (EU) eingesetzt wurde oder erst für die Wiederholungsuntersuchung (WU) entwickelt wurde. (Zur Untersuchungsabfolge siehe unten Punkt 3.1.) Itembeispiele zu den Aufgabentypen finden sich im Anhang.

## 3. Kontrolle und Stützung der Kontextvaliditätsannahme

Zur Prüfung der Kontextvaliditätsannahme bei regelgeleitet konstruierten Tests werden in der Literatur im wesentlichen vier Strategien vorgeschlagen:

- (a) die Kontrolle der Herstellungsprozedur,
- (b) die Entwicklung kontextvalider Paralleltests,

Tabelle 1: Übersicht über die Aufgabentypen (EU = Erstuntersuchung, WU = Wiederholungsuntersuchung)

Aufgaben zur Identifikation von Systemmerkmalen	
<b>Aufgabentyp Ia:</b> "Abhängigkeit" EU, 11 Items	Vorgegeben sind je vier Aussagen über eine Variable. Zuerst soll geprüft werden, ob diese Variable im System implementiert ist oder dort nicht vorkommt. Falls die Variable in der "Schneiderwerkstatt" vorkommt, soll außerdem angekreuzt werden, ob sie von <u>vielen</u> , <u>mehreren</u> oder <u>wenigen</u> anderen Variablen abhängig ist.
<b>Aufgabentyp Ib:</b> "Wirkung" EU, 11 Items	Wie Aufgabentyp Ib, gefragt wird aber nicht nach der Abhängigkeit, sondern ob die jeweilige Variable auf <u>vielen</u> , <u>einige</u> oder <u>wenige</u> andere Variablen einen Einfluß ausübt.
<b>Aufgabentyp IIa:</b> "Zusammenhänge verbal" EU und WU, 20 Items (15 gemeinsame Items in EU/WU)	Vorgegeben sind je sechs Aussagen über alle Möglichkeiten der Ausgestaltung einer direkten Relation zwischen zwei Variablen. Die richtige Aussage soll angekreuzt werden.
<b>Aufgabentyp IIb:</b> "Zusammenhänge figural", Typ 1: "Rekognition" EU, 8 Items	Vorgegeben sind jeweils drei graphische Darstellungen indirekter Relationen. Jede indirekte Relation enthält mindestens zwei direkte Relationen. Die Darstellung, welche die indirekte Relation (oder alle direkten Relationen) richtig wiedergibt, ist anzukreuzen.
<b>Aufgabentyp IIc:</b> "Zusammenhänge figural", Typ 2: "Reproduktion" WU, 7 Items	Vorgegeben sind je vier Variablen. Aufgabe ist es, die Relationen als "Pfeile" selbst einzuzichnen.
<b>Aufgabentyp III a:</b> "Eigenschaften" EU und WU 6 Items (EU), davon 5 in der WU	Vorgegeben sind je sechs "qualitative" Aussagen, die eine ausgewählte Variable und ihre Relationen betreffen. Jede zutreffende Aussage ist anzukreuzen.
<b>Aufgabentyp III b:</b> "Wichtigkeit" WU 21 Items	Für sieben ausgewählte Variablen sind alle möglichen Paarvergleiche angeführt. Bei jedem Paarvergleich ist diejenige Variable anzukreuzen, die für "wichtiger" erachtet wird.

- (c) die Prüfung von Vorhersagen zur Datenstruktur und
- (d) der Prä-Postvergleich.

Klauer (1984a, 16) beklagt, daß die meisten Autoren «bislang nahezu ausschließlich an die Überprüfung der Herstellungsprozedur denken» und weitere Prüfverfahren vernachlässigen. An dieser Stelle wird deshalb auf den Bericht der eher technischen Herstellungskontrolle verzichtet, gleichwohl die hier dargestellte regelgeleitete, testtheoretisch begründete und durch Experten bezüglich der Herstellung kontrollierte Konstruktionsprozedur erheblich zur Sicherung des Anspruchs auf Kontextvalidität beiträgt. Vielversprechend ist das Konzept der kontextvaliden Paralleltests (vgl. z. B. Schott, Neeb & Wieberg 1981), besonders für Wiederholungsmessungen etwa zur Untersuchung der Frage nach der Stabilität von Wissen. Der hier vorgestellte erste Konstruktionsversuch für einen Wissenstest im Rahmen komplexer Problemlösenszenarios blieb allerdings auf die Erstellung einer Testversion beschränkt. Zur Stützung der

Kontextvaliditätsannahme werden im folgenden Berichte über eine Prüfung der Vorhersage der Datenstruktur und ein Prä-Postvergleich angeführt.

