

---

# **Komplexes Problemlösen im Kontext angewandter Eignungsdiagnostik**

Inauguraldissertation

zur Erlangung des akademischen Grades eines Doktors der Philosophie  
der Philosophischen Fakultät der Ernst-Moritz-Arndt-Universität Greifswald

Kai Burmeister

---

Tag der Disputation

29.04.2009

Dekan

Prof. Dr. Matthias Schneider

Erstgutachter

Prof. Dr. Manfred Bornewasser

Gutachter

Prof. Dr. Hannelore Weber

PD Dr. Anette Hiemisch

Prof. Dr. Ricarda Bouncken

Dr. Marco Waage

---

# INHALT

<b>1. ZUSAMMENFASSUNG</b> .....	<b>7</b>
<b>2. EINLEITUNG</b> .....	<b>9</b>
<b>3. ANGEWANDTE EIGNUNGSDIAGNOSTIK</b> .....	<b>11</b>
3.1 Konstruktorientierte Verfahren .....	14
3.2 Simulationsorientierte Verfahren .....	20
3.3 Intelligenz, Wissen und Berufsleistung.....	23
<b>4. DAS FORSCHUNGSFELD AM BEISPIEL TAILORSHOP</b> .....	<b>29</b>
4.1 Der Tailorshop .....	29
4.1.1 Komplexität .....	33
4.1.2 Ökologische Validität .....	34
4.2 Die Studien von Putz-Osterloh und Lürer (1981) .....	36
4.3 Die Studie von Funke (1983) .....	39
4.4 Die Studie von Süß, Kersting und Oberauer (1991).....	41
4.5 Die Arbeit von Hasselmann (1993) .....	44
4.6 Die Studie von Wittmann, Süß und Oberauer (1996).....	47
4.7 Die Arbeit von Kersting (1999) .....	49
<b>5. RELIABILITÄT KOMPLEXER SZENARIEN</b> .....	<b>51</b>
5.1 Messung der Reliabilität.....	51
5.1.1 Retest-Reliabilität.....	51
5.1.2 Paralleltest-Reliabilität .....	52
5.1.3 Testhalbierung und Interne Konsistenz.....	53
5.2 Das Primat der Validität .....	54
5.3 Optimierung der Reliabilität.....	54
<b>6. INTELLIGENZ, WISSEN UND PROBLEMLÖSEN</b> .....	<b>57</b>
6.1 Der Zusammenhang mit Intelligenz.....	57
6.1.1 Strukturmodelle der Intelligenz .....	57
6.1.2 Verarbeitungskapazität.....	60
6.1.3 Arbeitsgedächtniskapazität.....	60
6.1.4 Der Zusammenhang von Intelligenz und Problemlöseleistung .....	61
6.1.5 Empirische Studien.....	64
6.1.6 Auswertung der Studien .....	69
6.2 Der Zusammenhang mit Wissen.....	74
6.2.1 Vorwissen und erworbenes Wissen .....	74
6.2.2 Implizites Wissen .....	76
<b>7. DER ZUSAMMENHANG MIT BERUFSLEISTUNG</b> .....	<b>79</b>
<b>8. INTEGRATION DER BEFUNDE UND ABLEITUNG VON HYPOTHESEN</b> .....	<b>83</b>

---

<b>9. DIE UNTERSUCHUNG</b> .....	<b>87</b>
9.1 Stichproben.....	87
9.2 Das Szenario AGRIMAN.....	88
9.2.1 Inhaltliche Einkleidung und Oberfläche.....	89
9.2.2 Wissensdiagnostik.....	92
9.2.3 Einordnung in die TAKS.....	94
9.2.4 Parameter in dieser Untersuchung.....	98
9.3 Weitere Instrumente.....	99
9.3.1 Planspielbasierte Gruppendiskussion.....	100
9.3.2 Fallstudienbasierte Gruppendiskussion.....	101
9.3.3 Interview.....	101
9.3.4 Rollenspiele.....	101
9.3.5 Intelligenztests.....	101
9.3.6 Persönlichkeitsfragebogen.....	102
9.4 Statistische Hypothesen.....	102
<b>10. AUSWERTUNG</b> .....	<b>105</b>
10.1 Deskriptive Auswertung.....	105
10.1.1 Diagnosephase I.....	105
10.1.2 Diagnosephase II.....	107
10.1.3 Handlungswissenstest.....	108
10.1.4 Deklarativer Wissenstest.....	110
10.2 Zusammenhang mit Wissen und Intelligenz.....	112
10.2.1 Zusammenhang mit Wissen.....	112
10.2.2 Zusammenhang mit Intelligenz.....	113
10.2.3 Zusammenhang mit Wissen und Intelligenz.....	114
10.3 Zusammenhang mit Motivation und Interesse.....	117
10.4 Zusammenhang mit Berufsleistung.....	121
10.4.1 Gesamtergebnis.....	121
10.4.2 Planspielbasierte Gruppendiskussion.....	123
10.4.2 Fallstudienbasierte Gruppendiskussion.....	125
10.4.3 Interview.....	127
10.4.4 Rollenspiele.....	128
10.4.5 Im Vergleich mit dem Intelligenztest.....	129
<b>11. DISKUSSION</b> .....	<b>133</b>
11.1 Sichtung der Ergebnisse.....	133
11.2 Bewertung der Ergebnisse.....	141
<b>12. LITERATUR</b> .....	<b>147</b>

---

## **DANKSAGUNG**

Joachim Stempfle und Oliver Hübner für die Hilfe bei der Entwicklung von AGRIMAN und dafür, dass diese Arbeit möglich wurde.

Anja König und Grete Baumann für die Hilfe bei der Organisation der Assessment Center Trainings.

Meinen Betreuern Prof. Dr. Manfred Bornwasser und Prof. Dr. Hannelore Weber für die unkomplizierte und schnelle Unterstützung.

---

---

# 1. ZUSAMMENFASSUNG

Trotz nachlassenden Interesses in der wissenschaftlichen Forschung erfreuen sich komplexe Problemlöseszenarien als sozial valider Ersatz für klassische Intelligenztests in der angewandten Eignungsdiagnostik ungebrochener Beliebtheit. Ziel der vorliegenden Arbeit ist die umfassende Validierung des für den Praxiseinsatz neu entwickelten komplexen Problemlöseszenarios AGRIMAN.

In einem theoretischen Teil wird das Forschungsfeld und seine Geschichte am Beispiel des bekanntesten Problemlöseszenarios TAILORSHOP erläutert sowie auf den Zusammenhang von komplexer Problemlöseleistung mit Intelligenz, Wissen und beruflicher Leistung eingegangen. Im empirischen Teil wird die Entwicklung und Validierung von AGRIMAN beschrieben. N=185 Probanden aus drei Stichproben bearbeiteten im Rahmen von neun Assessment Centern das neu entwickelte Szenario.

Die Ergebnisse bestätigen das Arbeitsmodell, dass Verarbeitungskapazität vermittelt über das erworbene systemspezifische Wissen die Problemlöseleistung beeinflusst. Außerdem ergaben sich mittelhohe Zusammenhänge zwischen Problemlöseleistung und über Assessment Center operationalisierter Berufsleistung.





---

## 2. EINLEITUNG

„Arbeiten zum komplexen Problemlösen folgen häufig einem Eröffnungsritual, in dem die wachsende Komplexität der Moderne und die damit einhergehenden neuen Anforderungen an ein systemisches Denken beschwörend den überschaubaren alten Zeiten gegenübergestellt werden.“ (Kersting, 1999). Diese Arbeit soll keine Ausnahme machen. Die Welt, Nationen, Städte, Unternehmen sind komplexe Systeme, deren Steuerung die Verantwortlichen an die Grenzen ihrer Leistungsfähigkeit bringt – und, wie leider immer wieder festzustellen ist, oft auch darüber hinaus.

Hier setzte die Forschung zum komplexen Problemlösen an. Computersimulationen sollten es ermöglichen, komplexe Realitätsbereiche wie etwa eine Kleinstadt oder ein Entwicklungsland im Labor „nachzubauen“, um das Handeln – und Scheitern – von Menschen bei solchen Aufgaben untersuchen und erklären zu können. Dabei standen zunächst kognitive Prozesse der Handelnden, insbesondere ihre Denkfehler, im Vordergrund.

Ein weiteres Forschungsfeld öffnete sich, als frühe Studien keinen systematischen Zusammenhang zwischen der Leistung bei der Steuerung solcher Simulationen und der Leistung in klassischen Intelligenztests zeigten. Intelligenztests wurden daraufhin als ungeeignet zur Vorhersage von Leistung in komplexen Situationen abgelehnt. Dörner (1986) stellte der klassischen Intelligenz die „operative Intelligenz“ als neues Konstrukt gegenüber. Diese sollte nicht nur Geschwindigkeit und Genauigkeit kognitiver Operationen umfassen, sondern auch Prozesse wie Umsicht im Sinne der Antizipation von Neben- und Fernwirkungen, Steuerungsfähigkeit der kognitiven Operationen und Verfügen über Heuristiken, für Dörner die „operativen“ Merkmale der Intelligenz. Komplexe Computersimulationen wurden als die Messinstrumente dieses neuen Konstrukts angesehen.

Diese Entwicklung betraf auch die angewandte Eignungsdiagnostik. Der Steuerung komplexer Systeme wird insbesondere im Managementbereich, aber auch in vielen anderen Berufen große Bedeutung beigemessen. Die ökonomische, reliable und valide Messung der dafür erforderlichen Fähigkeiten rückt in den Mittelpunkt des Interesses.

Nachdem neuere Studien konsistent substanzielle Zusammenhänge zwischen Intelligenz und der Leistung in komplexen Problemlösesimulationen nachweisen und die Existenz der „opera-

---

tiven Intelligenz“ damit massiv in Frage stellen, hat das Interesse an diesem Forschungsgebiet deutlich nachgelassen.

Demgegenüber steht eine ungebrochene Nachfrage nach simulationsbasierter Diagnostik in der Praxis, die von kommerziellen Anbietern gedeckt wird. Dies liegt daran, dass Intelligenztests trotz anerkannt hoher prädiktiver Validität aufgrund ihrer sehr geringen Akzeptanz in der Praxis kaum angewendet werden (können). Zwar gibt es Bereiche, in denen Intelligenztests flächendeckend eingesetzt werden, etwa in der Auswahl für den Polizeidienst (Neick, 2008). Insgesamt werden sie in Deutschland jedoch nur bei 35% der Auszubildenden, 9% der Trainees und 2% der Führungskräfte zur Auswahl eingesetzt (Schuler, 2005). Komplexe Problemlöseszenarien dagegen ermöglichen die sozial valide Diagnose kognitiver Fertigkeiten und stellen so wertvolle Bausteine in Assessment Centern dar.

Es besteht deshalb ein Bedarf nach in der Praxis sauber validierten komplexen Problemlöseszenarien. Leider fehlen zu den meisten kommerziell eingesetzten Simulationen wissenschaftliche Studien, die den Zusammenhang dieser Instrumente mit Intelligenz, Wissen und beruflicher Leistung untersuchen. Diese Lücke soll die vorliegende Untersuchung schließen.

Die Arbeit gliedert sich in zwei Teile, einen theoretischen und einen empirischen. Im theoretischen Teil werden komplexe Problemlöseszenarien zunächst in den Kontext angewandter Eignungsdiagnostik eingeordnet um dann das Forschungsfeld und seine Geschichte am Beispiel des bekanntesten Problemlöseszenarios – dem TAILORSHOP – zu erläutern. Daraufhin wird in eigenen Kapiteln auf die Reliabilität von Problemlöseszenarien sowie auf den Zusammenhang der Problemlöseleistung mit Intelligenz, Wissen und beruflichen Leistungsmaßen eingegangen.

Der empirische Teil beschreibt die umfassende Validierung des für den Praxiseinsatz neu entwickelten komplexen Problemlöseszenarios AGRIMAN. Er wird eingeleitet durch eine Beschreibung der Untersuchungsmethodik, in deren Rahmen auch auf das Szenario selbst und die Ziele bei seiner Entwicklung eingegangen wird. In der Auswertung der Studie wird auf deskriptive Daten, den Zusammenhang mit Intelligenz und Wissen sowie schließlich auf den Zusammenhang mit in Assessment Centern erhobenen beruflicher Leistung eingegangen.

---

### 3. ANGEWANDTE EIGNUNGSDIAGNOSTIK

Das komplexe Problemlöseszenario AGRIMAN wurde explizit zum Einsatz in der angewandten Eignungsdiagnostik entwickelt. Um eine umfassende Bewertung des Instruments zu ermöglichen, muss es in den breiteren Kontext etablierter eignungsdiagnostischer Instrumente eingeordnet werden. Schuler (2005) gibt den passenden Rahmen vor, indem er drei Ansätze der Eignungsdiagnostik unterscheidet: Den *Eigenschaftsansatz*, den *Simulationsansatz* und den *biografischen Ansatz*. Die Ansätze unterscheiden sich in der Methode der Merkmalerfassung und der Validierungslogik.

