

## Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen\*

Heinz-Martin Süß, Martin Kersting und Klaus Oberauer

Aus theoretischen Annahmen über den Einfluß von Intelligenz und bereichsspezifischem Wissen auf den Erfolg beim Lösen computersimulierter komplexer Probleme werden Hypothesen abgeleitet und geprüft. Die methodischen Voraussetzungen der Prüfung dieser Zusammenhangshypothesen werden diskutiert. Berichtet werden die Ergebnisse zweier Untersuchungen, die im Abstand von einem Jahr mit denselben Probanden durchgeführt wurden. 214 Oberstufenschülerinnen und -schüler von 13 Berliner Gymnasien nahmen an der ersten Untersuchung teil, 137 davon auch an der zweiten. Neben einem Test für das Berliner Intelligenzstrukturmodell wurden das Problem „Schneiderwerkstatt“ unter Intransparenzbedingung sowie ein spezifischer Wissenstest für dieses Problem mehrmals eingesetzt. Die interne Validität von Problemlösegütemaßen wird diskutiert und empirisch geprüft. Ergebnisse zur Stabilität von Intelligenz- und Problemlöseleistungen werden berichtet. Die Ergebnisse beider Untersuchungen bestätigen die Annahme, daß Wissen und Intelligenz substantielle Beiträge zur Vorhersage des Problemlöseerfolgs leisten.

Theoretical assumptions on the influence of intelligence and subject-specific knowledge on problem solving are discussed. On this basis, hypotheses are developed on the performance of solving computersimulated complex problems. The methodological assumptions necessary for testing these hypotheses are elaborated. The results of two investigations which were carried out within an interval of one year, on the same group of subjects are described. 214 high school pupils participated in the first study and 137 of them took part in the second one. In addition to a test for the "Berlin Model of Intelligence Structure" the system "Tailorshop" as well as the specific knowledge-test for this system were applied several times. The internal validity of the performance variables is discussed and empirically established. The stability of intelligence- and problem-solving-scores are reported. The results of both studies confirm the assumption that knowledge and intelligence contribute substantially to the prediction of problem solving success.

### 1. Einführung

Die Untersuchung menschlichen Handelns im Umgang mit computersimulierten Problemen ist seit mehr als 15 Jahren ein eigener, immer noch expandierender Forschungszweig der Psychologie. Dieses Forschungsgebiet etablierte sich in der deutschen Psychologie mit einer pointierten Kritik an der Validität klassischer Intelligenztests. Die als „Testintelligenz“ bezeichneten Fähigkeitskonstrukte, wurde behauptet, können Leistungen bei komplexen, „alltagsnahen“ Problemen nicht vorhersagen (Dörner, Kreuzig, Reither & Stäudel, 1983; Putz-Osterloh, 1981).

\* Dieser Beitrag wurde im Rahmen des von der „Kommission für Forschung und wissenschaftlichen Nachwuchs“ (FNK) der FU Berlin geförderten Forschungsprojektschwerpunktes „Intelligenz und Wissen“ erstellt. Prof. Dr. A. O. Jäger und Dipl. Psych. A. Beauducel danken wir für ihre kritisch-konstruktiven Beiträge zu einer früheren Fassung dieser Arbeit.

Nach einer Literaturübersicht von Kluwe, Misiak und Haider (in press) konnte nur in vier von 11 Untersuchungen ein bedeutender Zusammenhang zwischen Intelligenz- und Steuerungsleistungen ( $r > .40$ ) aufgezeigt werden; in den meisten Untersuchungen waren die Korrelationen nahezu Null. Ein Zusammenhang zu Intelligenzmaßen konnte nur unter Bedingungen gezeigt werden, in denen Variablen und Variablenverbindungen des simulierten Systems für die Probanden transparent waren. Lediglich Putz-Osterloh und Lürer (1981) konnten in einer Studie mit einem kleinen realitätsorientierten System (dem biologischen Tiefseesystem „Urgos“, einem 6 Variablensystem) mit  $r = .40$  auch unter Intransparenzbedingung einen bedeutsamen Zusammenhang feststellen. Sie halten jedoch „Urgos“ für ein System, das, verglichen etwa mit dem System „Schneiderwerkstatt“, in seiner Komplexität „drastisch reduziert“ und dadurch mit Intelligenztestaufgaben vergleichbar ist. Deshalb sehen sie ihren Befund als Bestätigung der Hypothese, daß klassische Intelligenztests nicht die Lösung von komplexen intransparenten Problemen vorhersagen können. Auch Kluwe et al. (in press) kommen zu dem Schluß, daß die vorliegenden Ergebnisse die Annahme eines Zusammenhangs nicht unterstützen. Allerdings weisen nach ihrer Meinung die Untersuchungen zahlreiche psychometrische Schwächen auf, so daß eine abschließende Bewertung bisher nicht möglich ist. Als Hauptmängel diskutieren sie Probleme der Bestimmung reliabler und intern valider Problemlösegütemaße und die ungeklärte Validität der Systeme.

In den vergangenen Jahren rückte neben der Frage nach dem Zusammenhang von Intelligenz und Problemlösen die Frage nach der Bedeutung von problemspezifischem Wissen im Umgang mit computersimulierten Systemen in den Vordergrund des Forschungsinteresses.

Hauptziel der beiden hier vorgestellten Untersuchungen war es, die Reliabilität und Stabilität von Leistungen beim Lösen alltagsorientierter computersimulierter Probleme und ihre Vorhersagbarkeit durch Intelligenz und Vorwissen zu untersuchen. Besonderes Gewicht haben wir auf die Bestimmung und Maximierung der psychometrischen Qualität der eingesetzten Meßinstrumente und Kriterien gelegt.

### 2. Theoretische Annahmen

#### 2.1. Intelligenz und Problemlösen

Unabhängig von der empirischen Frage nach der Existenz und Höhe von Zusammenhängen zwischen Intelligenztest- und Problemlöseleistungen muß die theoretische Interpretation potentieller Zusammenhangesbefunde vorab bestimmt werden. Drei Auffassungen konstituieren hier den Möglichkeitsraum:

1. Problemlöseszenarien werden als diagnostische Testaufgaben gesehen, die primär Informationen über kognitive Fähigkeiten liefern.

Im Fall eines ausbleibenden oder geringen Zusammenhangs zwischen Intelligenztest- und Problemlöseleistungen müßten Intelligenzstrukturmodelle – dieser Auffassung zufolge – um weitere Teilkonstrukte erweitert und computersimulierte Probleme in die Testbatterien für intellektuelle Fähigkeiten aufgenommen werden.

Eine solche Auffassung scheint dem Einsatz von computersimulierten Szenarien in der Eignungsdiagnostik zugrundezuliegen (vgl. Putz-Osterloh & Schroiff, 1987, Haselmann, 1990). Aus dieser Auffassung ergibt sich die Notwendigkeit, als Testaufgaben eingesetzte computersimulierte Szenarien ebenso wie Intelligenztestaufgaben zu validieren.

2. Die Bearbeitung von Problemlöseszenarien kann als Situation interpretiert werden, die intelligentes Handeln erfordert.

Der Zusammenhang zwischen Intelligenztestleistungen und Problemlöseleistungen ist dieser Auffassung zufolge ein Maß für die Kriteriumsvalidität der Intelligenztests. Von dieser Ansicht geht offenbar Putz-Osterloh (1981) aus, wenn sie den ausbleibenden Zusammenhang zwischen Intelligenztestergebnissen und dem Erfolg beim Management der „Schneiderwerkstatt“ als Indiz für die mangelnde Validität der Intelligenztests interpretiert. Eine solche Interpretation wird attraktiv durch die Annahme, daß computersimulierte Szenarien typische kognitive Anforderungen aus Alltagssituationen enthalten, die Intelligenztests zu prognostizieren beanspruchen. Es wird also angenommen, daß computersimulierte Szenarien „ökologisch valide“ sind. Diese „ökologische Validität“ der Szenarien ist aber bisher nur behauptet und nicht nachgewiesen. Zur Kritik des dabei involvierten „complexity bias“ siehe Kluwe et al. (in press).