### 3.1. Untersuchungsabfolge und weitere Instrumente

Als Bewährungsprobe des kontextvaliden konstruierten systemspezifischen Wissenstests «WIS» galten die Ergebnisse von 198 Schülerinnen und -schülern (Durchschnittsalter: 17.6 Jahre) aus der zweitägigen Erstuntersuchung (EU). – Am ersten Tag wurden die intellektuellen Fähigkeiten mit einem Test für das «Berliner Intelligenzstrukturmodell» (Jäger 1982) erfaßt; bereichsspezifisches Vorwissen von hohem Allgemeingrad wurde mit einem Wirtschaftslehrestest erhoben. Am zweiten Tag wurden sowohl der zur Erprobung anstehende Wissenstest «WIS-1» als auch der diesem Test zugrundeliegende Sachverhalt «Schneiderwerkstatt» mehrfach bearbeitet.

Der Untersuchungsplan sah zwei Experimental-Gruppen und eine Kontrollgruppe vor. Die Unterscheidung zwischen den Gruppen ist an anderer Stelle ausführlich dokumentiert (SuB, Oberauer & Kersting 1993) und kann für die vorliegende Frage mit wenigen Ausnahmen unberücksichtigt bleiben. In der EU wurde der Wissenstest (*«WIS-1»*) in den beiden Experimentalgruppen einmal vor (Meßzeitpunkt «M1») und einmal nach der ersten Problembearbeitung dargeboten (Meßzeitpunkt «M2»). Am Ende der Untersuchung (Meßzeitpunkt «M3») bekamen diese Gruppen eine im Umfang reduzierte Testform. Die Kontrollgruppe (Gruppe 3, N=44) bearbeitete den kompletten Wissenstest erst im Anschluß an die erste Systembearbeitung sowie ein weiteres Mal am Ende der Untersuchung, so daß für den ersten Meßzeitpunkt nur von den 154 Personen der Experimentalgruppen Daten zur Verfügung stehen. Vierundsechzig Prozent der Teilnehmer(innen) der Erstuntersuchung (N=137) nahmen ein Jahr später an der Wiederholungsuntersuchung (WU) teil. Für 24 von ihnen liegen allerdings keine Wissensdaten aus der Erstuntersuchung vor, da sie dort eine Vorform des Wissenstests bearbeitet hatten. Die teilweise überarbeitete und ergänzte Version des problemspezifischen Wissenstests (*«WIS-2»*) wurde ebenso wie in der EU einmal vor und einmal nach der ersten Problembearbeitung dargeboten.

### 3.2. Prüfung der erwarteten Schwierigkeitsstruktur

Einen Zugang zur Kontrolle der Kontenvalidität sieht Klauer (1984b, 69) darin, zu überprüfen, *«inwieweit bestimmte Aspekte der Testdaten alleine schon aus der Definition und Analyse der Grundgesamtheit der Aufgaben vorhergesagt werden können.»* Als Prüfstein der Kontenvalidität dient hier diejenige Struktur, welche durch die Systemtaxonomie in die Aufgabenmenge eingeführt wurde. Das Vertrauen in den Test und die Anwendung kann dann dadurch gestärkt werden, daß sich die vorgegebene Struktur der Fragen auch in den Antworten wiederfindet. Nach Analyse der Anforderungen wurde angenommen, daß es bei der Identifikationsleistung der Problemlösenden unterschiedliche «Präzisionsgrade» gibt. Diese Präzisionsgrade entsprechen in ihrem hierarchischen Aufbau den Ebenen der Systemmerkmale. Die auf höchster Präzisionsstufe geleistete Identifikation (das «quantitative Wissen») schließt der Aufgabenanalyse zufolge alle minderpräzisen Identifikationsleistungen mit ein.

Der Präzisionsgrad der Identifikationsleistungen einerseits, und die erfragten Systemmerkmale andererseits können – durch unterschiedliche Perspektiven der Fragestellungen und durch die Wahl der Iteminhalte – unabhängig voneinander variiert werden. Dies erlaubt die Ableitung von Vorher-

sagen z. B. über den hierarchischen Aufbau der Präzision einer Identifikationsleistung. Ein Beispiel für eine solche Vorhersage – und zugleich ein Beispiel für die unabhängige Variation von System- und Anforderungsmerkmalen – kann anhand der beliebig gewählten Variablenrelation *a-b* gegeben werden. Hält man das Systemmerkmal, eben diese Variablenrelation *a-b*, konstant und variiert nun systematisch die in bezug auf die Identifikation dieses Systemmerkmals geforderte Anforderung, so kann man folgende Prognose aufstellen: Weiß eine Person, *«daß die Erhöhung der Variable a um den Betrag x zu einer um y erhöhten Ausprägung der Variablen b führt»* (quantitativ), so sollte diese Person auch wissen, daß gilt: *«je mehr a desto mehr b»* (semi-quantitativ), daß *«a auf b wirkt»* (qualitativ), und daß *«eine – wie auch immer ausgerichtete – Verbindung a-b im System implementiert ist»* (semi-qualitativ). Aus dieser Einteilung ergab sich die allgemeine Prognose, daß Fragen, die «nur» «semi-qualitatives» Wissen erfordern, leichter zu lösen sind als Aufgaben, die «semi-quantitatives» Wissen erfordern. Das «qualitative Wissen» sollte sich in der Mitte dieses postulierten Itemschwierigkeitskontinuums wiederfinden lassen.