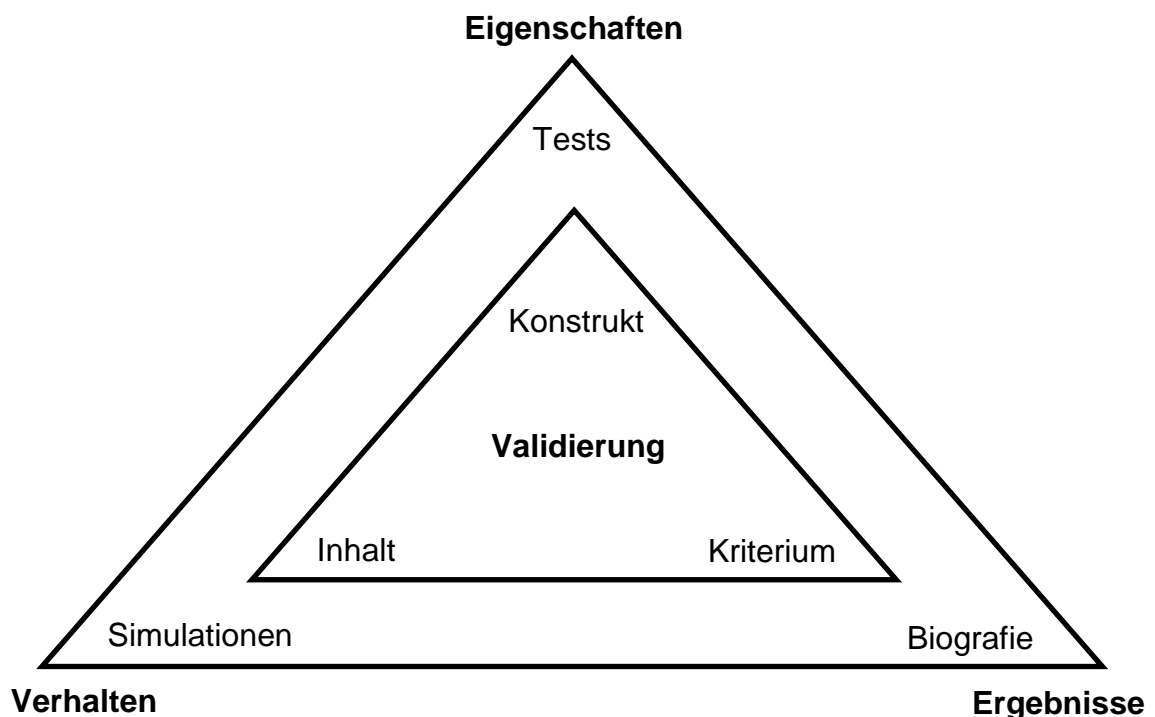


Abbildung 3.1. Die Ansätze der Eignungsdiagnostik und ihre Validierungslogik (Schuler, 2005)

Mit dem Eigenschafts- oder Konstruktansatz werden Merkmale erfasst, die als relativ stabil angenommen werden, vor allem kognitive Leistungsfähigkeit und Persönlichkeitseigenschaften. Dies geschieht über psychologische Testverfahren. Kriterium, ob die Erfassung des Konstrukts gelungen ist, ist die *Konstruktvalidität*.

---

Mit dem Simulationsansatz soll Verhalten erfasst werden, das in ähnlicher Form auch am Arbeitsplatz gefordert ist. Kriterium, ob dies gelungen ist, ist die inhaltliche Repräsentativität oder *Kontentvalidität*.

Der Biografische Ansatz basiert auf dem Prinzip, dass der beste Prädiktor zukünftigen Verhaltens einer Person das vergangene Verhalten dieser Person ist. Biografische Fragen sollen vergangene Leistungen erfassen, um zukünftige Leistung vorherzusagen. Kriterium, ob dies gelungen ist, ist die *prognostische Validität*.

Letztlich soll natürlich mit allen drei Ansätzen zukünftige Leistung vorhergesagt werden, das letztgültige Kriterium ist immer die zukünftige Berufsleistung. Ein konstruktvalider psychologischer Test hilft wenig, wenn das Konstrukt mit Berufserfolg nicht zusammenhängt. Und auch eine kontentvalide Simulation muss sich letztlich daran messen lassen, ob sie reales Verhalten vorhersagen kann.

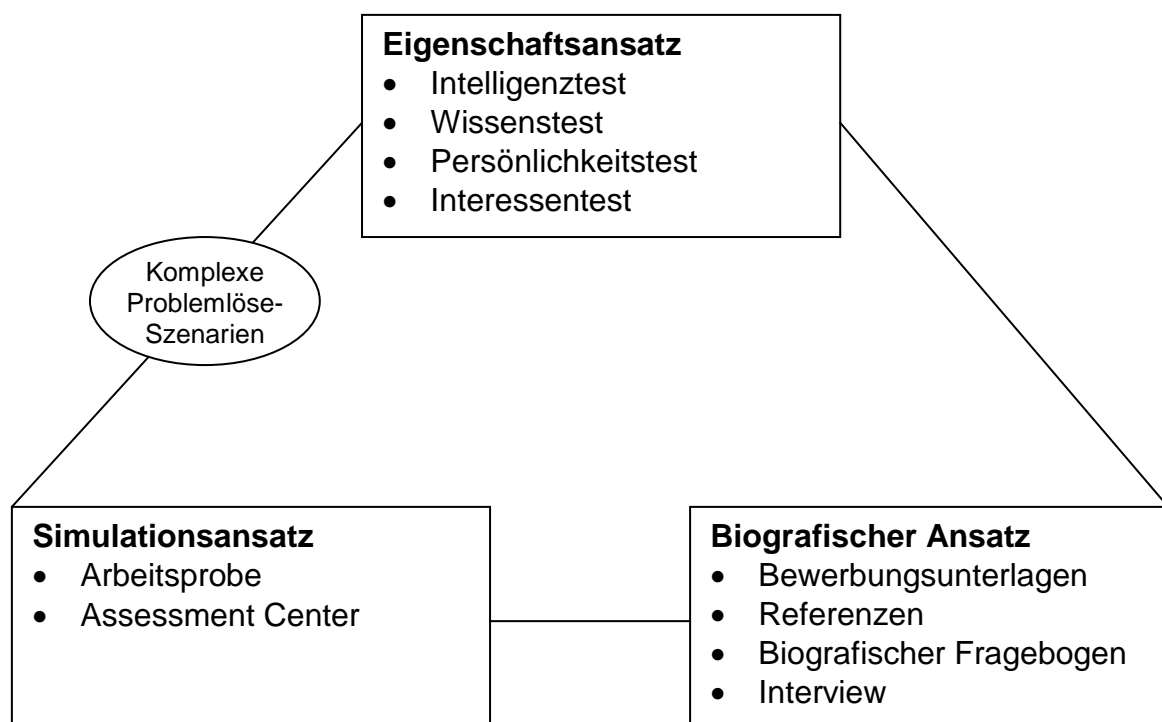


Abbildung 3.2. Etablierte Verfahren der Eignungsdiagnostik

Die Zuordnung der verschiedenen Verfahren der Eignungsdiagnostik zu den drei Ansätzen in Abbildung 3.2 macht die besondere Stellung komplexer Problemlöseszenarien deutlich. Von

---

Schuler (2005) werden sie dem Simulationsansatz zugeordnet, tatsächlich nehmen Sie eher eine Zwischenposition zwischen Simulations- und Eigenschaftsansatz ein.

Dem Simulationsansatz können sie zugeordnet werden, da sie ursprünglich entwickelt wurden, um einen komplexen Realitätsabschnitt zu modellieren, dadurch realistisches Verhalten hervorzurufen und dieses zu bewerten. Sie können jedoch auch dem Eigenschaftsansatz zugeordnet werden, da die Problemlöseleistung immer wieder mit einem von Intelligenz verschiedenen Konstrukt in Verbindung gebracht wurde, der komplexen Problemlösefähigkeit oder „operativen Intelligenz“. Diese doppelte Zuordnung spiegelt sich auch darin wieder, dass die Validierungslogiken beider Ansätze in Forschungsarbeiten zum komplexen Problemlösen thematisiert werden. Insbesondere frühe Arbeiten von Dörner und Putz-Osterloh betonen die Ökologische Validität bzw. Kontentvalidität der Verfahren. Dies wird in Abschnitt 4.1.2 kritisch beleuchtet. Spätere Arbeiten von Süß, Kersting und Oberauer (1991), Hasselmann (1993), Wittmann, Süß und Oberauer (1996) und Kersting (1999) thematisieren dagegen die mangelnde Konstruktvalidität komplexer Problemlöseszenarien. Darauf wird in den Abschnitten 4.4 bis 4.7 näher eingegangen.

In den folgenden Abschnitten soll auf den Eigenschaftsansatz und den Simulationsansatz sowie die zugehörigen Instrumente näher eingegangen werden. Tabelle 3.1 zeigt eine Gegenüberstellung beider Ansätze, in den Abschnitten 3.1 und 3.2 werden sie dann ausführlich vorgestellt. Dies geschieht zum einen, weil komplexe Problemlöseszenarien beiden Ansätzen zugerechnet werden können. Zum anderen wird in der Untersuchung zu AGRIMAN vor allem der Zusammenhang mit Instrumenten des Eigenschafts- und Simulationsansatzes untersucht. Die Sichtung empirischer Ergebnisse zur Validität gängiger eignungsdiagnostischer Instrumente hilft zudem bei der Bewertung der Ergebnisse zur Validität von AGRIMAN.

In Abschnitt 3.3 sollen die verschiedenen Instrumente im Rahmen eines integrativen Ansatzes zur Vorhersage von Berufsleistung in Beziehung zueinander gesetzt werden.

Komplexe Problemlöseszenarien bleiben in diesem Kapitel zunächst ausgespart, da auf sie in den Kapiteln 4 bis 7 ausführlich eingegangen werden soll.

Tabelle 3.1. Eigenschafts- und Simulationsorientiertes Verfahrenskonzept (Schuler, 2005)

	<b>Eigenschaftsorientiert</b>	<b>Simulationsorientiert</b>
<b>Bezug zum Arbeitsplatz</b>	Geringer. Eigenschaftsbezogene Items (Signs)	Höher. Tätigkeitsbezogene Aufgabe (Sample)
<b>Anforderungsbereich</b>	Eigenschaftshomogen	Eigenschaftsheterogen
<b>Ablaufstruktur</b>	Elementaristisch: Einzelitems	Ganzheitlich: Gesamtaufgabe mit Planungsanforderungen
<b>Rückmeldung &amp; Kontrolle</b>	Gering	Ablauf-Feedback Ergebnis-Feedback
<b>Verlaufscharakter</b>	Statisch	Dynamisch: Reaktionsabhängige Aufgabenentwicklung

### 3.1 Konstruktorientierte Verfahren

*Die klassische Methode, psychologische Konstrukte zu messen, ist der Test. Psychologische Tests stellen gewissermaßen den Prototyp konstruktorientierter Diagnoseverfahren dar. Sie können mit Brandstätter (1979, S.82) als standardisierte, routinemäßig anwendbare Verfahren zur Messung individueller Verhaltensmerkmale aufgefasst werden, aus denen Schlüsse auf Eigenschaften der betreffenden Person oder auf ihr Verhalten in anderen Situationen gezogen werden können. (Schuler, 2005).*

Die bekanntesten Testverfahren sind Intelligenztests und Persönlichkeitstests, daneben kommen aber auch Tests spezifischer kognitiver Fähigkeiten sowie Motivations- und Interessentests zum Einsatz. In der angewandten Eignungsdiagnostik wird nicht nur auf publizierte und über den Testhandel erhältliche Verfahren zurückgegriffen. Vielfach kommen Tests kommerzieller Anbieter zum Einsatz.

#### Intelligenztests

„Gemeinsamer Kern praktisch aller Intelligenzkonzepte ist die Qualität und Geschwindigkeit der Lösung neuer (also nicht routinebestimmter) Aufgaben.“ (Schuler, 2005). Diese Aussage macht den breiten Gültigkeitsanspruch des Intelligenzkonstrukts deutlich. Wird Intelligenz

---

diesem Anspruch gerecht, sollte es sich zur Vorhersage von Berufsleistung eignen – mit neuen Aufgaben werden die meisten Berufstätigen regelmäßig konfrontiert. In diesem Abschnitt soll zunächst die eignungsdiagnostische Bewährung von Verfahren zur Messung von Intelligenz untersucht werden, auf die theoretische Fundierung des Intelligenzkonstrukts und seiner Facetten wird dann in Abschnitt 6.1 eingegangen.

Bis 1960 wurden kognitive Verfahren in großer Häufigkeit eingesetzt (Schuler, 2005). Erste systematische Studien zeigten dann allerdings eine als unbefriedigend angesehene Validität zwischen .20 und .30 (Ghiselli, 1973). Dies ging einher mit genereller Kritik an eigenschaftsorientierten Verfahren und führte zu einer Umorientierung hin zu simulationsorientierten Verfahren - vor allem Assessment Centern.

Dieses Bild wandelte sich wieder durch die Arbeit der Schmidt & Hunter-Gruppe. Schmidt und Hunter reanalysierten die vorliegenden Einzelstudien unter Einschluss der Datenbasis von Ghiselli (1973) mit neueren metaanalytischen Techniken. Schmidt & Hunter (1998) zitieren die Arbeit von Hunter (1980), die auf eine prognostische Validität allgemeiner kognitiver Fähigkeitstests von .51 kommt. Diesen Ergebnissen und deren aggressiven Vermarktung durch die Schmidt & Hunter-Gruppe ist zu verdanken, dass Intelligenztests heute als validester Einzelprädiktor von Berufsleistung gelten.

Die Ergebnisse müssen allerdings kritisch hinterfragt werden. Ein zentraler Bestandteil der Methodik der Schmidt & Hunter-Gruppe ist die Verwendung statistisch korrigierter Validitätswerte. So werden Minderungskorrekturen zum Ausgleich der mangelnden Reliabilität des Kriteriums eingesetzt sowie eine statistische Korrektur der Varianzeinschränkung durch Vorselektion der Stichprobe vorgenommen. Zwei Zitate aus der Metaanalyse von Hunter & Hunter sollen das verdeutlichen (Hunter & Hunter, 1984, S. 79):

- *„Correction for error of measurement in job performance was done using the upper bound interrater reliability of .60 for ratings (King, Hunter & Schmidt, 1980).”*
- *„Except for ability, correlations are not corrected for restriction in range. Empirical values for range restriction are known only for ability tests, and most of the alternative predictors considered here are relatively uncorrelated with cognitive ability.”*

Zum einen wird also eine geringe Reliabilität des Kriteriums von nur .60 angenommen. Die empirische Validität wird dadurch in der Minderungskorrektur deutlich aufgewertet. Schwerer wiegt, dass die Varianzeinschränkung offenbar nur bei Tests allgemeiner kognitiver Fähigkei-

---

ten statistisch korrigiert wird. Auf dieser Basis die Überlegenheit solcher Tests festzustellen, erscheint zumindest fragwürdig.

Leider werden die unkorrigierten Validitäten in den Metaanalysen der Schmidt & Hunter-Gruppe nicht genannt. Süß (2006) zitiert allerdings verschiedene alternative Metaanalysen:

*„Zum Vergleich: Schmitt, Gooding, Noe & Kirsch (1984) kamen in ihrer Metaanalyse auf einen unkorrigierten Wert von .25, Bobko und Roth (1999) auf .30 (vgl. Cook, 2004). Salgado, Anderson, Moscoso, Bertua, de Fruyt & Rolland (2003) errechneten in einer Metaanalyse auf der Grundlage von 93 europäischen Primärstudien einen mittleren empirischen Validitätskoeffizienten von .29, korrigiert nach den Methoden von Schmidt und Hunter .62.“* (Süß, 2006).

Die empirisch ermittelten Validitäten schwanken also eher um .30, eine statistische Aufwertung auf über .50 oder gar über .60 wie bei Salgado et al (2003) erscheint zumindest zweifelhaft. Auf jeden Fall gilt es das zu berücksichtigen, wenn wie in dieser Arbeit unkorrigierte Validitäten anderer Instrumente im Vergleich beurteilt werden sollen.