3. Problemlöseleistungen können als Indikatoren für ein neues Konstrukt angesehen werden, welches sich auch theoretisch von schon etablierten Konstrukten der Intelligenz unterscheidet. Ein Kandidat für ein solches Konstrukt wäre beispielsweise Dörners „Operative Intelligenz“ (1986).

Dieser Auffassung zufolge entspricht der Zusammenhang zwischen Intelligenztest- und Problemlöseleistungen einem interessanten empirischen Zusammenhang zwischen dem Konstrukt „Intelligenz“ und dem neuen Konstrukt. Die Höhe des Zusammenhangs hat dann weder zwangsläufige Implikationen für die Validität eines der Konstrukte noch für Revisionen oder Bestätigungen der damit assoziierten Modelle. Voraussetzung für eine solche Interpretation ist jedoch, daß das neue Konstrukt hinreichend etabliert ist. Es muß also neben der reliablen Meßbarkeit, die auch für die Alternativen 1 und 2 notwendig ist, einen gewissen Grad an Stabilität über die Zeit und Generalität über verschiedene Problemsituationen haben.

Wir favorisieren die zweite Auffassung, weil sie die wenigsten Voraussetzungen enthält. Die Validierung von Intelligenztests an computersimulierten Szenarien hat allerdings nur dann auch praktische Relevanz, wenn die „ökologische Validität“ von Problemlöseszenarien belegt ist.

Interpretiert man die Bearbeitung von Problemlöseszenarien also als Situation, die intelligentes Handeln erfordert, ergeben sich folgende Fragen:

1. Wie reliabel sind Problemlöse gütemaße?
2. Welche Problemlöse gütemaße sind im Rahmen einer bestimmten Problemsituation adäquate Kriterien für Leistungen bei der Lösung des Problems?
3. Wieviel Varianz des Kriteriums Problemlöseleistung können Intelligenztestleistungen aufklären?

## 2.2. Intelligenz und Wissen beim Problemlösen

*Probleme* lassen sich analytisch zerlegen in eine *formale Problemstruktur* einerseits, eine *semantische Einkleidung* andererseits. Diese Zerlegung ist nicht nur für computersimulierte realitätsorientierte Probleme möglich (vgl. Hesse, 1982), sondern auch für Probleme der klassischen Problemlöseforschung. Kotovsky, Hayes und Simon (1985) haben das am Beispiel des „Tower of Hanoi“ gezeigt.

Die *kognitiven Voraussetzungen* der Probanden zur Lösung von Problemen lassen sich komplementär dazu in zwei Bereiche zerlegen: Den Bereich der *generellen intellektuellen Fähigkeiten* einerseits und den Bereich des *gegenstandsspezifischen Wissens* andererseits.

Wir nehmen an, daß es von der formalen Struktur eines Problems abhängt, welche Komponenten der Intelligenz in welchem Umfang zur Lösung des Problems notwendig sind.

Besteht das Problem in der Steuerung eines Systems vernetzter Variablen, so erfordert die formale Struktur dieses Problems unter anderem zwei Arten von Denkprozessen: *induktives Denken* und *deduktives Denken*. Mit Hilfe induktiver Denkprozesse können aus der Fülle der Variablenwerte und deren Veränderungen über die Zeit Regelmäßigkeiten extrahiert werden. Außerdem erlaubt das induktive Denken die Korrektur und Erweiterung des Vorwissens. Die Denkprozesse der Deduktion ermöglichen es, aus dem Wissen über Wenn-Dann-Zusammenhänge zwischen den Variablen konkrete Handlungsabsichten in einer bestimmten Situation für ein bestimmtes Ziel oder Subziel abzuleiten. Intelligenz ist also Voraussetzung für den Erwerb und die Anwendung von Systemwissen. Deduktives Denken ist außerdem möglicherweise für die Zerlegung eines gegebenen Ziels in Subziele relevant.

Induktives und deduktives Denken sind kognitive Leistungen, die traditionell in Intelligenztestaufgaben von großer Relevanz sind. Im Berliner Intelligenzstrukturmodell (BIS, vgl. Jäger, 1982) sind Aufgaben, die induktives und/oder deduktives Denken erfordern, der Operationsklasse „Verarbeitungskapazität“ (BIS-K) zugeordnet. Daraus ergibt sich die Erwartung, daß Intelligenztestleistungen Prädiktoren für Leistungen an computersimulierten Systemen sind. Wichtigster Einzelprädiktor des BIS sollte nicht die „Allgemeine Intelligenz“, das Integral aller Einzelfähigkeiten, sondern die operative Fähigkeit „Verarbeitungskapazität“ sein.

Auf der anderen Seite hängt es von der *semantischen Einkleidung* des Problems ab, welches spezifische Vorwissen die Problemlöseleistung mitbestimmt. Die semantische Einkleidung aktualisiert bei den Probanden Vorwissen aus dem entsprechenden Bereich. Richtiges Wissen über einen bestimmten Realitätsausschnitt ist in dem Maße nützlich für die Systemsteuerung, in dem das System den Realitätsausschnitt korrekt simuliert, den es aufgrund seiner semantischen Einkleidung zu simulieren vorgibt.

Der Steuerungserfolg bei computersimulierten Problemen mit einer semantischen Einkleidung hängt somit ab vom Umfang des für das System zutreffenden Wissens,

das die Problemlösenden schon zu Beginn der Bearbeitung an das Problem herantragen. Die Varianz des spezifischen Vorwissens über den simulierten Gegenstandsbereich ist dann weitgehend unabhängig von der Varianz der Intelligenz, wenn keine vergleichbaren Kontextbedingungen (Bedingungen für die Aneignung von Vorwissen) gegeben sind. Deshalb können Intelligenzmaße als Prädiktoren allein nicht die gesamte systematische Varianz der Kriterien aufklären. Ihre Prädiktionskraft ist für diesen Anwendungsbereich – wie für jeden, in dem gegenstandsspezifisches Wissen eine Rolle spielt – begrenzt. Diese Einschätzung konvergiert mit Ergebnissen der Expertise-Forschung, die die Bedeutung bereichsspezifischer Expertise für Problemlösen im Alltag hervorheben (Chi, Glaser & Farr, 1988). Für computersimulierte Systeme konnte beispielsweise Hesse (1982) zeigen, daß die Lösungsgüte bei einem computersimulierten System dann recht hoch mit Intelligenztestleistungen korrelierte, wenn das System ohne jede semantische Einkleidung präsentiert wurde. Die Steuerungsleistung bei demselben System mit einer plausiblen semantischen Einkleidung korrelierte nicht signifikant mit dem gleichzeitig eingesetzten Intelligenztest. Möglicherweise wurden die Effekte intellektueller Fähigkeiten durch die Effekte des Vorwissens überlagert.

### 3. Methodologische Voraussetzungen für die Untersuchung des Zusammenhangs von Intelligenz, Wissen und Problemlösen

#### 3.1. Zur Validierung von Intelligenzkonstrukten

Aus der Analyse der unterschiedlichen Realitätsnähe der Aufgabenanforderungen Indizien für die unterschiedliche Validität von „komplexen Problemen“ und Intelligenzmeßinstrumenten abzuleiten, ist nicht sinnvoll. Denn Inhaltsvalidität ist niemals ein Anliegen der Intelligenzforschung gewesen (Jäger, 1986). Aufgrund des generellen Geltungsanspruchs der Intelligenzkonstrukte wäre es auch nicht möglich, Aufgabenproben aus allen für bedeutsam erachteten Realitätsbereichen in Intelligenztests zu integrieren. Die Geltungsbereiche der Intelligenzmodelle müssen deshalb sukzessive durch Validierung an externen Kriterien ausgelotet werden.