Um überprüfen zu können, ob die empirische Datenstruktur diesen Annahmen entspricht, wurden die entsprechenden Itemschwierigkeits-Mittelwerte bestimmt. Geprüft wurde die Schwierigkeitsstruktur zum zweiten Meßzeitpunkt der Erstuntersuchung, da hier die umfangreichste Stichprobengröße (N=198) vorlag. Während sich für die Items zur semi-qualitativen Identifikation von Variablen und Variablenrelationen der «Schneiderwerkstatt» ein über alle Proband(innen) gemittelter Prozentsatz richtiger Antworten von 77,3% ergab, lag der entsprechende Wert für die qualitative Identifikation bei 46,7%. Die Identifikationsleistung, die der Anforderungsanalyse zufolge am anspruchsvollsten sein sollte, nämlich die Identifikation auf semi-quantitativem Niveau, erwies sich mit einem durchschnittlichen Wert von 29,0% auch tatsächlich als am schwierigsten.

Um diese Übereinstimmung der Datenstruktur mit der formulierten Erwartung zufallskritisch abzusichern, wurde eine Varianzanalyse mit Meßwiederholung durchgeführt. Grundlage der Varianzanalyse bildeten die relativen Häufigkeiten der von den Proband(innen) gelösten Items für jede dieser drei Wissensskalen («semi-quantitativ», «qualitativ» und «semi-qualitativ»). Der F-Test<sup>2</sup> erwies sich als hochsignifikant (F [2, 394]=464,3, p < .001). Auch die Kontraste waren hochbedeutsam, so daß die gesamte Schwierigkeitshierarchie als bestätigt angesehen werden kann. Die aufgrund der Definition und Analyse der Grundgesamtheit vorhergesag-

te Ähnlichkeitsstruktur konnte auch mit Hilfe der «smallest space analysis» dargestellt werden (siehe Kersting 1991, 133–141).

### 3.3. Prä-Postvergleich

Bei der Strategie der Prä-Postvergleiche werden Differenzwerte immer dann als Gütekriterien für nach den Regeln der Kontenvalidität konstruierte Tests herangezogen, wenn zwischen Vor- und Nachtest eine kontentrelevante Intervention geschaltet wird. Von dem hier vorgestellten systemspezifischen Wissenstest kann erwartet werden, daß er zwischen einer Vortestleistung ohne «Schneiderwerkstatt»-Erfahrung und einer Nachtestleistung mit Systemerfahrung differenzieren kann. Gegen eine Anwendung dieses Verfahrens kann aus methodenkritischer Sicht eingewandt werden, daß in diese Kontrolle empirisch ungeprüfte Annahmen mitgehen. In der skizzierten Argumentation «versteckt» sich die Hypothese, daß beim Umgang mit dem System in jedem Fall verbalisierbares Systemwissen erworben wird. Diese Hypothese ist umstritten (vgl. die Diskussion um die Bedeutung impliziten Wissens beim Problemlösen; z. B. Haider 1992). Ergäbe die skizzierte Prüfungsstrategie keine Prä-Postdifferenzen, könnte der Anspruch auf Kontenvalidität des Tests aufrechterhalten werden, indem diese Wissenserwerbshypothese in Frage gestellt wird. Zeigen sich aber Prä-Postdifferenzen, kann dieses Ergebnis durchaus als Hinweis auf die Kontenvalidität des Tests gelten. Die zu prüfende Annahme beim Prä-Postvergleich lautet: Wenn erstens die Wissenserwerbshypothese zutrifft und zweitens die Itemstichprobe das Item-

universum, also die in Frageform gebrachten Aussagen zur Identifikation von Systemmerkmalen repräsentiert, muß sich, so Herhig (1976, 142), *«aus der Variablen >Testzeitpunkt (vor/nach) <das Lösungsverhalten (falsch/richtig) vorhersagen lassen»*. Cox & Vargas (1966, zitiert nach Büscher 1975) sehen in einem vergleichbaren Verfahren eine alternative Technik zur Bestimmung der lehrzielorientierten Trennschärfe.