Wie die folgenden Abschnitte zeigen werden, gehören Intelligenztests mit einer empirischen Validität um .30 trotzdem zu den validesten Prädiktoren von Berufsleistung.

## **Wissenstests**

*„Fachkenntnisse, verstanden als Wissen über Fakten, Abläufe und Zusammenhänge innerhalb der eigenen Berufstätigkeit, gehören zu den wichtigsten Voraussetzungen beruflichen Erfolgs und deshalb auch zu seinen validesten Prädiktoren.“* (Schuler, 2005).

Diese Aussage dürften die meisten Personalverantwortlichen unterschreiben. Die Fachkenntnisse eines Bewerbers spielen deshalb eine wichtige Rolle in der angewandten Eignungsdiagnostik. Meist wird versucht, sie in Interviews zu erfragen. Ökonomischer und reliabler ist die Erfassung durch auf den Beruf zugeschnittene Wissenstests, die meist mit Multiple-Choice-Aufgabenformat arbeiten.

Dye, Reck & McDaniel (1993) haben die Validität von Wissenstests untersucht. In ihre Metaanalyse flossen 502 Korrelationskoeffizienten und eine Gesamtstichprobe von 363.528 Personen ein, wobei allerdings nur 164 Korrelationskoeffizienten und eine Stichprobe von 19.760 Personen den Zusammenhang mit Berufsleistung betrafen - der Rest untersuchte den Zusammenhang mit Trainingserfolg. Interessant ist vor allem der gefundene Moderatoreffekt:



---

Die unkorrigierte Validität fiel nur für auf den spezifischen Job zugeschnittene Wissenstests hoch aus, hier betrug der Zusammenhang mit Berufsleistung .31, ermittelt aus 59 Korrelationskoeffizienten mit einer Gesamtstichprobe von 3.965 Personen. Bei nicht auf den spezifischen Jobs zugeschnittenen Wissenstests fiel die unkorrigierte Validität dagegen vergleichsweise gering aus:  $r=.17$  für zumindest verwandte Tests,  $r=.16$  für Tests, die nichts mit dem Job zu tun hatten.

Auf den Job zugeschnittene Wissenstests sind also ähnlich valide wie Intelligenztests. Dies zeigt sich auch, wenn man die nach den Methoden der Schmidt & Hunter-Gruppe aufgewerteten Validitätskoeffizienten betrachtet. Schmidt & Hunter (1998) zitieren die Metaanalyse von Hunter & Hunter (1984), die eine korrigierte Validität von .48 ermittelt, Dye, Reck und McDaniel (1993) kommen auf eine korrigierte Validität von .45 für alle Wissenstests und sogar .62 für auf den spezifischen Job zugeschnittene Wissenstests.

### **Persönlichkeitstests**

Der Einsatz von Persönlichkeitstests in der Eignungsdiagnostik ist umstritten. Er geht von der Annahme aus, dass es stabile Persönlichkeitsmerkmale gibt, die entscheidend für beruflichen Erfolg sind. Diese Persönlichkeitsmerkmale sollen durch psychologische Testverfahren gemessen werden, die meist auf Items mit Likert-Skalenformat basieren.

Das aktuelle Modell der Persönlichkeit geht von fünf abstrakten, voneinander unabhängigen Persönlichkeitsdimensionen aus, den sogenannten „Big Five“. Diese Faktoren ließen sich in zahlreichen ganz unterschiedlichen Studien nachweisen, darunter lexikalische Ansätze, faktorenanalytische Reanalysen und interkulturelle Vergleichsstudien (Schuler, 2005).

---

Tabelle 3.2. Charakterisierung der „Big Five“

Extraversion	gesellig, gesprächig, großzügig, bestimmt, dominant, aktiv, impulsiv
Neurotizismus	ängstlich, deprimiert, verlegen, emotional, leicht verärgert, besorgt, unsicher
Verträglichkeit	freundlich, höflich, flexibel, vertrauensvoll, kooperativ, tolerant, versöhnlich, weichherzig
Gewissenhaftigkeit	verlässlich, sorgfältig, verantwortungsbewusst, planvoll, organisiert, leistungsorientiert, ausdauernd
Offenheit	einfallsreich, kultiviert, originell, vielseitig, intellektuell, aufgeschlossen, ästhetikbetont

In den meisten Persönlichkeitsverfahren werden diese abstrakten Dimensionen in konkretere Facetten aufgespalten. Über die Anzahl und Auswahl dieser Facetten besteht allerdings kein Konsens.

Folgt man der Konzeption von Persönlichkeitsdimensionen als stabil, erscheint die Annahme ihrer Validität nachvollziehbar: Welcher Chef wünscht sich nicht verlässliche, sorgfältige, verantwortungsbewusste, planvolle, organisierte, leistungsorientierte und ausdauernde, mit einem Wort gewissenhafte Mitarbeiter?

Empirisch untersucht wurde diese Annahme vor allem von Barrick, Mount und Judge (2001). Ihre Metaanalyse zweiter Ordnung basiert auf 11 Metaanalysen zur Validität der Big Five. Bei der Interpretation ihrer Ergebnisse ist wie auch bei Intelligenztests Vorsicht geboten. Barrick, Mount und Judge setzen eine doppelte Minderungskorrektur der Validität ein und korrigieren die angenommene Varianzeinschränkung statistisch. Insbesondere die doppelte Minderungskorrektur nicht nur der mangelnden Reliabilität des Kriteriums sondern auch des Prädiktors erscheint dabei nicht nachvollziehbar. Natürlich messen Persönlichkeitstests nicht perfekt, dies lässt sich aber schlicht nicht verhindern. Tabelle 3.3 zeigt die unkorrigierten und korrigierten Validitäten.

---

Tabelle 3.3. Unkorrigierte und korrigierte Validität der Big Five nach Barrick, Mount und Judge (2001)

Faktor	Empirische Validität	Korrigierte Validität
Extraversion	.08	.15
Neurotizismus	-.09	-.15
Verträglichkeit	.06	.11
Gewissenhaftigkeit	.12	.24
Offenheit	.03	.07

Als validester Prädiktor ergibt sich die Gewissenhaftigkeit. Die empirische Validität von .12 ist aber im Vergleich zu kognitiven Leistungstests gering. Zitiert wird häufig der korrigierte Wert von .24, manchmal auch der korrigierte Wert .27, der sich ergibt, wenn nur unabhängige Metaanalysen ausgewertet werden.

Ein möglicher Grund für die geringe Validität von Persönlichkeitstests zur Eignungsdiagnostik liegt sicher in der Fälschbarkeit der Daten. Anders als bei Leistungstests können sich Bewerber hier problemlos besser darstellen als sie eigentlich sind.

### **Weitere konstruktorientierte Verfahren**

Neben klassischen Intelligenztests, Wissenstests und Persönlichkeitstests gibt es noch einige weitere Verfahren, die zum Eigenschaftsansatz gezählt werden können.

Dazu zählen kognitive Leistungstests, die anders eingekleidet sind als Intelligenztests. Der von Schuler und Klingner (2005) entwickelte AZUBI-BK ist dafür ein Beispiel. Er basiert auf Aufgaben, die sich dem Berufsalltag in Büro- und kaufmännischen Tätigkeiten zuordnen lassen, etwa dem Vornehmen von Kategorisierungen oder der Vervollständigung eines Faxes. Dass der Test Intelligenz erfasst zeigt die Korrelation von  $r=.77$  mit einem Auszug des BIS-4 (Schuler, 2005). Trotzdem wird er von Bewerbern als deutlich informativer, transparenter und kontrollierbarer erlebt und besser akzeptiert als klassische Intelligenztests (Klingner & Schuler, 2004). Dies ist ein weiterer interessanter Ansatz, die Probleme mit der geringen sozialen Validität von klassischen Intelligenztests zu umgehen.

---

Neben Tests, die Teilkonstrukte allgemeiner Intelligenz erfassen, müssen noch Verfahren zur Messung der Aufmerksamkeit und Konzentration, etwa d2 oder FAKT, erwähnt werden.

Bei den nichtkognitiven Merkmalen schließlich sind vor allem Verfahren zur Messung von Motivation und Interessen erwähnenswert. Gerade Interessen werden von Praktikern aber eher überschätzt. Mögen sie bei der Wahl des Berufes eine gewichtige Rolle spielen, sind Sie zur Personalauswahl wenig geeignet: Schmidt & Hunter (1998) zitieren die Metaanalyse von Hunter & Hunter (1984), die nur eine korrigierte Validität von .10 errechnen. Einschränkend muss dazu allerdings angemerkt werden, dass diese Validität nur auf drei Studien mit 1.789 Teilnehmern basiert. Auch eine weitere von Hunter & Hunter (1984) zitierte Metaanalyse militärischer Studien kommt bei 15 Korrelationen allerdings nur auf eine mittlere Validität von .13.

## **3.2 Simulationsorientierte Verfahren**

Simulationsorientierte Verfahren sollen berufliche Aufgaben simulieren, um so direkt die Leistung in konkreten beruflichen Anforderungssituationen messen zu können. Dabei soll die Simulation die spätere Aufgabe möglichst genau und repräsentativ abbilden. Die prognostische Validität resultiert aus dieser Übereinstimmung.

### **Arbeitsprobe**

Bei der Arbeitsprobe wird tatsächliches berufliches Verhalten 1:1 abgebildet, um ohne Umweg über Konstrukte direkt die spätere Arbeitsleistung vorherzusagen. Im Rahmen einer Anforderungsanalyse werden zunächst die Berufsaufgaben genau beschrieben, um dann eine möglichst repräsentative Auswahl zu treffen und sie die Bewerber in einem standardisierten Verfahren bearbeiten zu lassen. Eine solche 1:1-Abbildung ist natürlich vor allem bei einfacheren, (psycho-)motorischen Aufgaben möglich. Komplexe Berufsbilder lassen sich dagegen nur schwer repräsentativ in einer Arbeitsprobe darstellen. Bedingung der meisten Arbeitsproben ist weiterhin, dass der Bewerber schon Vorerfahrung in der durchzuführenden Aufgabe besitzt.

---

Sind diese Kriterien erfüllt, lassen sich mit Arbeitsproben hohe Validitäten erzielen. Schmidt & Hunter (1998) zitieren die Metaanalyse von Hunter & Hunter (1984) mit einer Validität von .54, die sogar die von Intelligenztests noch übersteigt. Auch hier muss natürlich noch einmal betont werden, dass es sich nicht um die empirisch ermittelte Validität handelt, sondern um einen korrigierten Wert.

### **Assessment Center**

In einem Assessment Center werden...

- ...mehrere Teilnehmer...
- ...von mehreren trainierten Beobachtern...
- ... in simulierten Praxissituationen...
- ...nach vordefinierten Anforderungskriterien...
- ...beobachtet und bewertet.

Wie Arbeitsproben sollen Assessment Center die Anforderungen einer zukünftigen Tätigkeit möglichst repräsentativ simulieren. Eine saubere Anforderungsanalyse ist deshalb sehr wichtig. Diese soll nicht nur die Tätigkeit möglichst genau beschreiben, sondern dient auch der Erstellung eines Anforderungsprofils aus Kriterien, die für eine erfolgreiche Ausübung der Tätigkeit erforderlich sind. Auf dieser Basis werden Übungen erstellt, die für die Tätigkeit erfolgsrelevante Aufgaben abbilden und dazu dienen, die Anforderungskriterien zu messen. In der angewandten Eignungsdiagnostik läuft das meist auf eine Kombination prototypischer Assessment Center Bausteine hinaus. Typisch sind Gruppendiskussionen, Präsentationen, Postkorbaufgaben sowie Rollenspiele, die Mitarbeitergespräche und Konfliktgespräche simulieren. Oft werden diese Simulationen auch mit konstruktorientierten Testverfahren sowie biografischen Interviews kombiniert.

Das Verhalten der Teilnehmer in den simulierten Praxissituationen wird von Beobachtern erfasst und auf Basis der Kriterien des Anforderungsprofils bewertet. Die Beobachter rekrutieren sich meist aus Linienvorgesetzten zwei Stufen über der Zielposition. Es wird nicht in jeder Übung jedes Kriterium beobachtet, sondern ein vorher definierter Ausschnitt des Anforderungsprofils.

---

Der Arbeitskreis Assessment Center e.V. (2004) hat neun Qualitätsstandards für Assessment Center formuliert, die auch für diese Arbeit relevant sind und deshalb zitiert werden:

1. Vor der Entwicklung und Durchführung eines AC sind die Ziele und die Rahmenbedingungen des Auftrages sowie die Konsequenzen für die Teilnehmer verbindlich zu klären und zu kommunizieren.
2. Eignungsbeurteilung lässt sich nur mit einer exakten Analyse der konkreten Anforderungen sinnvoll gestalten.
3. Ein Assessment Center besteht aus Arbeitssimulationen. Um eine Prognose über die Eignung eines Bewerbers auf eine bestimmte Zielfunktion treffen zu können, müssen Aufgaben und Arbeitssituationen so realistisch wie möglich nachgestellt werden.
4. Grundlage für die Eignungsdiagnose ist eine systematische Verhaltensbeobachtung. Um zuverlässige und gültige Diagnosen zu gewährleisten, ist der Einsatz eines anforderungsbezogenen Beobachtungssystems zwingend notwendig.
5. Gut vorbereitete Beobachter, die das Unternehmen angemessen repräsentieren, sind am besten geeignet, fundierte und treffsichere Entscheidungen zu treffen. Ein umfassendes Training der Beobachter ist unverzichtbarer Bestandteil eines AC.
6. Systematische Vorauswahl und offene Vorinformation sind die Grundlage für den wirtschaftlichen und persönlichen Erfolg im AC. Deshalb müssen die Auswahlkriterien zur AC-Teilnahme bereits im Vorfeld offen kommuniziert und stringent für alle Personen angewendet werden.
7. Eine gute Planung und Moderation des AC gewährleisten einen transparenten und zielführenden Ablauf des Verfahrens.
8. Jeder AC-Teilnehmer hat das Recht auf individuelles Feedback, um so das Ergebnis nachvollziehen und daraus lernen zu können. Nach dem AC sind konkrete Folgemaßnahmen abzuleiten und umzusetzen.
9. Regelmäßige Güteprüfungen und Qualitätskontrollen stellen sicher, dass die mit dem AC angestrebten Ziele auch nachhaltig erreicht werden.