#### 3.2. Aggregation und Symmetrie als Voraussetzungen für die Validierung von Intelligenzkonstrukten

Ziel der Intelligenzstrukturforschung ist es, die hochgradig generellen kognitiven Fähigkeiten des Menschen möglichst umfassend zu konzeptualisieren und mit Tests zu erfassen. Strukturmodelle sind dann revisionsbedürftig, wenn gezeigt werden kann, daß sie dazu nicht oder nicht in ausreichendem Maße in der Lage sind.

Als Bewährungsprobe für die Güte von Fähigkeitskonstrukten gilt der Nachweis einer gelungenen Vorhersage von kognitiven Leistungen, und zwar über längere Zeiträume und äquivalente Situationen hinweg. Damit kann nicht die Leistung bei der Lösung einer einzelnen Aufgabe das Kriterium einer Prognose sein, sondern nur eine

Stichprobe von Aufgaben. Für jede Form der Validierung von Intelligenz-Meßinstrumenten an Problemlösegütemaßen müssen aggregierte Kriterienmaße verwendet werden: aggregiert über Meßwiederholungen typgleicher (paralleler) Aufgaben und – bei weiterreichendem Generalitätsanspruch – über repräsentative Aufgabentstichproben verschiedener Aufgabenklassen. Diese Aggregation erst sichert die Symmetrie der in Beziehung gesetzten Messungen (Wittmann & Matt, 1986; Wittmann, 1988). Für eine repräsentative Aufgabenauswahl fehlt aber auf der Seite der komplexen Probleme bislang jede Theorie. Auch eine konsensfähige Problemmklassifikation existiert nicht, wie etwa Kluwe et al. (in press) zu Recht beklagen.

#### 3.3. Reliabilität und interne Validität von Problemlösegüteindikatoren

Voraussetzung für die Bestimmung reliabler Problemlösegütemaße ist, daß den Probanden *eindeutige* Ziele vorgegeben werden. Demgegenüber wird im Hinblick auf die Sicherung der angestrebten „ökologischen Validität“ häufig gefordert, daß Ziele nicht spezifiziert oder zumindest polytelisch formuliert sein sollen (Dörner et al., 1983). Dagegen ist nichts einzuwenden, solange nicht gleichzeitig die Leistungen von Probanden miteinander verglichen und durch Prädiktoren vorhergesagt werden sollen. Denn nur bei präzisen Zielvorgaben kann man voraussetzen, daß alle Probanden konsistent das Ziel anstreben, nach dem ihre Leistung bewertet und verglichen wird.

Auch wenn präzise Ziele vorgegeben werden, dürften die Retest-Reliabilitäten der Problemlösegütemaße geringer sein als die von Intelligenztests. Jeder Simulationsdurchgang liefert lediglich „single act“-Kriterien (Fishbein & Ajzen, 1974), deren Vorhersagbarkeit wegen ihrer mangelnden Zuverlässigkeit a priori stark eingeschränkt ist. Mehrere Simulationsdurchgänge sind erforderlich, um die psychometrische Qualität der Problemlösegütekriterien durch Aggregation zu sichern.

Präzise Zielvorgaben sind auch Voraussetzung für die interne Validität der Problemlösegütemaße. Ein Problemlösegütemaß wird im folgenden dann als intern valide bezeichnet, wenn das Gütemaß die Steuerungsleistungen adäquat abbildet. Kluwe et al. (in press) diskutieren in Anlehnung an Sternberg (1982) diese Frage unter dem Aspekt der Quantifizierbarkeit. Wir halten ein Gütemaß dann für adäquat, wenn es das Ausmaß zweckrationalen Handelns in Bezug auf das vorgegebene Ziel erfaßt.

#### 3.4. Zum Zusammenhang von Intelligenz und Wissen

In der Geschichte der Intelligenzstrukturforschung wurden Intelligenz und Wissen lange Zeit begrifflich und empirisch nicht differenziert. Bedarf nach Differenzierung ergab sich im Kontext kulturvergleichender Untersuchungen, der Intelligenzfairneßdebatte (s. Möbus, 1978; Brambring, 1983) und den Versuchen, „culture free“ bzw. „culture fair“-Tests zu konstruieren (s. Groffmann, 1983). Oder sie war das Ergebnis der Modellentwicklung bei Cattell (1963, 1971), mit der Unterscheidung der „fluiden“ von der „kristallinen“, der wissensbasierten Seite der Intelligenz. Auch wenn die Geschichte der Intelligenzforschung zeigt, daß eine Trennung beider Komponenten

nicht vollständig möglich ist, sondern Erfahrung, Vorwissen und Übung alle Intelligenztestleistungen mitbestimmen (Groffmann, 1983), so werden die Aufgaben zur Erfassung „fluider“ Intelligenzkomponenten doch so konstruiert, daß spezifischem Wissen eine möglichst geringe Bedeutung zukommt. Im BIS drückt sich die Nicht-Trennbarkeit von Intelligenz und Wissen bei der Lösung einer Aufgabe in der Annahme aus, daß jede Intelligenzleistung durch operative *und* inhaltsgebundene Fähigkeitskomponenten bedingt ist. Die operative Modalität beschreibt den „fluiden“ Teil der Intelligenz, die Inhaltsmodalität (mit verbalen, numerischen und figuralen Fähigkeiten) den verbleibenden wissensbasierten Teil der Intelligenz. Aus diesem Grund muß stets zumindest eine moderate Korrelation von Intelligenz und Wissensmaßen erwartet werden. Ebenfalls in den Bedeutungshof des Konstrukts Intelligenz fällt die Lernfähigkeit und damit die Fähigkeit zum Erwerb, zur Strukturierung und Anwendung von Wissen. Damit kann erwartet werden, daß Intelligenz und Wissen umso höher korrelieren, je ähnlicher die Lernbedingungen für die Probanden sind. Daher ist die inkrementelle Varianz, die Wissen und Intelligenz jeweils zur Aufklärung von Problemlöseleistungen beitragen, interessanter als die Höhe einzelner Korrelationen.

## 4. Empirische Untersuchungen

### 4.1. Hypothesen

Auf der Grundlage der methodologischen Überlegungen lassen sich die theoretischen Annahmen zu empirisch prüfbareren Hypothesen präzisieren:

1. Gelingt es, Problemlöseleistungen reliabel und intern valide zu erfassen, dann leisten Intelligenzkonstrukte einen substantiellen Beitrag zur Vorhersage der Problemlöseleistungen.
2. Unter denselben Voraussetzungen sind auch allgemeines und spezifisches Vorwissen Prädiktoren von Problemlöseleistungen.
3. Beide Prädiktorgruppen erklären zusammen signifikant mehr Varianz am Kriterium als jede der beiden Gruppen alleine.
4. Auf der Seite der Intelligenz ist die Verarbeitungskapazität (BIS-K) der stärkste Prädiktor für Problemlöseleistungen.
5. Intelligenz und spezifisches Vorwissen interkorrelieren miteinander, der Zusammenhang ist aber eher gering.

Diese Hypothesen wurden in zwei Untersuchungen an derselben Probandenstichprobe im Abstand von einem Jahr geprüft.

### 4.2. Erstuntersuchung (EU)

Die Erstuntersuchung ist an anderer Stelle ausführlich dargestellt (Süß, Oberauer & Jäger, 1990; Süß & Kersting, 1990)<sup>1</sup> und wird deshalb hier nur kurz zusammengefaßt.

<sup>1</sup> Ausführliche Manuskripte können bei den Autoren angefordert werden.