Tabelle 2 zeigt die Ergebnisse der Varianzanalyse mit Meßwiederholung für die einzelnen «WIS-1»-Skalen (Erstuntersuchung) für die beiden zusammengefaßten Experimentalgruppen. Diese beiden Gruppen haben den Test einmal vor und einmal nach der Steuerung der «Schneiderwerkstatt» bearbeitet. Infolge der Systembearbeitung war für alle Skalen ein signifikanter Wissenszuwachs zu verzeichnen. Insgesamt sind diese Ergebnisse im Sinne der Prä-Posttest-Prüfstrategie ein weiterer Beleg für die Kontenvalidität der regelgeleiteten Testkonstruktion. Klauer (1983, 701) bemängelt an diesem Prüfungsansatz, daß den bis dato vorgenommenen Experimenten dieser Art eine Kontrollgruppe mit bloßer Testwiederholung fehlte. Skeptiker könnten einwenden, daß der Wissenszuwachs zum zweiten Wissensmeßzeitpunkt durch die wiederholte Bearbeitung der Wissensfragen bedingt ist. Um diesem Einwand zu hegegnen, sah der experimentelle Untersuchungsplan eine Kontrollgruppe vor, die den Wissenstest erstmalig nach der Systemsteuerung bearbeitete. Entsprechend der skeptischen Hypothese müßte diese wissenstestunerfahrene Gruppe schlechter abschneiden als die beiden Gruppen mit vorheriger Wissenstesterfahrung. Das Ergebnis stimmte

Tabelle 2: Varianzanalysen über den Wissenszuwachs: Erstuntersuchung, Gruppen I und II (N=154)

Skala zum Aufgabentyp	Mittelwert				
	M1	M2	df	F	p
I a "Fragen zur Abhängigkeit von Variablen"	69.39	73.43	1, 152	18.93	.00
I b "Fragen zur Wirkung von Variablen"	70.10	71.79	1, 151	4.55	.04
II a "Fragen zum Zusammenhang zwischen je zwei Variablen (verbal)"	47.13	56.06	1, 151	75.74	.00
II b "Fragen zum Zusammenhang mehrerer Variablen (figural)", Typ I	60.19	64.04	1, 150	7.63	.01
III "Fragen zu verschiedenen Eigenschaften einer Variablen und deren Relationen"	71.16	74.27	1, 152	26.29	.00

<sup>2</sup> Mittlere Prozentwerte der Anzahl richtiger Lösungen; "M1", "M2" = Meßzeitpunkte

hingegen mit der vorher vertretenen Hypothese überein: Die Erfahrungen mit der «Schneiderwerkstatt» schlugen sich unmittelbar in dem kontentvaliden Test nieder: Nach der Bearbeitung gab es keinen signifikanten Unterschied in den Wissensleistungen der Gruppen mit und ohne vorgeschalteten Wissenstest (für die Gesamtskala:  $t = -0.04$ , n. s.). Damit stützen auch die Ergebnisse aus dem Prä-Postvergleich die Annahme der Kontentvalidität des Tests.

#### 4. Zur Reliabilität der Wissensskalen

Die skizzierte Testentwicklung und die bislang berichteten Ergebnisse haben zwar in ausreichendem Maße Belege für die Kontentvalidität erbracht, ungeklärt ist jedoch die Frage der Reliabilität der Messungen. Die Bestimmung geeigneter Reliabilitätsmaße ist durchaus problematisch, da Paralleltestmessungen – bei zeitgleicher Messung sicherlich das adäquateste Reliabilitätsmaß – nicht vorliegen, und Koeffizienten für die innere Konsistenz der Skalen kaum als Reliabilitätsmaße geeignet sind. Auch die Stabilitätskoeffizienten innerhalb der Untersuchungen eignen sich nicht, da zwischen den Messungen bei der Systemsteuerung Gelegenheit zum Wissenserwerb bestand. Noch am ehesten geeignet sind die Stabilitätskoeffizienten über ein Jahr, berechnet als Korrelationen zwischen den beiden Skalen, die sowohl in der Erst- als auch in der Wiederholungsuntersuchung eingesetzt wurden (IIa und IIIa, siehe Tabelle 1). Die Korrelation zwischen der letzten Wissens-erhebung der Erstuntersuchung und der ersten Erhebung der Wiederholungsuntersuchung im darauffolgenden Jahr betrug für die aggregierten Leistungen  $r = 0.70$  ( $p < .01$ ,  $N = 113$ ). Da der Wissenstest signifikante Korrelationen zu einigen Facetten der Intelligenz aufweist (siehe Süß, Kersting & Oberauer 1993), wurde mit der Abspaltung einer über ein Jahr hinweg äußerst stabilen Intelligenzkomponente «Verarbeitungskapazität» (Stabilitätskoeffizient:  $r = 0.90$ , siehe Süß, Kersting & Oberauer 1991) eine statistische Korrektur vorgenommen. Die «bereinigte» Korrelation ist mit  $r = 0.62$  immer noch beachtlich hoch.