Assessment Center erreichen vergleichsweise hohe Validitäten. Schmidt & Hunter (1998) zitieren die Metaanalyse von Gaugler, Rosenthal, Thornton & Benson (1987), die eine korrigierte Validität von .37 berichten. Dieser Wert basiert auf 47 Studien, einer Gesamtstichprobe

---

von 12.235 Personen und 107 Validitätskoeffizienten. Die ungewichtete, unkorrigierte Validität beträgt .32, die nach Stichprobengröße gewichtete unkorrigierte Validität beträgt .29. Assessment Center erreichen damit ähnlich hohe unkorrigierte Validitätswerte wie Intelligenztests bei deutlich höherem Aufwand. Dazu muss allerdings angemerkt werden, dass die Qualität von Assessment Centern stark schwankt. Viele Verfahren, die in der Praxis eingesetzt werden, verstoßen eklatant gegen mehrere der oben zitierten Qualitätsstandards. Die Validität verschiedener Verfahren weist deshalb eine hohe Varianz auf. In der zitierten Metaanalyse schwanken die Werte zwischen -.04 und .74 mit einer Varianz von .03 – das entspricht einer Standardabweichung von .17. Die Qualität psychologischer Tests dagegen ist durchgängig relativ hoch. Ein Vergleich gemittelter Validitätswerte ist deshalb gegenüber der Assessment Center Methode tendenziell unfair.

Gering fällt die Konstruktvalidität von Assessment Center Verfahren aus. Schuler (2005) zitiert zahlreiche Studien, die alle zum gleichen Ergebnis kommen: Unterschiedliche Kriterien korrelieren hoch miteinander, wenn sie in der gleichen Übung erfasst werden. Wird das gleiche Kriterium in unterschiedlichen Übungen erfasst, korrelieren die verschiedenen Wertungen dagegen nur gering miteinander. Eine valide Konstrukterfassung nach dem Multitrait-Multimethod-Ansatz geschieht demnach nicht.

### **3.3 Intelligenz, Wissen und Berufsleistung**

Die Schmidt & Hunter-Gruppe hat sich nicht darauf beschränkt, die Validität verschiedener Verfahren der Personalauswahl metaanalytisch zu untersuchen. In zahlreichen Veröffentlichungen wurde vielmehr versucht, ein Kausalmodell zur Vorhersage von Berufsleistung zu entwickeln. Ihr Grundmodell geht davon aus, dass Intelligenz vermittelt über den Erwerb berufsrelevanten Wissens auf Arbeitsleistung wirkt. Intelligentere Menschen sind deshalb besser in ihrem Beruf, weil sie mehr berufsrelevantes Wissen erwerben. Berufsrelevantes Wissen ist in diesem Modell die Gesamtheit der Kenntnisse und Fähigkeiten, die zur erfolgreichen Ausübung eines spezifischen Berufes benötigt wird bzw. das Wissen über Fakten, Abläufe und Zusammenhänge innerhalb der eigenen Berufstätigkeit. Dieses einfache Modell wurde in weiteren Studien durch zusätzliche Variablen verfeinert, in seinem Grundaufbau aber immer bei-

---

behalten. Im Folgenden sollen drei besonders umfangreiche und aussagekräftige Studien vorgestellt werden.

### **Die Metaanalyse von Hunter**

Hunter (1986) zitiert als Basis des Grundmodells die Theorie von E. L. Thorndike, die in den 1920ern entwickelt wurde. Demnach ist Leistung durch den Umfang des erworbenen Wissens begrenzt. Wenn ein Berufstätiger nicht gelernt hat, was er in einer bestimmten Situation tun muss, kann er nicht angemessen reagieren. Wissen wiederum wird im Berufsleben in einem zweistufigen Prozess erworben. Wenn ein relevantes Ereignis eintritt muss es vom Berufstätigen zunächst als solches erkannt werden. Danach muss das im relevanten Ereignis enthaltene Lernpotenzial genutzt werden. Für beide Schritte ist Intelligenz entscheidend: Um eine kritische Situation zu erkennen muss die aktuelle Wahrnehmung mit bereits im Gedächtnis gespeicherter Information verknüpft werden. Um daraus zu lernen muss die Information so umstrukturiert werden, dass sie für zukünftige Wahrnehmungen relevant ist. Intelligenz ist nach dieser Theorie aber nicht nur für den Wissenserwerb, sondern auch für die Anwendung von Wissen entscheidend: Das passende Wissen muss aus dem Gedächtnis abgerufen und danach so modifiziert und ergänzt werden, dass es auf die spezifische Situation passt. Dementsprechend sagt die Theorie vorher, dass Intelligenz auch über das erworbene Wissen hinaus Leistung aufklärt.

Hunter (1983) prüfte diese Theorie in einer Metaanalyse. Er recherchierte Studien, die den Zusammenhang von Intelligenz, berufsrelevantem Wissen und Berufsleistung untersuchten. Dabei wurden nur Studien aufgenommen, die Berufsleistung über Arbeitsproben operationalisieren und nicht nur über Vorgesetztenbeurteilungen. Hunter sieht Arbeitsproben als direktes, objektives Maß von Arbeitsleistung, die Vorgesetztenbeurteilungen überlegen sind. 10 Studien mit insgesamt n=1790 Zivilisten und 4 Studien mit n=1474 Angehörigen des Militärs genügten Hunters Einschlusskriterien. Abbildung 3.3 zeigt die Ergebnisse einer Pfadanalyse der Daten, getrennt für zivile und militärische Daten.



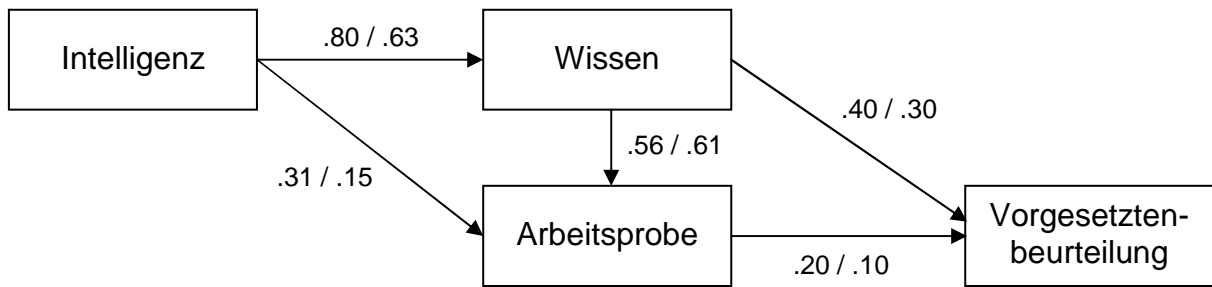


Abbildung 3.3. Pfadanalyse nach Hunter (1986). Pfadkoeffizienten vor dem Schrägstrich kommen aus den Studien mit Zivilisten, hinter dem Schrägstrich aus den militärischen Studien

Zu der Pfadanalyse muss angemerkt werden, dass die Varianzeinschränkung statistisch korrigiert wurde. Es zeigen sich sehr hohe Pfadkoeffizienten für den Einfluss von Intelligenz auf berufsrelevantes Wissen und von berufsrelevantem Wissen auf die Leistung in der Arbeitsprobe. Auch der Pfadkoeffizient zwischen Wissen und Vorgesetztenbeurteilung ist trotz der zweifelhaften Qualität dieses Kriteriums vergleichsweise hoch. Intelligenz wirkt nicht nur vermittelt über Wissen auf Arbeitsleistung, sondern es gibt auch einen direkten, wenn auch schwächeren Pfad. Diese Ergebnisse bestätigen alle Vorhersagen der Theorie von Thorndike.

### Die Arbeit von Schmidt, Hunter und Outerbridge

Die Arbeit von Schmidt, Hunter und Outerbridge (1986) erweitert das Grundmodell aus der Metaanalyse von Hunter (1983) um eine zusätzliche Variable: die Berufserfahrung in Monaten. Von den 14 Studien der Metaanalyse von Hunter (1983) geben nur die vier militärischen Studien Auskunft über diese Variable, so dass nur sie als Datenbasis der Arbeit von Schmidt, Hunter und Outerbridge dienen. Einschränkend muss weiterhin erwähnt werden, dass alle vier Studien aus einer Arbeit von Vineberg & Taylor (1972) stammen und sich hauptsächlich im untersuchten Beruf unterscheiden. Abbildung 3.4 zeigt die Pfadanalyse der Daten.

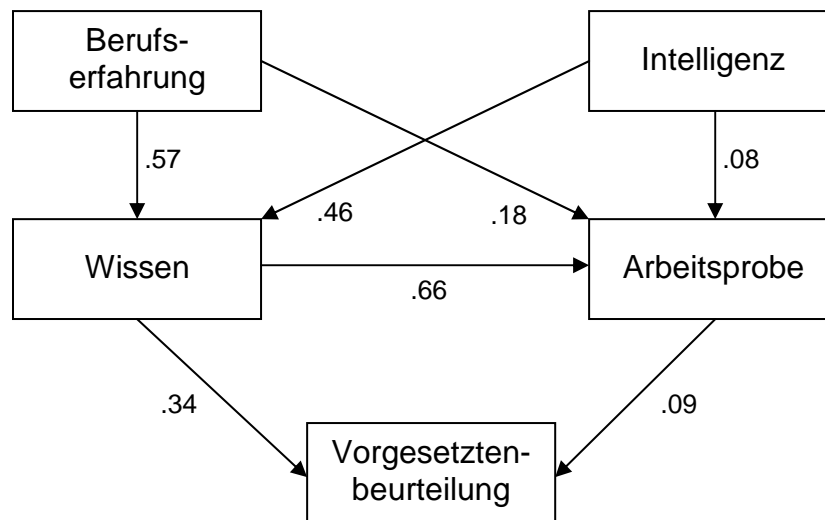


Abbildung 3.4. Pfadanalyse nach Schmidt, Hunter & Outerbridge (1986).

Die Pfadkoeffizienten fallen sehr ähnlich aus wie bei Hunter (1983), was angesichts der gleichen Datenbasis nicht verwunderlich ist. Unterschiede sind darauf zurückzuführen, dass die Werte in diesem Fall minderungskorrigiert wurden, aber keine statistische Korrektur der Varianzeinschränkung vorgenommen wurde. Interessant ist die Wirkung der zusätzlich eingeführten Variablen. Offensichtlich hängt das Ausmaß erworbenen berufsrelevanten Wissens nicht nur von der Intelligenz des Berufstätigen ab, sondern auch ganz erheblich von der Zeit, die er schon in dem betreffenden Job beschäftigt ist. Die Analyse von Schmidt, Hunter und Outerbridge zeigt allerdings, dass der Zusammenhang zwischen Berufserfahrung und beruflichem Wissen nur bis zu einer Dauer von 5 Jahren nachweisbar ist, danach wird offensichtlich kaum mehr zusätzliches berufsspezifisches Wissen erworben.

### Die Arbeit von Bormann, White, Pulakos und Oppler

Die Arbeit von Bormann, White, Pulakos und Oppler (1991) ergänzt das Modell von Hunter um weitere Variablen. Sie basiert auf einer Stichprobe von  $n=4.362$  Angehörigen der amerikanischen Streitkräfte aus neun verschiedenen Jobs. Abbildung 3.5 zeigt die Pfadkoeffizienten, die sich ergeben, wenn man das Modell von Hunter zugrundelegt.

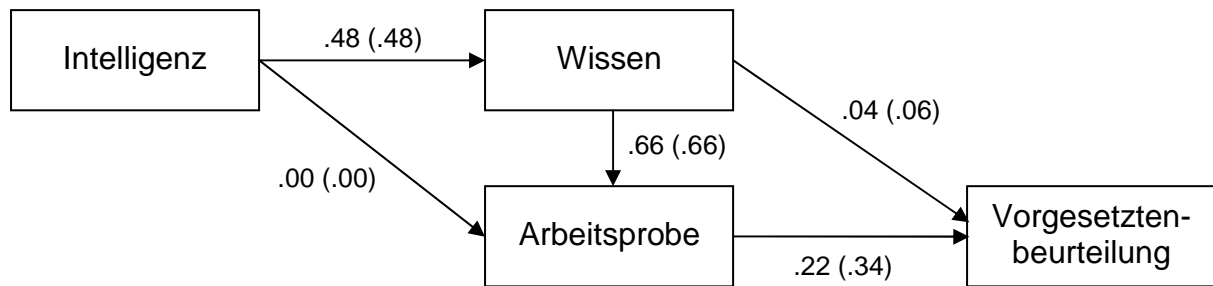


Abbildung 3.3. Pfadanalyse nach Bormann, White, Pulakos und Oppler (1991). Die Werte in Klammern sind statistisch korrigiert, um die Varianzeinschränkung auszugleichen.

Die hohen Pfadkoeffizienten zwischen Intelligenz und berufsrelevantem Wissen sowie berufsrelevantem Wissen und Leistung in der Arbeitsprobe lassen sich auch in dieser Stichprobe bestätigen. Unterschiede zeigen sich vor allem dahingehend, dass Intelligenz nur vermittelt über Wissen auf die Arbeitsleistung wirkt und darin, dass das berufliche Wissen in dieser Stichprobe kaum einen direkten Einfluss auf die Vorgesetztenbeurteilung hat.

In der Arbeit von Bormann, White, Pulakos und Oppler werden neben den bekannten Variablen auch die Gewissenhaftigkeit, Leistungsorientierung sowie Auszeichnungen und Disziplinierungen erfasst. Die Pfadkoeffizienten, die von diesen Variablen ausgehen, sind im dargestellten Gesamtmodell aber gering. Die Gewissenhaftigkeit wirkt mit einem Pfadkoeffizienten von .10 direkt auf die Vorgesetztenbeurteilung und mit einem Pfadkoeffizienten von -.23 auf die Zahl der Disziplinierungen, die dann wiederum mit einem Koeffizienten von -.18 die Vorgesetztenbeurteilung beeinflusst. Außerdem gibt es einen Pfad mit einem Koeffizienten von .14 zwischen Gewissenhaftigkeit und berufsspezifischem Wissen. Schmidt und Hunter (1998) schlagen als Erklärung für diesen Pfad vor, dass gewissenhaftere Berufstätige mehr Zeit mit ihrer eigentlichen Aufgabe verbringen und deshalb mehr lernen.