In der zweitägigen Untersuchung bearbeiteten 214 Schülerinnen und Schüler von 13 Berliner Oberschulen das System „Schneiderwerkstatt“ (in der Version 2.3 von J. Funke) einmal für 12 simulierte Monate und anschließend zweimal für sechs Monate. Die „Schneiderwerkstatt“ simuliert eine kleine Hemdenfabrik mit 24 Variablen, darunter 10 Eingriffs- und 7 verdeckte Variablen. Die „Schneiderwerkstatt“ wurde ausgewählt, da dieses System eine plausible semantische Einkleidung besitzt und für dieses System unter Intransparenzbedingung bisher nur negative Befunde zum Zusammenhang von Intelligenz und Steuerungsleistungen vorliegen, aus allerdings methodologisch unzureichenden Untersuchungen. Das System wurde unter Intransparenzbedingung und mit eindeutiger Zielvorgabe bearbeitet. Vor und nach der ersten Systemsteuerung und noch einmal am Ende der drei Durchgänge wurde mit einem speziell für das System konstruierten Wissenstest (WIS-1) das Systemwissen und das Handlungswissen der Probanden erhoben. Unter Systemwissen wurde dabei das Wissen über die Eigenschaften und Zusammenhänge der Variablen der „Schneiderwerkstatt“ verstanden. Das Handlungswissen ist demgegenüber Wissen über adäquate Steuerungseingriffe bei gegebenen Systemzuständen und Zielen. Der Test zur Erhebung des Systemwissens wurde nach den Methoden zur Konstruktion kontentvalider Meßinstrumente entwickelt (Kersting, 1991). Für das Instrument zur Erhebung des Handlungswissens kann hingegen keine Kontentvalidität beansprucht werden. Schon vor der Problembearbeitung wurden die intellektuellen Fähigkeiten mit einem Test für das Berliner Intelligenzstrukturmodell (BIS-2) sowie allgemeines Wirtschafts-Fachwissen durch einen standardisierten Wirtschaftslehretest (WLT; Deutsche Gesellschaft für Personalwesen e. V. sowie Krumm & Seidel, 1970) erhoben.

Entgegen unseren Erwartungen fanden sich keine Zusammenhänge zwischen den üblichen Problemlöse Güteindizes der „Schneiderwerkstatt“ und den verschiedenen Prädiktoren. Diese Güteindizes waren das Gesamtvermögen am Ende der Steuerung (Hörmann & Thomas, 1989) und das Trendmaß nach Funke (1983).

Einige Ergebnisse weckten jedoch – trotz Übereinstimmung mit dem vorgegebenen Ziel – Zweifel an der internen Validität dieser Gütemaße. So erreichte z. B. keine einzige Schülerin und kein einziger Schüler das vorgegebene Ziel, Steigerung des Gesamtvermögens, und mehr als die Hälfte der Probanden lag am Ende unter dem Ergebnis des „Nulllaufs“ (Systemlauf ohne Eingriff). Wir haben daher eine eingehende Aufgabenanalyse für das Problem durchgeführt. Dazu wurde das vorgegebene Ziel, Maximierung des Gesamtvermögens am Ende der Steuerung, in seine beiden Komponenten zerlegt:

1. *Die Maximierung des Verkaufs.* Die Maximierung des Verkaufs von Hemden ist die einzige Einnahmequelle, die das System vorsieht.
2. *Die Maximierung der Gewinnspanne pro verkauftem Hemd.*  
Für die Gewinnspanne zeigte sich, daß *alle* Probanden über die Monate gemittelt eine *negative* Gewinnspanne pro verkauftem Hemd erzielten. Da der Gewinn das Produkt aus der Anzahl verkaufter Hemden und der Gewinnspanne pro verkauftem

Hemd ist, ergab sich der Effekt, daß gute Problemlöser, die erfolgreich das Teilziel „Maximierung des Verkaufs“ verfolgten, damit nicht ihren Gewinn, sondern ihren Verlust maximierten. Da das System die Gewinnspanne nicht anzeigt, konnten die Probanden nicht erkennen, daß ihr „eigentlich“ vernünftiges Verhalten ihnen schadete. So konnte ein guter Problemlöser, der viele Hemden zu einem relativ geringen Verlust pro Hemd verkaufte, insgesamt genausoviel Verlust machen wie ein schlechter Problemlöser, der nur wenige Hemden verkaufte und dabei eine hoch negative Gewinnspanne pro Hemd realisierte. Das Gesamtvermögen und davon abgeleitete Maße konnten dadurch nicht zwischen guten und schlechten Problemlösern unterscheiden. Verschiedene empirische Tests bestätigten Folgehypothesen, die sich aus dieser Analyse ergaben. Wir haben daher unsere Hypothesen zum einen an den Teilgütekriterien für die beiden Subziele geprüft und zum anderen ein *neues Problemlösegütemaß* definiert. Dazu werden die Teilgütekriterien – Verkauf und Gewinnspanne pro verkauftem Hemd – nach Normalisierung *addiert*.

Mit den Teilgütekriterien und dem neuen Gütemaß fanden wir nun signifikante Zusammenhänge sowohl zum Intelligenzkonstrukt „Verarbeitungskapazität“, als auch zum allgemeinen und spezifischen Vorwissen. Während das allgemeine Vorwissen (Wirtschaftslehretest) mit jedem Bearbeitungsdurchgang an Prädiktionskraft verlor, nahm die Bedeutung des spezifischen Wissens erwartungsgemäß zu. Obwohl Wissen substantiell mit Intelligenz korrelierte (vor allem mit BIS-K, dessen Zusammenhang mit dem Wissensgesamtscore zum 1. Meßzeitpunkt  $r = .45$  betrug), wurden in multiplen Regressionen der neuen Problemlösegütemaße die inkrementellen Varianzbeiträge beider Prädiktorgruppen signifikant. Die multiplen Korrelationskoeffizienten erreichten  $R = .60$  für den ersten Durchgang bzw.  $R = .51$  für die aggregierten Gütemaße aus dem 2. und 3. Durchgang der Bearbeitung der „Schneiderwerkstatt“. Die Ergebnisse hielten einer Kreuzvalidierung im wesentlichen stand.

Alle diese Ergebnisse standen unter dem Vorbehalt, daß die vorgenommene Analyse der fehlenden internen Validität der herkömmlichen Gütemaße unter den spezifischen Bedingungen der Erstuntersuchung (EU) zutrifft.

### 4.3. Wiederholungsuntersuchung (WU)

Zentrale Ziele der Wiederholungsuntersuchung (WU) waren die Überprüfung (1) der Replizierbarkeit der EU-Ergebnisse, (2) der Stabilität der Intelligenz- und Steuerungsleistungen und (3) der Aufgabenanalyse, die dem neuen Problemlösegütemaß zugrundeliegt.

Zur Überprüfung der Aufgabenanalyse mußte das Problem so modifiziert werden, daß es den Probanden im Durchschnitt möglich wurde, das System unter Intransparenzbedingung zielentsprechend, d.h. mit Erfolg zu steuern. Dadurch sollte die Gewinnspanne der Probanden im positiven Bereich liegen. Nach der oben ausgeführten Hypothese wäre damit die interne Validität der traditionellen Gütemaße wiederhergestellt und so eine *Äquivalenz* dieser traditionellen Gütemaße mit dem neu definierten Gütemaß erreicht. Die Modifikation des Problems sollte insbesondere folgende Konsequenzen haben:

1. Während in der EU die Anzahl verkaufter Hemden negativ mit dem Gesamtvermögen korrelierte, weil durch den Verkauf Verluste maximiert wurden, sollen in der WU dagegen beide Teilgütekriterien positiv mit dem Gesamtvermögen korrelieren.

2. Im Gegensatz zur EU sollte in der WU das neu definierte Problemlösegütemaß so hoch mit den üblichen Gütemaßen (Gesamtvermögen am Ende und Trendmaß) korrelieren, daß die Maße als äquivalent bezeichnet werden können. Diese Äquivalenz sollte sich außerdem in gleichartigen Korrelationsmustern aller drei Problemlösegütemaße zu Außenkriterien ausdrücken.

Nur wenn diese Hypothesen bestätigt werden, können wir die Definition eines neuen Problemlösegütemaßes als gerechtfertigt akzeptieren.

#### 4.3.1 Stichprobe und Untersuchungsplan

Von den 214 Teilnehmern der Erstuntersuchung nahmen 137 (64,0% der EU-Stichprobe) an beiden Tagen der Wiederholungsuntersuchung teil. (Durchschnittliches Alter zum Zeitpunkt der WU: 18,6 Jahre,  $s = .67$ ; Geschlecht: 41,2% weiblich, 58,8% männlich). Signifikante Leistungsunterschiede zwischen Teilnehmern und Nichtteilnehmern der WU sind nur bei den BIS-K Ergebnissen zu verzeichnen: die WU-Teilnehmer waren signifikant besser. Von 10 Pbn, die an der WU teilgenommen haben, fehlten aufgrund einer technischen Panne die Problemlösedaten der EU. Bedingt durch den Untersuchungsplan der EU lagen nur für 90 Pbn der WU alle Wissensdaten der EU vor.