#### 5. Rückblick und Ausblick

Die vorgestellte regelgeleitete Testkonstruktion stellt den erfolgreichen Versuch dar, das vor allem in der Pädagogischen Psychologie entwickelte und etablierte Konzept der kontentvaliden Testkonstruktion für den Bereich der Problemlöseforschung und der Wissenspsychologie zu nutzen. Diese Form der Validierung ist auf dem Gebiet der Problemlöseforschung notwendig, solange die Voraussetzungen für die erfolgreiche Anwendung anderer Validierungsstrategien nicht gegeben sind. Bezüglich einer Kriteriumsvalidierung an den Steuerungsleistungen oder bezüglich einer Konstruktvalidierung gehört hierzu z. B. eine konsensfähige Konzeptualisierung von Wissen und die Klärung der Frage, wie problemspezifisches Wissen mit Steuerungsleistungen assoziiert oder dissoziiert ist. Der Ansatz der Kontentvalidität gewährleistet hier sozusagen «zwischenzeitlich» eine von diesen Fragestellungen unabhängige Methodenentwicklung und schafft Grundlagen für die Testbeurteilung. Sobald in diesen Fragestellungen Fortschritte erzielt werden, muß selbstverständlich auch für kontentvalide konstruierte problemspezifische Wissenstests zusätzlich der Nachweis der Kriteriums- und Konstruktvalidität erbracht werden. Solange diese weiteren Validierungen aber noch ausstehen, wird durch den Nachweis der Kontentvalidität ein testtheoretisch begründetes Instrument bereitgestellt, das als «Standardmeßverfahren» eingesetzt werden kann. So konnte mit dem hier vorgestellten Test gezeigt werden, daß das Wissen zur Identifikation von Systemmerkmalen signifikant bis hochsignifikant mit der intellektuellen Operationsklasse «Verarbeitungskapazität» (nach Jäger 1982) korreliert und daß die Variable «Wissen» einen – gegenüber der Intelligenz: inkrementellen – Beitrag zur Vorhersage der Varianz der Variablen «Steuerungsleistung» leistet (siehe Süß, Oberauer & Kersting 1993).

Mit der Technik einer kontentvaliden Testentwicklung sind Einschränkungen verbunden, die zum Teil grundsätzlicher Art sind. An die Grenzen stößt das Verfahren bei der Diagnose von Handlungswissen aufgrund der schwer zu bewerkstellenden Bestimmung von Grundmengen z. B. von Eingriffen, unspezifischen Heuristiken oder bereichsspezifischen Faustregeln. Auch die Diagnose qualitativ-struktureller Wissensrepräsentationen kann mit dem hier verfolgten Ansatz nicht geleistet werden.

#### Anmerkungen

- 1 Danken möchten wir Herrn Dipl.-Psych. André Beau-ducel, der mit seinen kritisch-konstruktiven Beiträgen und vielfältiger aktiver Unterstützung das Entstehen dieser Arbeit tatkräftig begleitet hat sowie drei anonymen Gutachter(innen) für ihre wertvollen Anregungen.
- 2 Die Berechnungen wurden mit dem Programm SYSTAT von Wilkinson (1992) durchgeführt. SYSTAT berechnet zusätzlich zum F-Test korrigierte Signifikanztests zur Korrektur von Verletzungen der Testvoraussetzungen (z. B. Annahme der Gleichheit der Kovarianzen). Auch diese voraussetzungsärmeren Tests – und zwar sowohl der eher liberale Test von Huyn-Feld als auch der eher konservative von Greenhouse-Geisser (siehe Kirby 1993) – erwiesen sich als hochsignifikant. Für die Anregung zu diesem Prüfverfahren danken wir Herrn Prof. Dr. Werner W. Wittmann herzlich!

#### Literatur

- American Psychological Association (1985). Standards for Educational and Psychological testing. Washington, DC: American Psychological Association.
- Andresen, N. & Schmidt, U. (1993). Zur Invarianz von Problemlösestilen über verschiedene Bereiche. Zeitschrift für experimentelle und angewandte Psychologie, 40, 1–17.
- Berry, D. C. & Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. Psychological Research, 49, 7–15.
- Berry, D. C. & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. British Journal of Psychology, 79, 251–272.
- Broadbent, D. E., Fitzgerald, P. & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. British Journal of Psychology, 77, 33–50.
- Büscher, P. (1975). Einige testtheoretische Ansätze kriteriumsbezogener Leistungsmessung. In Heller, K. A. (Hrsg.), Leistungsbeurteilung in der Schule. Heidelberg: Quelle und Meyer, 137–157.
- Cronbach, L. J. (1971). Test Validation. In Thorndike, R. L. (Hrsg.), Educational measurement. Washington: American Council on Education, 443–507.
- Cronbach, L. J. (1984). Essentials of Psychological Testing (4. Aufl.). New York: Harper & Row.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281–302.
- Feger, B. (1984). Die Generierung von Testitems zu Lehrtexten. Diagnostica, 30, 24–46.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? Diagnostica, 29, 283–302.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. Zeitschrift für Psychologie, 193, 443–465.
- Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. Sprache & Kognition, 9, 143–154.
- Funke, J. (1991). Keine Struktur im (selbstverursachten) Chaos? Erwiderung zum Kommentar von Stefan Strohschneider. Sprache & Kognition, 10, 114–118.
- Gigerenzer, G. (1981). Messung und Modellbildung in der Psychologie. München: Reinhardt.