---

## 4. DAS FORSCHUNGSFELD AM BEISPIEL TAILORSHOP

Nachdem komplexe Problemlöseszenarien im dritten Kapitel zunächst ausgespart wurden, sollen sie nun am Beispiel des TAILORSHOP vorgestellt werden. TAILORSHOP ist *das* archetypische komplexe Problemlöseszenario. Es wurde schon sehr früh von Professor Dietrich Dörner auf einem programmierbaren Taschenrechner entwickelt (Dörner, 1979) und seitdem in vielen Studien eingesetzt. Es war außerdem Inspiration für viele weitere, sehr ähnliche Simulationen. Da viele der wichtigsten Erkenntnisse zum komplexen Problemlösen eigentlich Erkenntnisse zum TAILORSHOP sind, lohnt es sich, dieses Programm in allen Details zu betrachten. Dies geschieht unter Punkt 4.1. Dort wird auch auf den Begriff der Komplexität und die postulierte ökologische Validität des Szenarios eingegangen. Unter den Punkten 4.2 bis 4.7 werden dann die Studien zum TAILORSHOP vorgestellt, die „Meilensteine“ des Forschungsfeldes darstellen. Diese Studien untersuchen alle für diese Arbeit entscheidenden Fragestellungen.

### 4.1 Der Tailorshop

*„Insgesamt besteht das System aus 24 netzwerkartig miteinander verknüpften Variablen, von denen 11 durch Maßnahmen der Pbn direkt beeinflusst sind. Aus der Abbildung wird deutlich, dass das „Kapital“ die „Kernvariable“ des Systems darstellt. Es steht mit 15 anderen Variablen in direkter Verbindung.*

*Einige Variablen unterliegen in ihrem Zustand Zufallsschwankungen (z.B. Produktion, Verkauf, Rohmaterialpreis). Daneben wirkt sich der Effekt von Maßnahmen nicht immer sofort auf den Zustand direkt benachbarter Variablen aus. So steigt z.B. die Produktion der Maschinen erst einen Monat später, nachdem die Reparaturkosten erhöht und damit der Zustand der Maschinen verbessert worden ist.*

*Das System ist so programmiert, dass die Schneiderwerkstatt ohne Eingriffe Konkurs anmelden müsste. Denn es muss Rohmaterial eingekauft und die Reparaturkosten erhöht werden,*

soll die bestehende Produktionskapazität und damit der Gewinn des Betriebes bei der bestehenden Betriebsgröße gehalten werden.“ (Putz-Osterloh, 1981).

So wird der TAILORSHOP, die Computersimulation eines frühkapitalistischen, hemdenproduzierenden Unternehmens, in einer frühen Studie vorgestellt. Spätere Studien übernehmen diese Beschreibung meist weitgehend textgleich. Außerdem wird oft eine Vernetzungsgrafik ähnlich der in Abb. 4.1 gezeigt.

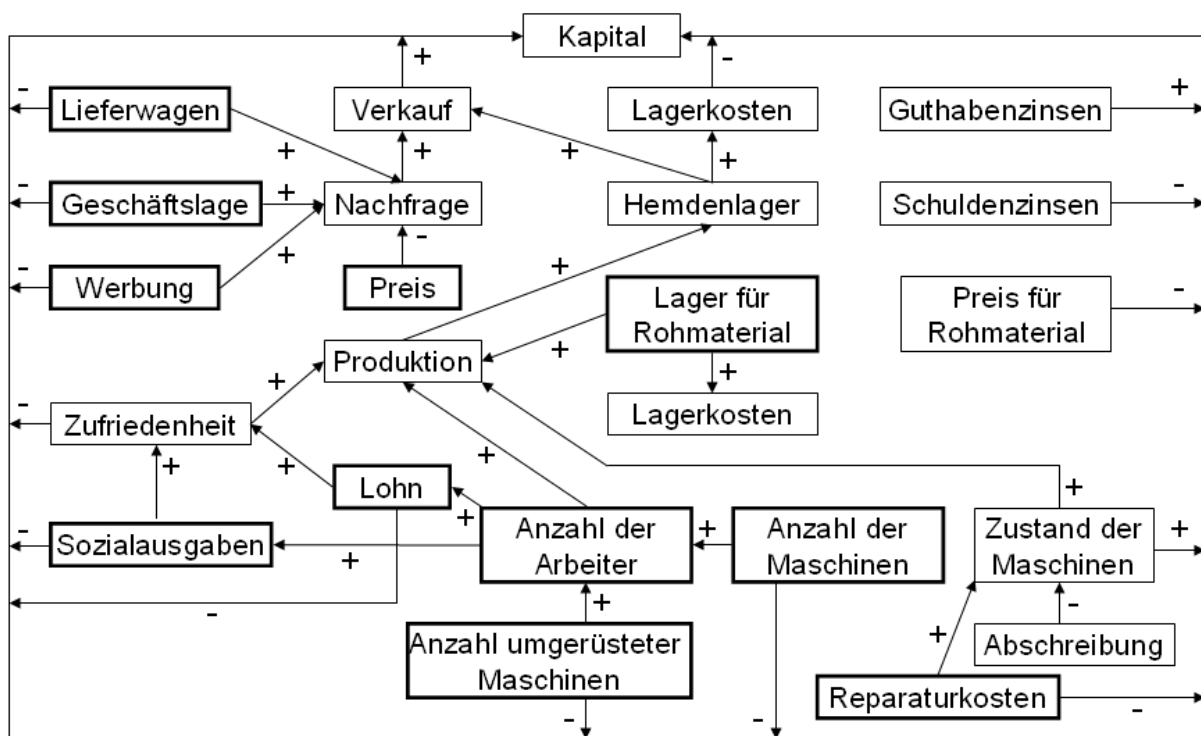


Abbildung 4.1. Vernetzungsgrafik des TAILORSHOP. Fett umrandete Variablen können vom Nutzer manipuliert werden.

Gespielt werden bis zu 24 simulierte Monate. Der Nutzer kann in jedem Monat 11 der 21 Variablen (3 der von Putz-Osterloh genannten 24 Variablen sind eigentlich Konstanten) relativ frei einstellen. Seine Einstellungen haben Auswirkungen auf die anderen 10 Variablen, insbesondere natürlich auf das Kapital.

Funke (1986) beschreibt den TAILORSHOP sehr umfassend, alle folgenden Informationen sind aus seinem Buch entnommen. Tabelle 4.1 zeigt die Startwerte der verwendeten Variablen, Tabelle 4.2 zeigt die vorkommenden Konstanten und ihre Werte.

Tabelle 4.1. Startwerte der TAILORSHOP-Variablen.

<b>Variable</b>	<b>Startwert</b>
Kapital	15774.659
Produktion	403.932
Mögliche Produktion	403.932
Rohmaterial (im Lager)	16.068
Rohmaterialpreis	3.994
Hemden (im Lager)	80.761
Hemdenpreis	52.000
Nachfrage	766.636
Verkauf	407.216
Reparaturkosten	1200.000
Maschinenkapazität	47.040
Werbung	2800.000
Lohn (pro Arbeiter)	1080.000
Sozialausgaben (pro Arbeiter)	50.000
Zufriedenheit	0.981
50erMaschinen	10
50erArbeiter (Anzahl Arbeiter an 50er-Maschinen)	8
100erMaschinen	0
100erArbeiter (Anzahl Arbeiter an 100er-Maschinen)	0
Lieferwagen	1
Geschäftslage	Cityrand

Tabelle 4.2. Werte der TAILORSHOP-Konstanten.

<b>Konstante</b>	<b>Wert</b>
Minimale Lohnkosten	850
Mietkosten Vorstadt	500
Mietkosten Cityrand	1000
Mietkosten City	2000
Kaufpreis Lieferwagen	10000
Unterhalt Lieferwagen	500
Verkaufspreis Lieferwagen	$8000 - (\text{Alter} * 100)$
Kaufpreis 50er-Maschine	10000
Kaufpreis 100er-Maschine	20000
Verkaufspreis 50er-Maschine	$(\text{Maschinenkapazität}/50) * 8000$
Verkaufspreis 100er-Maschine	$(\text{Maschinenkapazität}/50) * 16000$
Lagerkosten pro Rohmaterial	0.50
Lagerkosten pro Hemd	1.00
Guthabenzins	0.0025
Schuldenzins	0.0065

Die Werte der Variablen verändern sich jeden Monat entsprechend der folgenden Formeln:

Variable	Formel
Zufriedenheit	$Min\left(1.7, \left(0.5 + \frac{Lohn - 850}{550} + \frac{Sozialausgaben}{800}\right)\right)$
Mögliche Produktion	$\sqrt{Zufriedenheit} * \begin{pmatrix} Min(50erMaschinen, 50erArbeiter) \\ * (Maschinenkapazität + 4rnd - 2) \\ + Min(100erMaschinen, 100erArbeiter) \\ * (2 * Maschinenkapazität + 6rnd - 3) \end{pmatrix}$
Produktion	$Min(MöglicheProduktion, Rohmaterial)$
Hemden	$Hemden + Produktion - Verkauf$
Rohmaterial	$Rohmaterial - Produktion$
Nachfrage	$\frac{g * \left( Min\left(\frac{Werbung}{5}, 900\right) + Lieferwagen * 100 \right) + 100rnd - 50}{2}$ $+ 280 * 1.25 * e^{\left(\frac{Preis^2}{4250}\right)}$ wobei g=1 für Vorort, g=1.1 für Cityrand, g=1.2 für City
Verkauf	$Min(Hemden, Nachfrage)$
Rohmaterial-Preis	$2 + 6.5rnd$
Maschinen-Kapazität	$Min\left(50, \left(0.9 * Maschinenkapazität + 0.017 * \frac{Reparaturkosten}{Maschinenzahl}\right)\right)$
Kapital	$Kapital$ $- (Lohn + Sozialausgaben) * (50erArbeiter + 100erArbeiter)$ $- Produktion$ $- 0.5 * Rohmaterial$ $- Hemden$ $- Werbung$ $- 500 * Lieferwagen$ $- Mietkosten$ $- Reparaturkosten$ $+ Verkauf * Preis$
Kapital	$Kapital * Zinsen$ wobei Zinsen=0.0025 wenn Kapital>0, Zinsen=0.0065 wenn Kapital<0

rnd beschreibt eine Pseudozufallszahl zwischen 0 und 1. Pseudozufall bedeutet, dass die Zufallszahlenfolge einmal generiert wurde und für alle Probanden konstant ist.



---

## 4.1.1 Komplexität

Am TAILORSHOP lässt sich gut demonstrieren, was ein komplexes Problem ausmacht. Im Gegensatz zur klassischen Problemdefinition über Anfangszustand, Zielzustand und Barriere sind komplexe Probleme nach Dörner et al (1983, S. 19f) gekennzeichnet durch die ihnen immanente Komplexität, Intransparenz, Vernetztheit, Eigendynamik, Polytelie und gegebenenfalls Unbestimmtheit. Was bedeutet das in diesem konkreten Beispiel?

Komplexität	Die TAILORSHOP-Simulation enthält 21 Variablen und 18 Konstanten, die bei der Problemlösung bedacht werden müssen.
Vernetztheit	Die Variablen und Konstanten sind untereinander durch die genannten Gleichungen verbunden. Viele Variablen werden dabei von mehr als einer anderen Variable beeinflusst, z.B. hängt die Nachfrage von Werbung, Lieferwagen, Preis und Geschäftslage ab.
Eigendynamik	Variablen können sich eigenständig verändern, beispielsweise nimmt die Maschinenkapazität von allein kontinuierlich ab.
Intransparenz	Die Formeln, die die Variablen untereinander vernetzen, sind für den Problemlöser unbekannt. Außerdem gibt es verdeckte Variablen wie die mögliche Produktion und die Nachfrage.
Polytelie	Der Beobachter muss mehrere, einander widersprechende Ziele verfolgen. So müssen etwa die Nachfrage und die Produktion gesteigert, gleichzeitig aber möglichst wenig Geld ausgegeben werden.
Unbestimmtheit	Nach Dörner können sogar die Ziele selbst unbestimmt sein, etwa wenn dem Problemlöser nur die Instruktion gegeben wird, das Unternehmen „möglichst gut zu führen“. Beim TAILORSHOP ist das Ziel allerdings eindeutig die Maximierung des Kapitals.

Ein allgemein anerkannter Vorschlag zur Definition und Messung der Komplexität unterschiedlicher Simulationen steht nach wie vor aus. Oft wird die Zahl der Variablen als grobes Maß verwendet. Es besteht jedoch ein Konsens, dass dies unzureichend ist – so hat die Anzahl und Art der Verknüpfung unter den Variablen einen mindestens ebenso großen, wenn nicht gar größeren Einfluss auf die Komplexität einer Simulation.

---

## 4.1.2 Ökologische Validität

*„One can use computers to simulate reality by programming them to represent models of political or economic systems, for example. It is possible to define the social, psychological, economic and ecological relations of a small city as a network of interrelations and then simulate this with a computer. The computer acts then – more or less in accordance with reality – like a small town.“* (Dörner, 1987, S.97)

Der Anspruch komplexer Problemlöseszenarien zielt klar auf die realistische Simulation eines bestimmten Realitätsabschnitts ab. Ziel der Forschung zum komplexen Problemlösen war es ja, zu untersuchen, wie Menschen sich in komplexen Systemen verhalten. Dazu sollten diese Systeme möglichst detailgetreu im Labor nachgebaut werden. Der Anspruch höherer Validität komplexer Problemlöseszenarios leitet sich oft direkt aus dem Postulat höherer Realitätsnähe, oder auch „ökologischer Validität“, ab.

Zweifel an diesem Anspruch kommen schnell auf, wenn man die Algorithmen der verwendeten Problemlösesimulationen im Detail inspiziert. Dies soll hier am Beispiel TAILORSHOP geschehen, dessen Algorithmen unter 4.1 vorgestellt wurden.

1. Lohn und Sozialausgaben können vom Akteur in weiten Grenzen frei festgelegt und jeden Monat beliebig geändert werden, ohne dass die Schwankungen Auswirkungen etwa auf die Zufriedenheit haben.
2. Die Entlassung von Mitarbeitern hat keinen Einfluss auf die Zufriedenheit.
3. Kündigungen kommen nicht vor.
4. Maximale Zufriedenheit kann erreicht werden, indem die Sozialausgaben komplett gestrichen und der Lohn auf 1516.82 eingestellt werden (Funke, 1986).
5. Die Maschinenkapazität ist eine globale Variable, die sich monatlich verschlechtert, ohne dass das Alter der Maschinen Einfluss darauf hat. Wenn man also zu alten Maschinen mit geringer Maschinenkapazität neue Maschinen kauft, haben diese sofort ebenfalls eine schlechte Kapazität. Selbst der Verkauf und Neukauf aller Maschinen führt nur zu Verlusten, da die neuen Maschinen die schlechte Kapazität übernehmen.