Den Untersuchungsplan der Wiederholungsuntersuchung gibt Tab. 1 wieder. Alle hier dargestellten Ergebnisse beziehen sich auf die Stichprobe der WU-Teilnehmer (auch die EU-Ergebnisse beziehen sich – mit Ausnahme von Tab. 3 – nur auf die Probanden, die auch an der WU teilgenommen haben). Unterschiede zwischen den experimentellen Gruppen bleiben hier unberücksichtigt. Dieses Vorgehen erscheint zulässig, da die durch experimentelle Variation möglicherweise induzierten Gruppenunterschiede sich entweder gar nicht oder nur gegen die Zusammenhangshypothesen auswirken dürften. (Eine nach Gruppen getrennte Auswertung ergibt übrigens dieselben Zusammenhangsmuster).

#### 4.3.2 Instrumente und Methoden:

##### Modifikationen des Systems „Schneiderwerkstatt“

Das System „Schneiderwerkstatt“ wurde gegenüber der Erstuntersuchung (EU) modifiziert, um die Hypothesen der Aufgabenanalyse zu prüfen. Diese Vereinfachung der Systemsteuerung wurde – ohne wesentliche Eingriffe in die formale Struktur der Variablenvernetzung – durch vier Maßnahmen erreicht:

1. Die Kosten für die Investitionen wurden im Verhältnis zu ihrem Nutzen verringert. Verstärkt wurde die Wirkung der Werbung auf die Nachfrage.

2. Die Probanden bekamen solche Informationen, die einem Manager in der Realität zur Verfügung stehen, die aber – aus für uns unverständlichen Gründen – den Probanden in der „Schneiderwerkstatt“ üblicherweise nicht gegeben werden. Das betrifft beispielsweise die Kosten für die Anschaffung von Maschinen.

*Tabelle 1:* Ausschnitt aus dem Untersuchungsplan der Wiederholungsuntersuchung

Gruppe 1 (EXPL) N=49 (35,8%)	Gruppe 2 (INSTR) N=48 (35,0%)	Gruppe 3 (KON) N=40 (29,2%)
1. Tag	1. Tag	1. Tag
BIS-3	BIS-3	BIS-3
2. Tag	2. Tag	2. Tag
SWS-Einführung + 2 Übungsmonate	SWS-Einführung + 2 Übungsmonate	SWS-Einführung + 2 Übungsmonate
selbständige Exploration der SWS	Studium der SWS- Vernetzungsgrafik	1
WIS-D1	WIS-D1	WIS-D1
SWS-A	SWS-A	SWS-A
SWS-B	SWS-B	SWS-B
WIS-D2	WIS-D2	WIS-D2

EXPL	Explorationsgruppe (Gruppe 1): Die Pbn dieser Gruppe konnten das System innerhalb einer vorgegebenen Zeit von 30 Minuten frei erkunden.
INSTR	Instruktionsgruppe (Gruppe 2): Der zweiten Gruppe wurde für 30 Minuten die Vernetzungsgrafik des Systems anhand eines Tafelbildes erläutert und anschließend zum Selbststudium überlassen. Bei der darauffolgenden Systemsteuerung verfügten die Probanden nicht mehr über die Vernetzungsgrafik.
KON	Kontrollgruppe (Gruppe 3).
BIS-3	Test zum Berliner Intelligenzstrukturmodell, Version 3
SWS	Simulationsprogramm „Schneiderwerkstatt“, Berliner Version; programmiert in Turbo-C von J. Faulhaber. A und B bezeichnen die Bearbeitungsdurchgänge mit jeweils neuen Startwerten
WIS	Wissenstest zur „Schneiderwerkstatt“, Version 2. D1 und D2 bezeichnen die Meßzeitpunkte.
WLT	Wirtschaftslehretest. DGP e. V., Hannover und V. Krumm/G. Seidel. (nur in der Erstuntersuchung erhoben).

3. Die Programmbedienung wurde vereinfacht und gegen Eingriffsfehler zusätzlich abgesichert.

4. Die Probanden hatten die Möglichkeit, die Variablenwerte des Vormonats einzublenden.

Es soll betont werden, daß diese Veränderungen weder die Intransparenz bezüglich verborgener Variablen und -Zusammenhänge noch den generellen Charakter des Systems als „komplexes Problem“ mit den für diese Probleme postulierten Anforderungen beeinträchtigen.

#### Weitere eingesetzte Instrumente

Zur Messung der Intelligenz wurde ein neuer, verbesserter Test für das BIS (BIS-3) eingesetzt, der auf der Basis der EU-Ergebnisse aus der Vorläuferversion BIS-2 weiterentwickelt wurde. Trotz der durchgeführten Modifikationen kann – mit einer konservativen Schätzung – die Stabilität der Intelligenztestleistungen bestimmt werden.

Der Wissenstest wurde anhand der EU-Ergebnisse überprüft und modifiziert. Items, die ursprünglich lediglich zur Sicherung der Inhaltsvalidität aufgenommen wurden, jedoch keine Varianz aufwiesen, wurden ausgetauscht. Ansonsten wurde an dem für WIS-1 erfolgreich angewandten Konstruktionsprinzipien zur Sicherung der Kontenvalidität festgehalten.

In diesem problemspezifischen Wissenstest (WIS-2) wurde das Systemwissen der Probanden durch Fragen zu Eigenschaften einzelner Variablen und zu Zusammenhängen von Variablen erhoben. Die Fragen zu Zusammenhängen von Variablen wurden dabei zum einen im Multiple-Choice-Verfahren mit verbal formulierten Wenn- Dann-Aussagen zu je zwei Variablen erhoben, und zum anderen mußten die Probanden in mehreren Aufgaben je vier Variablen des Systems durch Pfeile, die den Variablenbeziehungen entsprachen, verbinden. Erweitert wurden die Methoden zur Erfassung von Handlungswissen. Dazu wurden die Aufgaben zum Eingriffswissen aus der Erstuntersuchung verbessert und an die Modifikationen der „Schneiderwerkstatt“ adaptiert. Bei diesen Aufgaben sollen die Probanden zu standardisiert vorgegebenen Systemzuständen und wechselnden vorgegebenen Zielen unter sechs Eingriffsalternativen die kurzfristig wirksamste auswählen. Als zusätzlichen Aufgabentyp zur Erfassung von Handlungswissen wurden 10 Items mit bereichsspezifischen Heuristiken (Faustregeln) vorgegeben, die als richtig oder falsch beurteilt werden mußten.

#### 4.3.3 Auswertungskonzeption

Als Problemlösegütemaße für die „Schneiderwerkstatt“ wurden die traditionellen Gütemaße („Gesamtvermögen“ und „Trendmaß“) und das neue Gütemaß herangezogen. Sowohl der Zusammenhang der Gütemaße untereinander als auch deren Zusammenhang mit den Prädiktoren werden berichtet. Für einige nachfolgende Auswertungsschritte wurde aus Gründen der besseren Vergleichbarkeit mit EU-Ergebnissen das neue Gütemaß verwendet.

Die Bildung der Wissensskalen folgte der Auswertung der EU (s. Kersting, 1991): Um eine Gleichgewichtung der Einzelitems zu erreichen, wurden die Items zunächst über Pbn, Gruppen und Meßzeitpunkte z-standardisiert. Skalen wurden vorläufig nur für die fünf Aufgabentypen gebildet. Während die drei Aufgabentypen zum Systemwissen zusammengefaßt wurden, wurde bei den beiden Aufgaben zum Handlungswissen auf eine Zusammenfassung vorläufig verzichtet, weil ihre Interkorrelation zu gering und ihr Verhältnis theoretisch zu wenig bestimmt ist.