- Haider, F. (1992). Implizites Wissen und Lernen. Ein Artefakt? Zeitschrift für experimentelle und angewandte Psychologie, 39, 68–100.
- Herbig, M. (1976). Praxis lehrzielorientierter Tests. Düsseldorf: Schwann.
- Hörmann, H. J. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexem Problemlösen. Sprache & Kognition, 8, 23–31.
- Hübner, R. (1989). Methoden zur Analyse und Konstruktion von Aufgaben zur kognitiven Steuerung dynamischer Systeme. Zeitschrift für experimentelle und angewandte Psychologie, 36, 221–238.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. Diagnostica, 28, 195–226.
- Jäger, A. O. (1986). Validität von Intelligenztests. Diagnostica, 32, 272–289.
- Kersting, M. (1991). Wissensdiagnostik beim Problemlösen. Entwicklung und erste Bewährungsprobe eines kontentvalide konstruierten problemspezifischen Wissenstests. Unveröffentlichte Diplomarbeit im Forschungsprojektschwerpunkt «Intelligenz und Wissen» am Institut für Psychologie, Berlin: FU Berlin.
- Kirby, K. N. (1993). Advanced Data Analysis with SYSTAT. New York: Van Nostrand Reinhold.
- Klauer, K. J. (1983). Kriteriumsorientierte Tests. In Feger, H. & Bredenkamp, J. (Hrsg.), Enzyklopädie der Psychologie (Bd. 3. Messen und Testen). Göttingen: Hogrefe, 693–726.
- Klauer, K. J. (1984a). Kontentvalidität. Diagnostica, 30, 1–23.
- Klauer, K. J. (1984b). Über Parallelität, Reliabilität und Validität kontentvalider Paralleltests. Diagnostica, 30, 67–80.
- Kluwe, R. H. (1988). Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen. In Mandl, H. & Spada, H. (Hrsg.), Wissenspsychologie. Weinheim: Psychologische Verlags Union, 359–385.
- Kluwe, R. H., Misiak, C. & Haider, H. (1991). Modelling the Process of Complex System Control. In Milling, P. M. & Zahn, E. O. K. (Hrsg.), Computer-based management of complex systems. Proceedings of the 1989 International Conference of the System Dynamics Society. Berlin: Springer, 335–342.
- Leutner, D. (1989). Implementation und experimentelle Evaluation von Lernhilfen im computersimulierten Planspiel «Hunger in Nordafrika». In Schrettenbrunner, H. (Hrsg.), Software für den Geographieunterricht. Lüneburg: Selbstverlag des Hochschulverbandes für Geographie und ihre Didaktik e.V., 81–109.
- Lienert, G. A. (1961). Testaufbau und Testanalyse. Weinheim: Beltz.
- Loevinger, J. (1965). Person and population as psychometric concepts. Psychological Review, 72, 143–155.
- Morris, N. M. & Rouse, W. B. (1985). The effects of type of knowledge upon human problem solving in a process control task. IEEE Transactions on Systems, Man and Cybernetics, 15, 698–707.
- Moser, K. (1987). Inhaltsvalidität als Kriterium psychologischer Tests. Diagnostica, 33, 110–122.
- Plötzner, R., Spada, H., Stumpf, M. & Opwis, K. (1990). Learning qualitative and quantitative reasoning in a microworld for elastic impacts. European Journal of Psychology and Education, 4, 501–516.
- Plötzner, R. & Spada, H. (1992). Analysis-based learning on multiple levels of mental domain representation. In De Corte, E., Linn, M. C., Mandl, H. & Verschaffel, L.



(Hrsg.). Computer-based learning environments and Problem Solving. Berlin: Springer, 103–127.

Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme? Zeitschrift für Psychologie, 193, 63–84.

Putz-Osterloh, W., Bott, B. & Houben, I. (1988). Beeinflusst Wissen über ein realitätsnahes System dessen Steuerung? Sprache & Kognition, 7, 240–251.