- 
6. Der Produktpreis kann jeden Monat beliebig festgelegt werden, ohne dass Schwankungen das Kaufverhalten beeinflussen.
  7. Die Anschaffung von Lieferwagen erhöht die Nachfrage. Sonst haben sie keinen Effekt.

Dazu kommen zahlreiche Verknüpfungen, die von Probanden aufgrund ihres Vorwissens vermutet werden (z.B. Hörmann & Thomas, 1989), aber nicht im System implementiert sind. Beispielhaft sollen vier davon aufgeführt werden.

1. Saisonale Schwankungen in der Nachfrage werden nicht simuliert.
2. Für Lieferwagen müssen keine Arbeiter als Fahrer eingestellt werden.
3. Es sind keine „Springer“ an Maschinen nötig.
4. Auch wenn das Geschäft in der City liegt werden nicht weniger Lieferwagen benötigt.

Nach Inspektion der Algorithmen dürfte offensichtlich sein, dass der TAILORSHOP *keine* realistische Simulation einer Hemdenfabrik darstellt. Diese exemplarische Analyse eines archetypischen Problemlöseszenarios ließe sich – mit ähnlichem Ergebnis – für alle aktuell eingesetzten Szenarien wiederholen. Die realitätsnahe Modellierung eines komplexen Realitätsausschnittes erfordert eben doch mehr als einen einzelnen Psychologen mit Programmierkenntnissen.

Wenn sich nun die diagnostische Überlegenheit von komplexen Problemlöseszenarien nicht aus ihrer Realitätsnähe ableiten lässt, wird meist argumentiert, dass sie zumindest gleiche Anforderungen an den Bearbeiter stellen wie reale Problemsituationen:

*„Viele Alltagsprobleme sowie solche, mit denen Manager konfrontiert werden, sind jedoch typischerweise durch Intransparenz gekennzeichnet. So ist selten eindeutig nachvollziehbar, durch welche Ursachen eine aktuelle Situation entstanden ist, und somit ist nicht eindeutig prognostizierbar, was daraus künftig folgen kann. Die Probleme sind komplex und vernetzt: Entscheidungen haben neben Haupt- auch meist daran geknüpfte Nebenwirkungen, die möglicherweise erst spät sichtbar werden. Und grundsätzlich sind Probleme nicht statisch, sondern dynamisch: Die Außenwelt wartet nicht, bis ein Manager in beliebig langer Zeit alle*

---

*Argumente für oder gegen eine Entscheidung gegeneinander abgewogen hat.“ (Putz-Osterloh, 1990, S.196f.)*

Das mag plausibel klingen. Dennoch lassen sich schon bei kurzem Nachdenken Unterschiede entdecken, die Zweifel an dieser Analogie nähren: Der Zeitraffereffekt der Szenarien, die diskreten Zeittakte, die nur einen Eingriff im Monat ermöglichen, der größere Überblick über alle Variablen, deren genaue numerische Werte bekannt sind, die Abwesenheit sinnlich-körperlicher Erfahrung und sozial vermittelter Information, die Spielsituation. Die Analogie psychologischer Prozesse bei der Bearbeitung von Alltagsproblemen und komplexen Problemlösesimulationen einfach zu postulieren reicht nicht aus.

Was bleibt ist die Augenscheinvalidität (face validity – oder auch „faith validity“) komplexer Problemlöseszenarien. Das klingt abschätzig, doch sie ist der Grund für die höhere Akzeptanz der Szenarien im Vergleich zu Intelligenztests, und damit der eigentlich entscheidende Grund für deren Einsatz in der Praxis. Die fehlende Realitätsnähe und die ungeklärte Frage, ob psychologische Prozesse bei der Bearbeitung komplexer Simulationen denen bei der Bearbeitung realer Arbeitssituationen entsprechen machen lediglich deutlich, dass eine saubere Validierung der eingesetzten komplexen Problemlöseszenarien an Außenkriterien zwingend erforderlich ist. Das soll diese Arbeit im empirischen Teil leisten.

Zunächst sollen die wichtigsten Studien zum TAILORSHOP vorgestellt werden.

## **4.2 Die Studien von Putz-Osterloh und Lürer (1981)**

*„Entgegen der Erwartungen ergab sich bei der auswertbaren Teilstichprobe von n=48 Personen kein bedeutsamer Zusammenhang zwischen Problemlöse- und Intelligenztestleistung.“ (Putz-Osterloh, 1981)*

Das ist das Fazit, das Wiebke Putz-Osterloh aus ihrer ersten Studie zum TAILORSHOP „Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg“ zieht. Zwei Jahre zu-

---

vor hatte Frau Putz-Osterloh das Problemlöseverhalten beim Lösen von Aufgaben des Raven-APM Intelligenztests analysiert und folgende Problemlöseprozesse nachgewiesen:

- Herausfiltern von relevanten Merkmalen als Trägern von Veränderungen
- Analyse der Veränderung von Variablen und deren Beziehung / Relation untereinander
- Äquivalenzprüfungen zwischen verschiedenen Variabilitäten per Analogieschluss
- Aufstellen einer Regel zur Beschreibung von einander äquivalenten Variabilitäten
- Anwendung von Regeln zur Konstruktion einer Lösung

Dem stellte Sie Problemlöseszenarien gegenüber, die *zusätzlich* das Aufstellen und Ableiten von Zielen, das aktive Einholen von Informationen und die Auswahl geeigneter Maßnahmen zum Erreichen der Ziele erfordern. Putz-Osterloh erwartete Zusammenhänge zwischen Intelligenz und Problemlöseleistung nur, wenn vor allem die beim Intelligenztest erforderlichen Problemlöseprozesse wichtig sind. Dies sei der Fall, wenn das konkrete Problemlöseziel und alle möglichen Maßnahmen vorgegeben werden und ein „Kleinsystem“ mit einer überschaubaren Zahl von Variablen zu steuern ist.

Um diese Hypothese zu prüfen bearbeiteten in ihrer ersten Studie  $n=83$  studentische Versuchspersonen verschiedener Fachrichtungen für 24 simulierte Monate den TAILORSHOP und den Raven-APM Intelligenztest. Aufgrund von Versuchsleiterfehlern mussten 35 Versuchspersonen nachträglich ausgeschlossen werden, was die auswertbare Stichprobe auf  $n=48$  reduzierte. Eine Besonderheit der Studie von Putz-Osterloh war, dass die Versuchspersonen nicht selber am Computer arbeiteten, sondern alle Maßnahmen einem Versuchsleiter mitteilen mussten, der diese dann für sie in den Computer eingab.  $N=13$  Versuchspersonen arbeiteten dabei unter einer *transparenten Bedingung*, ihnen wurde der Wert aller Systemvariablen zu Beginn jedes Monats mitgeteilt.  $N=35$  Versuchspersonen arbeiteten unter einer *intransparenten Bedingung*, sie mussten den Wert aller Systemvariablen vom Versuchsleiter erfragen. Eine weitere Besonderheit war, dass alle Versuchspersonen zum lauten Denken aufgefordert wurden.

Als Maß für den Erfolg bei der Bearbeitung des Systems wurde die Anzahl von Monaten gewählt, in denen sich das Kapital erhöht. Die Rangkorrelation nach Kendall dieses Gütemaßes mit dem Intelligenzwert betrug

- 
- Transparenzbedingung (N=13):  $r=-.01$
  - Intransparenzbedingung (N=35):  $r=-.22$

Außerdem ergaben die Daten, dass Personen unter Transparenzbedingung nicht erfolgreicher waren als unter Intransparenzbedingung.

*„Das Kleinsystem der Schneiderwerkstatt scheint danach in seinen Anforderungen eher einem hochkomplexen System als dem Intelligenztest von Raven zu ähneln.“* (Putz-Osterloh, 1981) ist die Schlussfolgerung von Putz-Osterloh aus einer umfassenden Prozessanalyse.

Im gleichen Jahr wiederholte Putz-Osterloh die Studie mit veränderten Bedingungen. Diesmal bearbeiteten N=70 studentische Versuchspersonen den TAILORSHOP. N=35 arbeiteten unter *Instruktionsbedingung*, die der transparenten Bedingung der Vorgängerstudie entsprach. N=35 (von denen zwei aufgrund unvollständiger Daten ausfielen) bearbeiteten die *Grafische Bedingung*, das heißt Sie erhielten zusätzlich ein Schaubild vergleichbar mit der Abbildung 4.1, nur ohne die verdeckten Variablen und die Vorzeichen. Auf diese Weise sollte die Simulation weiter einem Intelligenztest angenähert werden. Der TAILORSHOP lief diesmal nur für 15 Monate, als Intelligenztest diente auch hier der Raven-APM. Wieder wurden die Maßnahmen der Versuchspersonen vom Versuchsleiter eingegeben und die Versuchspersonen zum lauten Denken aufgefordert.

Die Rangkorrelationen zwischen Intelligenz und Problemlöseleistung betrug diesmal:

- Instruktionsbedingung (N=35):  $r=.00$
- Grafische Bedingung (N=33):  $r=.31$

Außerdem lies sich an den Daten ablesen, dass die Versuchspersonen unter grafischer Bedingung durchschnittlich erfolgreicher waren als die unter Instruktionsbedingung.

*„Die logisch wie auch inhaltlich schwer verständlichen Nullkorrelationen zwischen Intelligenztest- und Problemlöseleistung erfahren damit eine erste Aufklärung. Erfolge beim Lösen komplexer dynamischer Probleme werden aus Ergebnissen herkömmlicher Intelligenztests erst dann vorhersagbar, wenn bestimmte Transparenzbedingungen bei der Problemstellung eingehalten werden.“* (Putz-Osterloh & Lüer, 1981).

---

Diese beiden Studien waren in mehr als einer Art und Weise interessant:

- Sie verwendeten eine komplexe Problemlösesimulation zur Messung eines Konstruktes, der komplexen Problemlösefähigkeit.
- Sie warfen die Frage nach dem Zusammenhang dieser Fähigkeit mit Intelligenz auf.
- Sie fanden keinen Zusammenhang mit Intelligenz und trugen so maßgeblich zu der frühen Lehrmeinung bei, die Leistung im Umgang mit komplexen und intransparenten Systemen würde nicht mit Intelligenz zusammenhängen.

### 4.3 Die Studie von Funke (1983)

*„Entgegen der vielfach behaupteten und auch empirisch belegten These des fehlenden Zusammenhangs von Intelligenz und Problemlösen zeigt sich nun doch das Gegenteil. Allerdings wird deutlich, dass es auch auf die Art der Lösungsgüte-Bestimmung ankommt, welche These gestützt wird.“* (Funke, 1983)

So das Fazit von Joachim Funke in seiner Studie „Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor?“, die eine Replikationsstudie der unter Abschnitt 4.2 beschriebenen Studien von Putz-Osterloh darstellt.

N=53 studentische Versuchspersonen (von denen fünf aufgrund von technischen Gründen ausfielen) bearbeiteten für 12 Monate den TAILORSHOP. Diese Versuchspersonen waren aus einer größeren Stichprobe von n=105 Personen vorselektiert worden: Die 105 Personen hatten den Intelligenztest von Raven bearbeitet, und nur das untere und obere Quartil war zur Bearbeitung des TAILORSHOP eingeladen worden.

Wie in der Studie von Putz-Osterloh und Lüer gab es eine *Intransparenzbedingung*, die der Instruktionsbedingung entsprach sowie eine *Transparenzbedingung*, die der Grafischen Bedingung entsprach. Ein wesentlicher Unterschied zu den Studien von Putz-Osterloh bestand darin, dass die Versuchspersonen nicht mit einem Versuchsleiter kommunizieren mussten. Stattdessen stellte ein Computer zu Beginn jeden Monats den Wert jeder Variablen des Systems dar und fragte dann alle möglichen Maßnahmen nacheinander ab. Außerdem wurden die Versuchspersonen nicht zu lautem Denken aufgefordert.

---

Neben dem in den Studien von Putz-Osterloh verwendeten Gütemaß der Anzahl der Anstiege im Kapital (von Funke TRENDPO genannt) wurde ein weiteres, TRENDFU genanntes Maß verwendet, das die Monate mit Anstieg im *Gesamtkapital* zählte. Das Gesamtkapital addierte dabei zum aktuellen Kapital den Verkaufswert der Maschinen und Lieferwagen (berechnet wie in Abschnitt 4.1 beschrieben) sowie den Konkurswert der Rohstoffe und Hemden im Lager (von Funke auf 2,- bzw. 20,- pro Stück festgelegt).

Die Auswertung von Funke zeigt, dass die Wahl des Gütemaßes einen entscheidenden Einfluss auf die Ergebnisse hat. Die Korrelation Eta der Intelligenz (operationalisiert als Gruppenzugehörigkeit zum unteren oder oberen Intelligenzquartil) sowie dem traditionellen Gütemaß TRENDPO betrug nur  $\eta=.14$ , was als eine Bestätigung der Befunde von Putz-Osterloh interpretiert werden könnte. Wählt man jedoch das Gütemaß TRENDFU, so beträgt die Korrelation  $\eta=.28$  – es zeigt sich ein substanzieller und signifikanter Zusammenhang zwischen Intelligenz und Problemlöseleistung.

Funke erklärt diese Befunde mit der *mangelnden Reliabilität* des in den Studien von Putz-Osterloh verwendeten Gütemaßes TRENDPO. Unterstützung findet diese These in einer Retest-Studie mit  $N=14$  Probanden, in der die Retest-Reliabilität von TRENDPO nur  $r=.20$  betrug, die von TRENDFU dagegen  $r=.79$ . Das ist plausibel: Das Maß TRENDPO bewertet Investitionen grundsätzlich erst einmal negativ, da sie das Kapital verringern. Notverkäufe von Investitionsgütern dagegen werden positiv bewertet. Beides macht wenig Sinn, wenn man die erfolgreiche Systemsteuerung messen will.

Diese Studie war in mehrfacher Hinsicht wichtig für die Forschung:

- Sie widersprach der damals gängigen Meinung, Intelligenz und Problemlöseerfolg würden nicht zusammenhängen.
- Sie stellte die wichtige Frage nach der Reliabilität der Problemlösesimulationen als Messinstrument. Diese Frage wird in Kapitel 5 dieser Arbeit aufgegriffen.
- Sie beseitigte methodische Mängel früherer Studien wie den Einfluss von Versuchsleitereffekten.