### 4.4 Ergebnisse

#### 4.4.1 Stabilität der BIS-Skalen

Tabelle 2 gibt die Stabilitätskoeffizienten der 8 BIS-Skalen über ein Jahr wieder. Die Stabilitätskoeffizienten sind mit Korrelationen von bis zu .90 für die Skala Verarbeitungskapazität insgesamt sehr hoch. Die geringste Stabilität haben Merkfähigkeit (BIS-M) und Einfallsreichtum (BIS-E). Während für BIS-M die geringere Reliabilität der Skala (als spezifisches Problem von Gedächtnistests) die Ursache sein sollte, ist die geringere Stabilität der E-Skala multipel bedingt: Schwächen der Aufgaben, generelle Probleme der psychometrischen Kreativitätsmessung und Stabilitätsgrenzen des intendierten Konstrukts könnte mögliche Ursachen sein. Die Koeffizienten sind insgesamt im Vergleich zu den vorliegenden Stabilitätskoeffizienten in der Literatur (s. etwa Amthauer, 1973) und zu früheren Tests für das BIS (Baukasten-A/B; Schmidt et al. 1986) beachtlich hoch.

Tabelle 2: Stabilitätskoeffizienten der BIS-Skalen (Korrelationen der Ergebnisse aus EU und WU)

Stabilität	B	M	E	K	F	V	N	AI
(N = 137)	.86	.74	.65	.90	.78	.80	.87	.88

B = Bearbeitungsgeschwindigkeit, M = Merkfähigkeit, E = Einfallsreichtum, K = Verarbeitungskapazität, F = figurales (anschauungsgebundenes) Denken, V = verbales (sprachgebundenes) Denken, N = numerisches (zahlegebundenes) Denken, AI = Allgemeine Intelligenz (Integral aller Teilfähigkeiten).

#### 4.4.2 Interne Validität der Problemlösegütemaße

Ziel der Modifikationen des Systems Schneiderwerkstatt war es, das System so zu modifizieren, daß die Probanden unter den gegebenen Bedingungen das System zielentsprechend, d. h. mit Erfolg, steuern können.

Tabelle 3: Leistungen der Probanden nach 12monatigem Management der SWS (Angaben in Prozent)

SWS-Bearbeitungsdurchgang	EU			WU	
	SWS-A	SWS-B <sup>1</sup>	SWS-C <sup>1</sup>	SWS-A	SWS-B
Steuerungsleistungen					
Insgesamt Gewinn erwirtschaftet	0,0	0,9	0,4	52,6	73,7
Mehr als die Hälfte der Monate mit Gewinn	0,8	3,4	5,1	57,7	75,2
Kein Monat mit Gewinn	69,8	73,2	65,5	7,3	5,1
Besser als der Nulllauf	0,0	65,5	46,4	98,5	100

EU: N = 214; WU: N = 137 <sup>1</sup> Durchgang B und C der EU: 6 Simulationsmonate

SWS Simulationsprogramm „Schneiderwerkstatt“, Berliner Version; programmiert in Turbo-C von J. Faulhaber. A und B bezeichnen die Bearbeitungsdurchgänge mit jeweils neuen Startwerten.

Als Indikatoren für den Erfolg dieser Maßnahme können zunächst deskriptive Kennwerte herangezogen werden. Tabelle 3 gibt eine Übersicht über einige wichtige Ergebnisse. Danach kann festgestellt werden, daß die Modifikationen erfolgreich waren. Die Daten der Wiederholungsuntersuchung ermöglichen also die Überprüfung der Gültigkeit der Aufgabenanalyse der Erstuntersuchung, die der Definition des neuen Problemlösegütemaßes in der EU zugrunde lag.

Tabelle 4: Interkorrelationen der Teilgütekriterien und Korrelationen mit den Problemlösegütemaßen in EU und WU

	Erstuntersuchung			Wiederholungsuntersuchung		
	Verk-A	Gekap-A	Trend-A	Verk-A	Gekap-A	Trend-A
PLG-A		.23*	.30**		.86**	.83**
GewSp-A	.51**	.61**	.62**	.17	.33**	.32**
Verk-A		-.20	-.10		.71**	.63**
	Verk-BC	Gekap-BC	Trend-BC	Verk-B	Gekap-B	Trend-B
PLG-BC <sup>1</sup> bzw. B		.23*	.34**		.80**	.80**
GewSp-BC <sup>1</sup> bzw. B	.26*	.74**	.72**	.23*	.14	.16
Verk-BC <sup>1</sup> bzw. B		-.33**	-.15		.56**	.58**

EU: N = 127; WU: N = 137; \* p < .05 \*\* p < .01;

<sup>1</sup> für die EU wurde jeweils das Aggregat aus Durchgang B und C zugrundegelegt; PLG = neues Problemlösegütemaß, GewSp = Gewinnspanne pro verkauftem Hemd, Verk = Anzahl verkaufter Hemden, Gekap = Gesamtkapital nach dem letzten Monat, Trend = Trendmaß; A, B und C kennzeichnen die SWS-Bearbeitungsdurchgänge; BC = Aggregat aus Durchgang B und C der EU.

Die Ergebnisse in Tabelle vier und sechs bestätigen die Prognosen:

1. Die Teilgütekriterien „Verkauf“ und „Gewinnspanne pro verkauftem Hemd“ korrelieren in der WU positiv mit den traditionellen Gütekriterien.
2. Das neu definierte PLG korreliert hoch mit den beiden traditionellen Gütekriterien. Die Korrelationsmuster aller drei Gütekriterien mit Wissen und Intelligenz sind äquivalent.

Damit wird durch die hier vorgenommene Analyse die Annahme fehlender Validität traditioneller Gütemaße unter den spezifischen Bedingungen der EU bestätigt. Die Definition eines neuen Gütemaßes war notwendig und gerechtfertigt. Für die WU ist die Äquivalenz der Gütekriterien und auf das neue PLG verzichtet worden.

#### 4.4.3 Reliabilität der Problemlösegütemaße

Die Reliabilität der Steuerungsleistungen konnte gegenüber der Erstuntersuchung nochmals verbessert werden. Betrug die durchschnittliche Interkorrelation zwischen den drei SWS-Durchgängen in der EU noch r = .57, so stieg die Korrelation

in der WU auf  $r = .74$  (s. Tab. 5). Da – wie in der EU – sich die Startwerte zwischen den beiden Simulationsdurchgängen unterschieden, kann dieser Wert als sehr hoch eingestuft werden.

Tabelle 5: Reliabilität und Stabilität der Steuerungsleistungen

	PLG-BC (EU)	PLG-B (WU)	PLG-AB (WU)
PLG-A (EU)	.63**		
PLG-A (WU)		.74**	
PLG-ABC (EU)			.46**

#### 4.4.4 Stabilität der Steuerungsleistungen

Die Korrelation zwischen den aggregierten Steuerungsleistungen aus EU und WU beträgt  $r = .46$ . Berücksichtigt man die durchgeführten Modifikationen des Systems und die Zeitspanne von einem Jahr zwischen den beiden Untersuchungen, so ist dieser Zusammenhang beachtlich. Verglichen mit der Stabilität der Intelligenzskalen ist ein deutlicher Unterschied zu verzeichnen, der mit den durchgeführten Programmänderungen allein nicht begründbar ist. Der wesentliche Grund dafür dürfte die geringere Reliabilität der Problemlösegütemaße sein, die trotz Aggregation nur mit einer Testlänge von drei (EU) beziehungsweise 2 (WU) unabhängigen Messungen geschätzt werden konnte.

#### 4.4.5 Intelligenz und Steuerungsleistungen

Damit sind die Voraussetzungen gegeben, um den Zusammenhang zwischen BIS- und Steuerungsleistungen zu überprüfen. Die Korrelationsmuster von EU (berechnet für die WU-Teilnehmer) und WU sind sehr ähnlich (Tab. 6): Mäßig hohe, aber substantielle Korrelationen von über  $r = .30$  zwischen dem neuen PLG und BIS-K. Dies gilt in der Wiederholungsuntersuchung jedoch erwartungsgemäß, anders als in der EU, auch für die klassischen Gütemaße, das Gesamtvermögen und das Trendmaß. Dieses Ergebnis bestätigt die Interpretation der EU-Befunde.