Putz-Osterloh, W. & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. Zeitschrift für experimentelle und angewandte Psychologie, 28, 309–334.

Reichert, U. & Dörner, D. (1988). Heuristiken beim Umgang mit einem «einfachen» dynamischen System. Sprache & Kognition, 7, 12–24.

Resnick, L.B. & Ford, W.W. (1981). The Psychology of Mathematics for Instruction. Hillsdale, NJ: Erlbaum.

Roid, G.H. & Haladyna, Th.M. (1982). A Technology for Test-Item Writing. New York: Academic Press.

Schott, F. (1984). Regelgeleitete Itemkonstruktion. Ein Verfahren zur Definition von Itemuniversa und deren kontextualer Abbildung in Itemmengen für Tests und Treatments. Diagnostica, 30, 47–66.

Schott, F., Neeb, K.-E. & Wieberg, H.-J.W. (1981). Lehrstoffanalyse und Unterrichtsplanung. Eine praktische Anleitung zur Analyse von Lehrstoffen, Präzisierung von Lehrzielen, Konstruktionen von Lehrmaterialien und Überprüfung des Lernerfolgs. Braunschweig: Westermann.

Spada, H. & Reimann, P. (1988). Wissensdiagnostik auf kognitionswissenschaftlicher Basis. Zeitschrift für Differentielle und Diagnostische Psychologie, 9, 183–192.

Strohschneider, S. (1991a). Kein System von Systemen! Kommentar zu dem Aufsatz Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen von Joachim Funke. Sprache & Kognition, 10, 109–113.

Strohschneider, S. (1991b). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. Diagnostica, 37, 353–371.

Süß, H.M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. Diagnostica, 37, 334–352.

Süß, H.M., Kersting, M. & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz. Zeitschrift für Differentielle und Diagnostische Psychologie, 14, 189–203.

Süß, H.M., Oberauer, K. & Kersting, M. (1993). Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme. Sprache & Kognition, 12, 83–97.

Tergan, S.O. (1989a). Psychologische Grundlagen der Erfassung individueller Wissensrepräsentationen Teil I: Grundlagen der Wissensmodellierung. Sprache & Kognition, 8, 152–165.

Tergan, S.O. (1989b). Psychologische Grundlagen der Erfassung individueller Wissensrepräsentationen Teil II: Methodologische Aspekte. Sprache & Kognition, 8, 193–202.

Wilkinson, L. (1992). SYSTAT Statistics Version 5. Evanston, Illinois: SYSTAT Inc.

Dipl.-Psych. Martin Kersting, Deutsche Gesellschaft für Personalwesen, Grassistraße 12, D-04107 Leipzig, Telefon (0341) 717 02 64  
 Dr. Hein-Martin Süß, Universität Mannheim, Lehrstuhl Psychologie II, Schloß, Ehrenhof Ost, D-68131 Mannheim, Telefon (0621) 292 57 80

Anhang A: Beispielim Aufgabentyp I a, "Abhängigkeit"

Produktion

- a) Kommt in unserer Hemdenfabrik nicht vor.
- b) Die Höhe der Produktion ist von vielen anderen Variablen abhängig.
- c) Die Höhe der Produktion ist von mehreren anderen Variablen abhängig.
- d) Die Höhe der Produktion ist von wenigen anderen Variablen abhängig.

Anhang B: Beispielim Aufgabentyp I b, "Wirkung"