---

## 4.4 Die Studie von Süß, Kersting und Oberauer (1991)

*„Die Steuerungsleistung eines komplexen computersimulierten Systems hängt also auch unter Intransparenz mit dem Intelligenzkonstrukt „Verarbeitungskapazität“ zusammen. Die Hypothese, dass ein solcher Zusammenhang nur unter Transparenzbedingungen zu finden ist (Putz-Osterloh, 1981), ist damit entkräftet.“ (Süß, Kersting & Oberauer, 1991).*

So das Fazit der Studie „Intelligenz und Wissen als Prädiktoren für Leistungen bei computer-simulierten komplexen Problemen“, die in mehr als einer Hinsicht die ursprünglichen Annahmen zum komplexen Problemlösen erschütterte und aufgrund ihrer methodischen Sauberkeit und großen Stichprobe darin, wenn überhaupt, nur schwer zu widerlegen ist.

Die Studie bestand aus zwei Untersuchungen im Abstand von einem Jahr. An der ersten, zweitägigen Untersuchung nahmen N=214 Oberstufenschülerinnen und -schüler teil. Sie bearbeiteten einmal für 12 Monate und danach zweimal für 6 Monate den TAILORSHOP. Vor und nach dem ersten Durchgang sowie nach den drei Durchgängen wurde mit dem eigens dafür konstruierten Wissenstest WIS das Systemwissen (Wissen über Eigenschaften und Zusammenhänge von Variablen) und Handlungswissen (Wissen über adäquate Steuerungseingriffe bei gegebenem Systemzustand) erhoben. Schon vor der Problembearbeitung wurde außerdem der BIS-2 Intelligenztest und ein Test zum allgemeinem Wirtschaftswissen (WLT) bearbeitet.

Als Gütekriterium wurden zunächst das Gesamtvermögen am Ende der Steuerung sowie das Trendmaß nach Funke TRENDFU verwendet. Die Korrelationen zwischen der Intelligenzfacette Verarbeitungskapazität des Berliner Intelligenzstrukturmodells und den beiden Gütemaßen im ersten Durchgang fiel mit  $r=.13$  bzw.  $r=.15$  sehr gering aus. Für den zweiten und dritten Lauf sowie für alle anderen Intelligenzfacetten waren die Zusammenhänge sogar noch geringer. Ein Ergebnis also, das leicht als Bestätigung der Ergebnisse von Putz-Osterloh hätte gelten können.

Eine genauere Inspektion der Daten zeigte allerdings, dass der TAILORSHOP die Schüler offenbar überfordert hatte: Keiner von ihnen schaffte es, das Gesamtvermögen zu steigern,

---

und 70% schafften dies in keinem einzigen Monat. In einer Aufgabenanalyse wurde deshalb das Ziel der erfolgreichen Steuerung des TAILORSHOP in zwei Unterziele zerlegt:

1. Die Maximierung des Verkaufs.
2. Die Maximierung der Gewinnspanne pro Hemd.

Wie die Daten zeigten, hatte kein Proband über die Monate gemittelt eine positive Gewinnspanne erzielt. Da der Gewinn aber das Produkt aus Verkauf und Gewinnspanne darstellt, wird bei negativer Gewinnspanne durch Maximierung des Verkaufs der Verlust maximiert. Ein schlechter Problemlöser, der 200 Hemden mit je 100 Euro Verlust verkauft erwirtschaftet so das gleiche Ergebnis wie ein besserer Problemlöser, der 400 Hemden mit je 50 Euro Verlust verkauft. Kapitalgebundene Gütemaße können demnach bei negativer Gewinnspanne nicht zwischen guten und schlechten Problemlösern differenzieren.

Um dieses Problem zu beheben wurde ein *neues Gütemaß* entwickelt, bei dem Verkauf und Gewinnspanne z-transformiert, normalisiert und dann *addiert* wurden. Das gemittelte neue Problemlösegütemaß aus allen drei Durchgängen korrelierte mit der Verarbeitungskapazität aus dem BIS-2 zu  $r=.40$ , ein im Vergleich zu früheren Untersuchungen sehr hoher Wert. Die Korrelation mit dem BIS-2-Gesamtwert Allgemeine Intelligenz fiel mit  $r=.18$  deutlich geringer aus, die Intelligenzfacetten Bearbeitungsgeschwindigkeit, Merkfähigkeit und Einfallsreichtum waren also schlechtere Prädiktoren der Problemlösegüte.

An der Wiederholungsuntersuchung ein Jahr später nahmen  $N=137$  der 214 Probanden aus dem Vorjahr teil. Sie bearbeiteten zweimal für je zwölf Monate eine modifizierte Variante des TAILORSHOP. Die Modifikationen hatten zum Ziel, das System einfacher steuerbar zu machen. Dies wurde durch Senkung der Investitionskosten und Erhöhung der Wirkung von Werbung erreicht. Zusätzlich erhielten die Probanden zu Beginn mehr Information, etwa über die Anschaffungspreise von Maschinen oder darüber, wie viele Arbeiter pro Maschine benötigt werden. Sie konnten außerdem stets die Variablenwerte aus dem Vormonat abrufen. Diesmal bearbeiteten die Probanden den Intelligenztest BIS-3 sowie vor und nach den beiden Durchgängen den Systemwissenstest WIS.

---

Der Zusammenhang zwischen Verarbeitungskapazität und dem gemittelten neuen Problemlösegütemaß aus beiden Durchgängen betrug in der Wiederholungsuntersuchung  $r=.34$ . Bei Aggregation aller fünf Durchgänge aus Erstuntersuchung und Wiederholungsuntersuchung stieg der Zusammenhang sogar auf  $r=.47$ . Dass die Modifikation des TAILORSHOP Wirkung zeigte ist daran zu erkennen, dass in der Wiederholungsuntersuchung auch das aggregierte Gesamtkapital zu  $r=.31$  und das aggregierte Trendmaß TRENDFU zu  $r=.32$  mit der Verarbeitungskapazität korrelierten.

Neben der entscheidenden Rolle der Intelligenz deuteten die Ergebnisse der Studie von Süß, Kersting und Oberauer auch auf eine wichtige Rolle des Wissens hin. So ergaben sich sowohl in der Erstuntersuchung mit  $r=.38$  als auch in der Wiederholungsuntersuchung mit  $r=.35$  substantielle Zusammenhänge zwischen Problemlöseleistung und allgemeinen Wirtschaftswissen. Auch der *vor* der eigentlichen Systemsteuerung durchgeführte Test zum systemspezifischen Wissen korrelierte zu  $r=.40$  in der Erstuntersuchung bzw. zu  $r=.51$  in der Wiederholungsuntersuchung hoch mit der Problemlöseleistung.

Die multiple Korrelation von Intelligenz und Wissen auf der einen Seite und der Problemlöseleistung auf der anderen Seite betrug in der Erstuntersuchung  $R=.60$ , in der Wiederholungsuntersuchung  $R=.59$ .

Schließlich sind die Daten zur Stabilität der Problemlöseleistung interessant und ernüchternd. In der Erstuntersuchung korrelierten die drei Durchgänge untereinander durchschnittlich mit  $r=.57$ , die beiden Durchgänge der Wiederholungsuntersuchung korrelierten zu  $r=.74$ . Eher traurig ist dagegen die Stabilität des Problemlösegütemaßes über ein Jahr, also die Korrelation zwischen Erstuntersuchung und Wiederholungsuntersuchung: Sie beträgt lediglich  $r=.46$ . Werden Intelligenz und Wissen aus der Problemlöseleistung auspartialisiert, beträgt die Stabilität der Residuen nur noch  $r=.13$ .

Die Ergebnisse dieser Studie sind in mehrfacher Hinsicht bedeutend.

- Wie schon bei Funke konnte gezeigt werden, dass ein falsch gewähltes Gütemaß der Systemsteuerung die Ergebnisse massiv verfälschen kann. Die frühen Ergebnisse von Putz-

---

Osterloh, die bei Intransparenz keinen Zusammenhang zwischen Intelligenz und Problemlöseleistung fand, konnten so erklärt werden.

- Es wurde ein neuer wichtiger Bereich umfassend untersucht, nämlich die Frage nach dem Zusammenhang zwischen (Vor)wissen und Steuerungsleistung. Darauf wird in Kapitel 6 genauer eingegangen.
- Es wurde erneut die Frage nach der Reliabilität der Steuerungsleistung aufgeworfen. Mit  $r=.46$  fällt die Stabilität über ein Jahr gering aus.
- Schließlich wird die Existenz eines eigenen Konstrukts „Komplexe Problemlösefähigkeit“ massiv angezweifelt, da die Residuen der Steuerungsleistung nach Auspartialisierung von Wissen und Intelligenz keine Stabilität mehr aufweisen

## 4.5 Die Arbeit von Hasselmann (1993)

*„In dieser Arbeit soll die eben nahegelegte Nutzbarkeit komplexer, computersimulierter Problemstellungen als Instrument der Eignungsdiagnostik im Managementbereich untersucht werden.“* (Hasselmann, 1993)

So Dieter Hasselmann in der Einleitung zu seiner Dissertation „Computersimulierte komplexe Problemstellungen in der Management-Diagnostik“, die viele Punkte beinhaltet, die auch Thema dieser Arbeit sind.

Hasselmann stellt an eine in der Managementdiagnostik einzusetzende Simulation die folgenden Anforderungen, die auch für diese Arbeit wegweisend sind:

- Das Szenario sollte eine mittlere Größe haben, damit zum einen die Bearbeitungszeit zwei Stunden nicht überschreitet, und zum anderen die Kriterien komplexer Probleme erfüllt werden sowie Realitätsbezug und innere Plausibilität zumindest befriedigend sind.
- Das Szenario sollte einen wirtschaftlichen Kontext haben, um Akzeptanz und Identifikation zu fördern.

- 
- Umfangreiches Fach- oder Expertenwissen sollte nach Möglichkeit nicht Voraussetzung einer erfolgreichen Problembearbeitung sein, da nicht angestrebt wird, diese Komponenten zu erheben.
  - Das Ergebnis sollte nicht in erheblichem Umfang durch reine Gedächtnisleistung determiniert sein, frühere Zeittakte sollten also abrufbar sein.

Die Simulation, die diese Punkte erfüllen soll, ist die TEXTILFABRIK, die eine Weiterentwicklung des TAILORSHOP darstellt. Hasselmann geht dabei insbesondere auf die von Funke geäußerte Kritik ein. So wird das Gesamtkapital statt des Flüssigkapitals als Gütemaß verwendet, die Anschaffung neuer Maschinen verbessert den durchschnittlichen Maschinenzustand und Lieferwagen erhöhen nicht mehr die Nachfrage. Es wird allerdings auch der etwas merkwürdige Effekt eingeführt, dass die Wirkung der Werbeausgaben mit steigender Maschinenzahl geringer wird. Simuliert werden soll hier eigentlich, dass mit steigender Expansion der Produktion immer werberesistentere Kunden überzeugt werden müssen. Warum dies an der Maschinenzahl festgemacht wird, anstatt es in die Werbewirkungsfunktion zu implementieren, bleibt rätselhaft.

Die zentrale Leistung von Hasselmann stellt die Validierung dieser TAILORSHOP-Version an Außenkriterien des Berufserfolges dar. N=21 Führungsnachwuchskräfte einer norddeutschen Großbank bearbeiteten im Rahmen eines Personalentwicklungs-Assessment-Centers die TEXTILFABRIK. Etwa zwei Jahre später wurden Kriterien der betrieblichen Bewährung erhoben.

Der Kapitalendwert korrelierte mit der Position zu  $r=.49$ , mit dem Gehalt zu  $r=.53$ . Zudem korrelierte es zu  $r=.49$  mit dem durch den Personalbetreuer eingeschätzten Potential. Interessanterweise betrug die Korrelation mit der durch den Personalbetreuer bewerteten tatsächlichen Arbeitsleistung nur  $r=.16$ , mit dem eingeschätzten Nutzen für das Unternehmen gar nur  $r=.09$ .

Hasselmann berechnete auch den Zusammenhang zwischen dem Gesamtkapital und den im Assessment Center erhobenen Kriterien. Hier ergaben sich substantielle Zusammenhänge mit den Kriterien Problemlösefähigkeit ( $r=.45$ ), Entscheidungsfähigkeit ( $r=.39$ ), Überzeugungsfä-

---

higkeit ( $r=.26$ ) und Belastbarkeit ( $r=.38$ ). Korrelationen um Null ergaben sich für die Kriterien Teamgeist, Kontaktfähigkeit und Verbaler Ausdruck.

Hasselmann behandelt in seiner Arbeit darüber hinaus die Fragen nach der Reliabilität und der Konstruktvalidität der TEXTILFABRIK.

$N=41$  Studenten verschiedener Fachrichtungen bearbeiteten die TEXTILFABRIK zwei mal im Abstand von 14 Tagen. Die Korrelation beider Messzeitpunkte betrug  $r=.47$  für das Endkapital,  $r=.73$  für das Trendmaß TRENDFU.

Um die Generalisierbarkeit der Problemlöseleistung auf weitere Simulationen zu prüfen verwendete Hasselmann zwei weitere Simulationen, den HEIZOELHANDEL, einer der TEXTILFABRIK in Umfang, Oberfläche und Struktur sehr ähnlichen Simulation, und den REIFENHANDEL, ein Kleinsystem mit nur 7 Systemvariablen.

$N=22$  Studenten bearbeiteten zunächst den HEIZOELHANDEL und dann nach einer mehrwöchigen Pause die TEXTILFABRIK. Die Korrelation der beiden Kapitalendwerte betrug  $r=.48$ . Jeweils  $N=10$  Probanden bearbeiteten den REIFENHANDEL in verschiedenen Schwierigkeitsgraden und danach die TEXTILFABRIK. Nur für eine der drei Varianten ergab sich eine substantielle Korrelation von  $r=.51$ .

Die Studie von Hasselmann stellt die umfassende Validierung eines komplexen Problemlöseszenarios an Außenkriterien dar:

- Der Zusammenhang mit Außenkriterien der Berufsleistung (Position, Gehalt, Vorgesetztenrating und Assessment Center Score) wird untersucht. Dies ist eine zentrale Forderung, die in Abschnitt 4.1.2 aufgestellt wurde.
- Erneut rückt die Reliabilität ins Blickfeld. Auch hier ergeben sich mit Werten zwischen  $r=.47$  und  $r=.73$  nur schlechte bis mittelmäßige Werte.
- Schließlich wird die Konstruktvalidität untersucht. Die TEXTILFABRIK korreliert mit zwei anderen Simulationen nur um  $r=.50$ . Dies widerspricht der Annahme eines homogenen Konstrukts der Problemlösefähigkeit.