Gleichzeitig muß festgehalten werden, daß die Steigerung der Reliabilität der Problemlösegütemaße gegenüber der Erstuntersuchung *nicht* zu einer Steigerung der Vorhersagbarkeit der Steuerungsleistungen durch BIS-K führte. Die Höhe der Korrelationen ist eher etwas geringer als in der EU.

In der WU sind auch teilweise signifikante Korrelationen der Steuerungsleistungen mit BIS-E (Einfallsreichtum) zu verzeichnen. Dieser Befund war durch Hypothesen nicht vorhergesagt und auch in der EU nicht festzustellen.

Durch Aggregation der Ergebnisse aus EU und WU konnte die Reliabilität der Messungen für Intelligenz und Problemlösegütemaße nochmals erhöht werden. Die Korrelation zwischen BIS-K und dem Gesamtgütemaß stieg auf  $r = .47$ .

Tabelle 6: Korrelationen von Problemlösegütemaßen und Teilgütekriterien aus EU und WU mit Vorwissen und BIS-Leistungen

EU:	WLT	WIS-D1	B	M	E	K	F	V	N	AI
PLC-ABC (EU)	.28**	.47**	-.01	.08	.05	.42**	.09	.06	.29**	.19
GeKap-ABC (EU)	-.04	.04	-.07	.01	.00	.11	-.03	.10	-.01	.02
Trend-ABC (EU)	.04	.22	-.02	.17	.09	.21*	.12	.20	.06	.15
WU:	WLT <sup>1</sup>	WIS-D1	B	M	E	K	F	V	N	AI
PLG-AB (WU)	.39**	.49**	.10	-.07	.23*	.34**	.18	.10	.23*	.21*
GEKap-AB (WU)	.36**	.49**	.09	.00	.20*	.31**	.16	.13	.22*	.21*
Trend-AB (WU)	.35**	.46**	.11	-.07	.22*	.32**	.15	.08	.24*	.20*
EU/WU (aggr.):	WLT <sup>1</sup>		B	M	E	K	F	V	N	AI
PLG-EU/WU	.36**	--	.07	.04	.18	.47**	.19	.10	.34**	.26*

EU: N = 127 (nur WU-Teilnehmer, für WIS-D1: N = 87); WU: N = 137;

<sup>1</sup> WLT-Ergebnisse aus der EU;

BIS- und WIS-Ergebnisse sind jeweils aus derselben Untersuchung wie die Problemlöseleistungen;

Legende der BIS-Skalen: s. Tab. 2; Legende der Problemlösegütemaße: s. Tab. 4;

Ergänzungen: AB bzw. ABC = über Durchgänge A und B (WU) bzw. A, B und C (EU) aggregierte Werte;

PLG-EU/WU = über EU und WU aggregiertes Problemlösegütemaß; WLT = Wirtschaftslehretest; WIS-D1 = Vorwissenstest Gesamtscore.

Da für die aggregierten Daten aus EU und WU auch für die BIS-Zellen ausreichend reliable Daten vorliegen sollten, lohnt auch ein Blick in die Korrelationsmatrix von PLG und jenen Zellen des BIS, die aus der Kreuzklassifikation der operativen Fähigkeit „Verarbeitungskapazität“ und den drei inhaltsgebundenen Fähigkeiten zustandekommen: Die höchste Korrelation liegt für BIS-KN (Verarbeitungskapazität mit numerischen Material) mit  $r = .49$  vor. Dieser Wert ist wegen der nicht optimalen Reliabilität der BIS-Zellen eher eine Unterschätzung des tatsächlichen Zusammenhangs.

Tabelle 7: Korrelation der Problemlöseleistungen mit den BIS-Zellen

EU/WU (aggr.):	KN	KF	KV
PLG-EU/WU	.49**	.39**	.31**

#### 4.4.6 Intelligenz und Wissen beim Problemlösen

Als Ergebnis der EU (Süß & Kersting, 1990) wurde festgestellt, daß allgemeines und spezifisches Vorwissen hoch mit den Steuerungsleistungen korrelierten und neben den Intelligenzskalen einen inkrementellen Varianzbeitrag zur Vorhersage

der Problemlösegütemaße leisteten. Diese Ergebnisse sollten sich in der Wiederholungsuntersuchung replizieren lassen. Tabelle 8 zeigt die multiplen Regression des aggregierten PLG auf die Prädiktoren. Die multiple Korrelationen „schrumpfen“ durch ihre Kreuzvalidierung nur geringfügig. Sowohl das Systemwissen als auch die Skala „Faustregeln“ (s. oben), die einen Aspekt des Handlungswissens darstellt, erhöhte die multiple Korrelation signifikant; nur das Eingriffswissen brachte keine zusätzliche Varianzaufklärung mehr. Die Befunde der EU wurden damit bestätigt. Allerdings ist auch hier deutlich, daß die Gesamtvorhersagbarkeit der Steuerungsleistungen durch unsere Prädiktoren trotz der reliableren Gütekriterien der WU gegenüber der EU nicht verbessert werden konnte.

Tabelle 8: Vorhersage der Problemlöseleistungen (PLG-AB) in der WU

Hierarchische Regressionsanalyse	r	R	R <sup>2</sup>	R <sup>2</sup> <sub>diff</sub>	Sign F <sub>diff</sub>	Kreuzvalidierung (Zufalls-splittung)	Gruppe A	Gruppe B
							N = 69	N = 68
K (WU)	.34							
M (WU)	-.07	.40	.16					
WLT	.39	.50	.25	.09	< .01	R vor Korrektur	0.70	0.50
Systemwissen	.43					R nach Korrektur	0.60	0.44
Faustregeln	.32							
(Eingriffswissen) <sup>1</sup>	.20	.59	.35	.10	< .01			

N = 137; (1) für Vorhersage redundant;  
 Regressionsgleichung: PLG-A = .20 \* K - .12 \* M + .21 \* WLT + .24 \* Systemwissen + .20 \* Faustregeln + .09 \* Eingriffswissen

Im Gegensatz zur EU zeigte sich in der WU der BIS-Faktor M (Merkfähigkeit) als Suppressor für den Zusammenhang von BIS-K und der Problemlöseleistung. In der EU hatte der Faktor B (Bearbeitungsgeschwindigkeit) diese Rolle eingenommen.

#### 4.5 Zusammenfassung der empirischen Befunde

Die vorgestellten Ergebnisse bestätigen die getroffenen Annahmen: Nicht etwa die Allgemeine Intelligenz, sondern die operative Fähigkeit „Verarbeitungskapazität“ des BIS ist der beste Prädiktor für die Steuerungsleistungen in der Schneiderwerkstatt. Die Korrelationen von  $r = .47$  mit BIS-K und von  $r = .49$  mit der Zelle KN des BIS sind auch in ihrer absoluten Höhe beachtlich.

Derart hohe Korrelationen können allerdings erst auf der höchsten Aggregationsstufe aufgezeigt werden, wenn die PLG-Werte aus EU und WU zusammengefaßt werden. Dies bestätigt einen Aspekt des Symmetriekonzepts von Wittmann (1988) für den Zusammenhang von Intelligenz und Problemlösen.

Die Steuerungsleistung eines komplexen computersimulierten Systems hängt also auch unter Intransparenz mit dem Intelligenzkonstrukt „Verarbeitungskapazität“ zusammen. Die Hypothese, daß ein solcher Zusammenhang nur unter Transparenzbedingung zu finden ist (Putz-Osterloh, 1981), ist damit entkräftet.