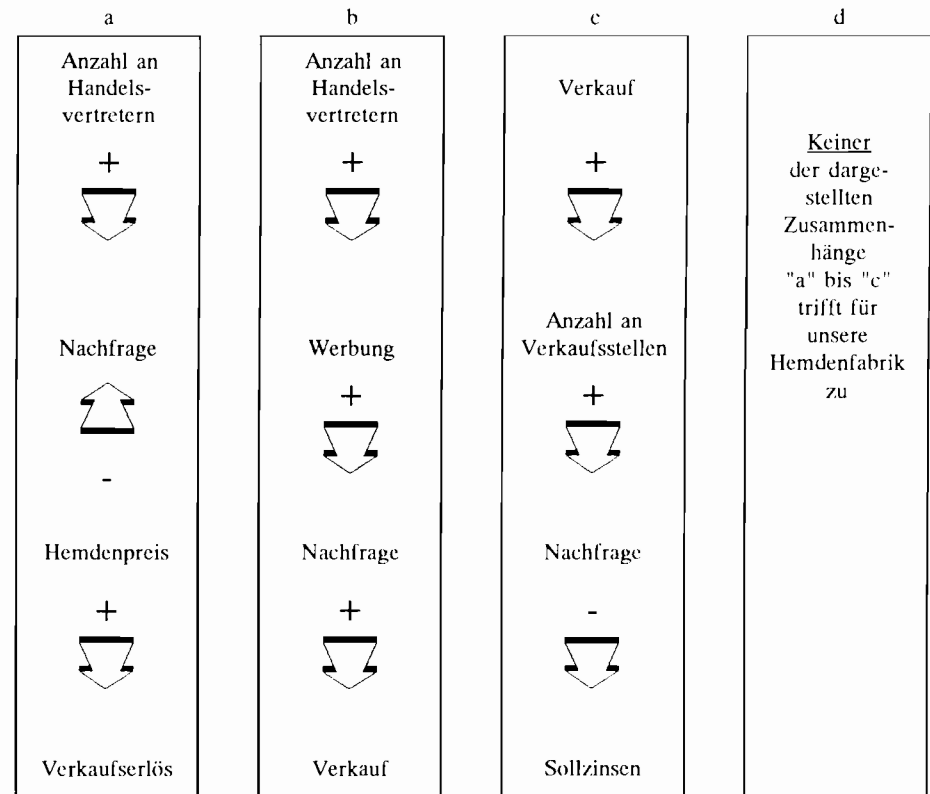
Hemdenpreis

- a) Kommt in unserer Hemdenfabrik nicht vor.
- b) Der Hemdenpreis beeinflusst viele andere Variablen.
- c) Der Hemdenpreis beeinflusst einige andere Variablen.
- d) Der Hemdenpreis beeinflusst wenige andere Variablen.

Anhang C: Beispielim Aufgabentyp II a, "Zusammenhänge verbal"

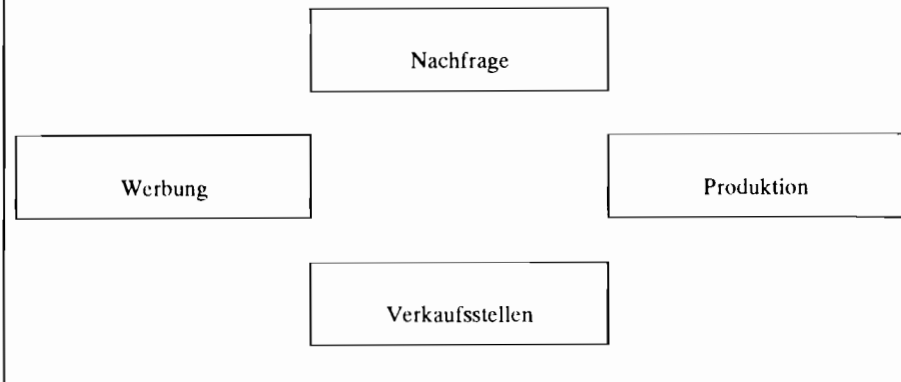
- a) Eine Erhöhung der Produktionskapazität (Anzahl der maximal herstellbaren Hemden) steigert die Arbeitsmotivation.
- b) Eine Erhöhung der Produktionskapazität senkt die Arbeitsmotivation.
- c) Eine Erhöhung der Arbeitsmotivation steigert die Produktionskapazität.
- d) Eine Erhöhung der Arbeitsmotivation senkt die Produktionskapazität.
- e) Die Produktionskapazität und die Arbeitsmotivation wirken wechselseitig aufeinander.
- f) Keine der Aussagen a)- e) ist richtig.

Anhang D: Beispielim Aufgabentyp II b, "Zusammenhänge figural" / Typ 1 "Rekognition"



Anhang E: Beispielim Aufgabentyp II c, "Zusammenhänge figural" / Typ 1 "Reproduktion"

Bitte tragen Sie die Verbindungen zwischen den Variablen als Pfeile ein. Vergessen Sie nicht, die Richtung des Zusammenhangs durch ein Plus- oder Minuszeichen zu markieren.



Anhang F: Beispielim Aufgabentyp III a, "Eigenschaften" (Mehrfachantworten erlaubt)

Ausgaben für Soziales

- Die Ausgaben für Soziales wirken auf den Lohn.
- Die Ausgaben für Soziales wirken über die Arbeitsmotivation auf die Produktionskapazität.
- Die Ausgaben für Soziales wirken zusammen mit dem Lohn auf die Arbeitsmotivation.
- Die Ausgaben für Soziales steigen und sinken ganz von alleine.
- Die Ausgaben für Soziales bleiben immer gleich hoch.
- In unserer Hemdenfabrik trifft keine der Aussagen a) - e) auf die Ausgaben für Soziales zu.

Anhang G: Beispielim Aufgabentyp III b, "Wichtigkeit"

Kosten für's Hemdenlager - Rohmaterial im Lager

- Die Variable Kosten für's Hemdenlager ist wichtiger als die Variable Rohmaterial im Lager
- Die Variable Rohmaterial im Lager ist wichtiger als die Variable Kosten für's Hemdenlager