Weiterhin kann festgehalten werden: allgemeines und spezifisches Vorwissen liefern zusätzlich zur Intelligenz einen substantiellen Beitrag zur Vorhersage der Steuerungsleistungen. Besonders beachtenswert ist in diesem Zusammenhang, daß das allgemeine Wirtschaftslehrewissen, das nur in der Erstuntersuchung erhoben wurde, mit den ein Jahr später erhobenen Leistungen der modifizierten SWS in gleicher Höhe korreliert wie mit den Gütemaßen aus der Erstuntersuchung.

In der WU konnte außerdem das Ergebnis der Aufgabenanalyse der EU, das uns als Grundlage zur Bestimmung eines neuen Problemlösegütemaßes diente, bestätigt werden. Ohne diese Analyse wären auch unsere EU-Ergebnisse als weiterer Hinweis auf die Invalidität von Intelligenztestleistungen in die Literatur eingegangen. Offen bleibt, ob möglicherweise bereits früher publizierte Ergebnisse mit der alten Version der SWS aus demselben Grund Nullkorrelationen mit Intelligenzmaßen erzielten.

#### Literatur

- Amthauer, R. (1973). *I-S-T 70. Intelligenz-Struktur-Test*. Göttingen: Hogrefe.
- Brambring, M. (1983). Spezielle Eignungsdiagnostik. In K. J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik*. Göttingen, Toronto, Zürich: Hogrefe.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Chi, M. T. H.; Glaser, R. & Farr, M. J. (1988). *The Nature of Expertise*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Deutsche Gesellschaft für Personalwesen e. V. (1986). *Differentieller Kenntnistest, Subtest Wirtschaft*. Hannover.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290–308.
- Dörner, D. & Kreuzig, H. W. (1983). Problemlösefähigkeit und Intelligenz. *Psychologische Rundschau*, 34, 185–192.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Fishbein, M. & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59–74.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29, 283–302.
- Groffmann, K. J. (1983). Die Entwicklung der Intelligenzmessung. In K. J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik*. Göttingen: Hogrefe.
- Hasselmann, D. (1990). Der Einsatz computersimulierter, komplexer Probleme in betrieblichen Assessment Center-Veranstaltungen. In D. Frey (Hrsg.), *Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990 (S. 64–65)*. Göttingen: Hogrefe.
- Hesse, F. W. (1982). Effekte des semantischen Kontextes auf die Bearbeitung komplexer Probleme. *Zeitschrift f. Exp. & Angew. Psychologie*, 29, 62–91.
- Hörmann, H. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexem Problemlösen. *Sprache & Kognition*, 8, 23–31.

- 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 Bewährungskontrolle eines kontextvaliden konstruierten problemspezifischen Wissenstests*. Diplomarbeit im Forschungsprojektschwerpunkt „Intelligenz und Wissen“. FU Berlin, FB Erziehungs- und Unterrichtswissenschaften, Institut für Psychologie.
- Kluwe, R. H., Misiak, C. & Haider, H. (in press). The control of complex systems and performance in intelligence tests. In H. Rowe (Hrsg.), *Future perspective in intelligence research and measurement*. Hillsdale, N. J.: Erlbaum.
- Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17, 248–298.
- Krumm, V. & Seidel, G. (1970). *Wirtschaftslehretest (BWL)*. Weinheim: Beltz.
- Möbus, C. (1978). Zur Fairness psychologischer Intelligenztests: Ein unlösbares Trilemma zwischen Zielen von Gruppen, Individuen und Institutionen. *Diagnostica*, 24, 191–234.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189, 79–100.
- Putz-Osterloh, W. & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift f. Exp. & Angew. Psychologie*, 28, 309–334.
- Putz-Osterloh, W. & Schroiff, M. (1987). Komplexe Verhaltensmaße zur Erfassung von Hochbegabung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 207–216.
- Schmidt, J. U., Brocke, B., Jäger, A. O., Doll, J. & König, F. (1986). *Entwicklung eines Tests für das Berliner Intelligenzstrukturmodell (Arbeitsbericht 4 des Forschungsprojektschwerpunkts „Produktives Denken/Intelligentes Verhalten“)*. Berlin: FU Berlin, FB Erziehungs- und Unterrichtswissenschaften, Institut für Psychologie.
- Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*, pp. 225–307. Cambridge: University Press.
- Süß, H.-M. & Kersting, M. (1990). *Bedeutung und Erfassung von Wissen für das Lösen komplexer Probleme*. Referat, gehalten auf dem 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990.
- Süß, H.-M., Oberauer, K. & Jäger, A. O. (1990). *Zur prädiktiven Validität von Wissen und Intelligenz für Steuerungsleistungen an computersimulierten Systemen*. Referat, gehalten auf dem 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990.
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309–329.
- Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology*. 2nd ed. New York: Plenum, 505–560.

*Anschrift der Verfasser:*

Dr. Heinz-Martin Süß  
 Institut für Psychologie  
 Freie Universität Berlin  
 Habelschwerdter Allee 45  
 1000 Berlin 33

## Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme\*

Stefan Strohschneider

Die Debatte um die Zusammenhänge zwischen der Leistung beim Bearbeiten komplexer Probleme und der Leistung in Intelligenztests ist nach wie vor zu keinem schlüssigen Ergebnis gelangt. In dem Beitrag werden weitere empirische Befunde zu diesem Thema aus drei verschiedenen Studien berichtet, wobei zur Messung der Intelligenz jeweils der „Berliner Intelligenzstrukturmodell-Test“ verwendet wurde. Als komplexe Probleme kamen zwei Versionen der Entwicklungshelfer-Simulation MORO und das abstrakt formulierte, komplexe Kleinsystem VEKTOR zum Einsatz. Während sich in der ersten MORO-Untersuchung keinerlei Zusammenhänge zwischen Intelligenz und Problemlöseerfolg finden lassen, führt die Vorgabe konkreter Handlungsziele in der zweiten MORO-Untersuchung zu deutlich höheren gemeinsamen Varianzanteilen. Die Leistungen der Vpn beim Umgang mit dem abstrakten VEKTOR-Problem schließlich sind fast durchgängig mit ihren Intelligenztestwerten korreliert. In der Diskussion werden „einfaktorielle“ Erklärungsansätze kritisiert und die gefundenen Korrelationsunterschiede auf die mit der Vorgabe von Handlungszielen und dem Ausschluß der Nutzbarkeit von Vorwissen verbundene Reduktion der Freiheitsgrade, die die komplexen Probleme der individuellen Vorgehensweise der Vpn lassen, zurückgeführt.

The debate concerning the relations between complex problem solving and intelligence testing has not yet come to a conclusive solution. The article reports further empirical results on this topic, coming from three different studies. The Ss' intelligence was measured with the test for the "Berlin Intelligence-Structure-Model". The Ss' ability to deal with complex problems was observed in either one of two versions of the "Peace-Corps-Worker Simulation" MORO or the dynamic 8-variable system VEKTOR, which is an abstract problem, not designed to simulate any specific part of reality. While there are no correlations between intelligence and problem solving in the first MORO-study, simply giving the Ss exact goal states for several key-variables of this system in the second study leads to distinctly higher common proportions of variance. Finally, the Ss' success in controlling VEKTOR is very closely related to their intelligence. In the discussion, "one-factor-explanations" of such results are criticized. It is argued that the correlational differences have to be attributed to the presentation of exact goal states and the limitations on the use of general knowledge resulting in a reduction of the number of degrees of freedom, the complex problems allow for individual problem solving strategies to develop.

### Einleitung und Fragestellung

Die Frage nach der Vorhersagbarkeit des Erfolges von Menschen beim Umgang mit komplexen Problemen durch ihre Leistungen beim Bearbeiten von Intelligenztests war in den letzten Jahren im Bereich der deutschsprachigen Psychologie

\* Die Untersuchungen, über die hier berichtet wird, wurden von der Deutschen Forschungsgemeinschaft im Rahmen des Projektes „Systemdenken“ (Do 200/5) und des „Leibniz-Programmes“ (Prof. Dörner) gefördert. Für wertvolle Kommentare zu früheren Versionen dieser Arbeit danke ich Prof. Dr. D. Dörner und Prof. Dr. A. O. Jäger.