Leider muss die Aussagekraft der Studie eingeschränkt werden, da die Stichprobengrößen für eine saubere Validierung deutlich zu gering sind. Gerade für die Validierung am Außenkrite-

---

rium Berufserfolg ist ein Stichprobenumfang von  $N=21$  einfach nicht genug. Die berichteten Zusammenhänge haben damit lediglich explorativen Wert.

## 4.6 Die Studie von Wittmann, Süß und Oberauer (1996)

„Determinanten komplexen Problemlösens“ heißt die Arbeit von Werner W. Wittmann, Heinz-Martin Süß und Klaus Oberauer, die von der Anzahl eingesetzter Instrumente her wohl die umfangreichste Studie zu diesem Thema darstellt.

$N=92$  junge Erwachsene, davon 76 Studierende der Universität Mannheim, nahmen an der Untersuchung teil. Eine Teilmenge von 57 Personen konnte auch für den zweiten Teil wiedergewonnen werden.

Im ersten Teil bearbeiteten die Teilnehmer zwei Durchgänge zu je zwölf Monaten des TAILORSHOP in der von Süß weiterentwickelten Variante (siehe Abschnitt 4.4). Zuvor war ihnen der systemspezifische Wissenstest WIS vorgelegt worden. Das zweite, in zwei Durchgängen bearbeitete Szenario hieß POWERPLANT, ein Kleinsystem, das nur aus vier Variablen bestand. Auch diesen beiden Durchgängen wurde ein systemspezifischer Wissenstest vorgeschaltet. Nach den beiden Szenarien wurde noch die Kurzform des Wirtschaftskundlichen Bildungstests WBT eingesetzt.

Am zweiten Tag wurde das dritte Problemlösenszenario eingesetzt, die Wirtschaftssimulation LEARN. LEARN wurde von Wirtschaftswissenschaftlern als Trainingsprogramm entwickelt und besteht aus mehr als 2000 Variablen. Auch hier wurde ein systemspezifischer Wissenstest vorgeschaltet.

Schon in einer anderen Untersuchung hatten die Teilnehmer den BIS-4 Intelligenztest und einen Test zur Arbeitsgedächtniskapazität (WMC) bearbeitet.

Die Auswertung zeigt einen sehr hohen Zusammenhang zwischen Intelligenz, Wissen und Problemlöseleistung. Die Verarbeitungskapazität aus dem Berliner Intelligenzstrukturtest korreliert zu  $r=.48$  mit dem TAILORSHOP, zu  $r=.39$  mit POWERPLANT und zu  $r=.44$  mit LEARN. Noch besser lässt sich die Problemlöseleistung in dieser Studie durch das Vorwissen vorhersagen: Der WBT korreliert zu  $r=.43$ ,  $r=.27$  und  $r=.28$  mit den drei Szenarien, das jewei-

---

lige szenariospezifische Vorwissen zu  $r=.60$  mit TAILORSHOP und  $r=.56$  mit POWERPLANT. Interessanterweise kann man aus dem systemspezifischen Vorwissen eines Szenarios auch sehr gut die Leistung in den beiden anderen Szenarien vorhersagen. Beispielsweise korreliert das Vorwissen zu POWERPLANT zu  $r=.46$  mit dem TAILORSHOP-Ergebnis.

Sehr interessant ist die Konstruktvalidierung der Szenarien. Wie schon in der Studie von Hasselmann fiel die Korrelation der Szenarios untereinander maximal mittelhoch aus. Der höchste Zusammenhang ergibt sich mit  $r=.48$  zwischen TAILORSHOP und POWERPLANT. Ähnlich wie in der Untersuchung von Süß, Kersting und Oberauer können die Autoren zeigen, dass dieser Zusammenhang praktisch verschwindet, wenn man Intelligenz und Wissen auspartialisiert. Schon durch die Auspartialisierung der Intelligenz allein verschwindet der Zusammenhang zwischen LEARN und den anderen beiden Szenarien. Der Zusammenhang von TAILORSHOP und POWERPLANT bleibt jedoch mit  $r=.36$  signifikant.

Wie die Studie von Süß, Kersting und Oberauer konnten hier mehrere, den ursprünglichen Annahmen zum komplexen Problemlösen widersprechende Ergebnisse gefunden werden:

- Erneut wurde ein hoher Zusammenhang der Problemlöseleistung mit Intelligenz gefunden, und zwar bei allen drei eingesetzten Simulationen.
- Erneut wurde die Frage nach dem Einfluss des Vorwissens in den Mittelpunkt gerückt. Zwei der drei Simulationen weisen Zusammenhänge mit den systemspezifischen Vorwissenstests auf, die die mit Intelligenz sogar überschreiten.
- Ähnlich wie in der Studie von Süß, Kersting und Oberauer werden massive Zweifel an einem eigenständigen Konstrukt „Komplexe Problemlösefähigkeit“ geweckt. Die Korrelation verschiedener Szenarien ist wie in der Studie von Hasselmann (1993) lediglich mittelhoch, und verschwindet, wenn Intelligenz und Vorwissen auspartialisiert werden.

Wittmann und Süß (1999) beschreiben die gleiche Untersuchung mit einer größeren Stichprobe von  $N=136$  Personen. Leider werden nur multiple Korrelationen zwischen Intelligenz, Arbeitsgedächtniskapazität und Problemlösemaßen angegeben.



---

## 4.7 Die Arbeit von Kersting (1999)

*„Ausgehend von diesen Überlegungen lag es nahe, eine Studie zur prädiktiven Kriteriumsvalidierung anzustellen, in der – zur Klärung der Frage der Systemspezifität – mehr als nur ein Szenario eingesetzt wird, und in der gleichzeitig etablierte diagnostische Verfahren zu verwandten oder gar identischen Fähigkeiten, nämlich Intelligenz und Wissen, zum Einsatz kommen.“ (Kersting, 1999)*

Martin Kerstings „Überlegungen“ in den Kapiteln vier bis zehn des auf seiner Dissertation basierenden Buchs „Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien?“ stellen einen lesenswerten Verriss der Idee dar, komplexe Szenarien zur Diagnose der Problemlösefähigkeit einzusetzen. Seine fundierten Einwände stellen letztlich die Schlussfolgerung aus den in den Abschnitten 4.4 und 4.6 beschriebenen Studien dar.

Der empirische Teil beschreibt die prädiktive Kriteriumsvalidierung zweier komplexer Problemlöseszenarien. N=104 berufserfahrene Führungskräfte der Polizei bearbeiteten für zwölf Monate den TAILORSHOP in der von Süß überarbeiteten Variante und danach den zugehörigen systemspezifischen Wissenstest WIS. Außerdem bearbeiteten sie ebenfalls für zwölf Monate das von U. Funke entwickelte komplexe Problemlöseszenario DISKo (U. Funke, 1992). DISKo lehnt sich an den TAILORSHOP an und simuliert mit 17 direkt und 24 indirekt beeinflussbaren Variablen eine Chipfabrik. Außerdem wurde mit dem DKT-W ein Kenntnistest zum Wirtschaftswissen und natürlich ein Intelligenztest eingesetzt: N=64 Personen bearbeiteten den BIS-4, N=40 Personen einen unveröffentlichten Intelligenztest des DGP.

Für N=73 Polizisten wurden durchschnittlich ein Jahr und siebeneinhalb Monate später Vorgesetztenbeurteilungen der Berufsleistung erhoben.

Die Auswertung der Prädiktoren ergab eine erneute Bestätigung früherer Befunde.

- Verarbeitungskapazität und TAILORSHOP-Kapitalendwert hingen zu  $r=.21$  miteinander zusammen. Der DISKo-Kapitalendwert korrelierte nicht mit Intelligenz, was sich analog zu der Studie von Süß, Kersting und Oberauer auf eine Überforderung der Probanden zurückführen ließ. Ein neu definiertes Gütemaß, das Verkauf und Gewinnspanne nicht multiplizierte sondern addierte, korrelierte zu  $r=.26$  mit der Verarbeitungskapazität.

- 
- Die Zusammenhänge mit den systemspezifischen Wissenstests betragen  $r=.22$  für den TAILORSHOP und  $r=.32$  für DISKo.
  - Beide Szenarien korrelierten untereinander nur zu  $r=.23$ , was die Konstruktvalidität erneut in Frage stellt.

Die prädiktive Validierung ergab substantielle Korrelationen mit dem Vorgesetztenurteil. Das TAILORSHOP-Ergebnis korrelierte zu der vom Vorgesetzten eingeschätzten beruflichen Problemlöseleistung zu  $r=.37$ , das DISKo-Ergebnis zu  $r=.33$ . Die vom Vorgesetzten eingeschätzte Kooperationsfähigkeit konnte erwartungsgemäß durch die Problemlösesimulationen kaum vorhergesagt werden.

Kerstings Arbeit greift verschiedene Punkte aus früheren Studien wieder auf:

- Der Zusammenhang mit Außenkriterien der Berufsleistung wird untersucht. Dies ist eine zentrale Forderung, die in Abschnitt 4.1.2 aufgestellt wurde. Im Gegensatz zu Hasselmann verwendet Kersting eine große Stichprobe. Bei der Interpretation der Höhe der Zusammenhänge muss allerdings bedacht werden, dass hier nur die berufliche Problemlöseleistung als Kriterium dient, nicht die Berufsleistung insgesamt.
- Es lassen sich Zusammenhänge zwischen Intelligenz und Wissen auf der einen Seite sowie Problemlöseleistung auf der anderen Seite finden.
- Wie bei Funke (1983) zeigt sich am Szenario DISKo noch einmal, dass eine falsche Operationalisierung der Problemlöseleistung die Ergebnisse verfälscht. Wieder ist die zu große Schwierigkeit des Szenarios das Problem.
- Erneut korrelieren zwei Problemlöseszenarien nur schwach miteinander. Auch auf Basis dieser Untersuchung muss die Generalität des Konstrukts Problemlöseleistung verworfen werden.

---

## 5. RELIABILITÄT KOMPLEXER SZENARIEN

Die präsentierten Studien zum TAILORSHOP haben gezeigt, wie wichtig die Reliabilität komplexer Szenarien als Voraussetzung ihrer Validität ist. Wie in den folgenden Abschnitten gezeigt wird ist es schwer, die Reliabilität komplexer Problemlöseszenarien zu bestimmen – sie kann aber sehr wohl anhand theoretischer Überlegungen optimiert werden.

### 5.1 Messung der Reliabilität

Zur Messung der Reliabilität des Verfahrens werden in der klassischen Testtheorie die Verfahren der Testwiederholung, des Paralleltests, der Testhalbierung oder der Internen Konsistenz verwendet.

#### 5.1.1 Retest-Reliabilität

Retest-Reliabilität bezeichnet die Methode, ein und dasselbe diagnostische Instrument bei der gleichen Personengruppe mehrmals in einem gewissen Abstand anzuwenden und die Ergebnisse zu korrelieren.

Diese Methode kann relativ einfach auf Problemlöseszenarios übertragen werden, folglich wurde dies in der Vergangenheit am häufigsten versucht. Für eine Übersicht über vorhandene Studien siehe z.B. Hasselmann (1993), Funke (1995), Kersting (1999, S103ff). Tabelle 5.1 zeigt ebenfalls eine Übersicht.

Tabelle 5.1. Reteststudien zu komplexen Problemlöseszenarien

Quelle	Szenario	Intervall	N	$r_{tt}$
Strohschneider (1986)	Moro	11 - 55 Tage	n=25	$r_{tt}=.26-.44$ (5 Messzeitpunkte)
Schoppek (1996)	Jogi'91	sofort	n=31	$r_{tt}=.38, r_{tt}=.53$ (Zwei Gütemaße)
Schoppek (1991)	Feuer	sofort	n=22	$r_{tt}=.44-.82$ (5 Messzeitpunkte)
Putz-Osterloh (1991)	Feuer	sofort	n=50	$r_{tt}=.60-.89$
Putz-Osterloh (1991)	Feuer	sofort	n=80	$r_{tt}=.84$

---

Putz-Osterloh und Haupts (1990)	Feuer	sofort	n=30	$r_{tt}=.48$
Funke (1983)	Schneiderwerkstatt	2 Wochen	n=14	$r_{tt}=.8$ (Trendmaß)
Süß (1996)	Schneiderwerkstatt	1 Jahr	n=137	$r_{tt}=.46$
Hasselmann (1993)	Textilfabrik	2 Wochen	n=52	$r_{tt}=.43$ (Kapital) $r_{tt}=.73$ (Trendmaß)
Hasselmann (1993)	Heizölhandel	2 Wochen	n=14	$r_{tt}=.67-.74$
Locher (1997)	Heizölhandel	1 Tag	n=20	$r_{tt}=.12$ (Kapital) $r_{tt}=.58$ (Trendmaß)
U. Funke (1993)	DISKo	sofort	n=10	$r_{tt}=.54$

Man sollte sich allerdings davor hüten, die resultierenden Ergebnisse unreflektiert als Maß der Reliabilität der Szenarios zu sehen. Zentrales Merkmal komplexen Problemlösens ist die aktive Auseinandersetzung mit einer dynamischen Situation. Ziel ist es, möglichst viel Wissen über die Situation zu sammeln um sie mit Hilfe dieses Wissens erfolgreich zu steuern. Die als stabil angenommene Fähigkeit zum Umgang mit komplexen Problemen hilft also dabei, im Umgang mit dem vorliegenden spezifischen komplexen Problem immer besser zu werden. Das bedeutet, dass die angenommene stabile Problemlösefähigkeit nur schwer noch einmal mit dem gleichen komplexen Problem gemessen werden kann.

### 5.1.2 Parallelttest-Reliabilität

Um die Parallelttest-Reliabilität zu bestimmen, müssen zwei äquivalente Szenarios vorgegeben und die Ergebnisse korreliert werden. Dies soll Wiederholungseffekte minimieren. Die Frage, wie man äquivalente Szenarios baut, ist bisher jedoch nicht schlüssig beantwortet worden. Das gleiche Szenario inhaltlich anders einzukleiden oder mit anderen Startwerten vorzulegen fällt eher unter Retest-Messung. Zwei unterschiedliche Szenarios zu korrelieren ist dagegen eher eine Untersuchung der Generalität bzw. Konstruktvalidität der Problemlöseleistung.

Einen anderen Ansatz testete Funke (1995). Er gab 14 Probanden gleichzeitig das Szenario „DISKo“ und die Parallelversion „Videofabrik“ vor, die als deutsche und französische Niederlassung eines Konzerns vorgestellt wurden und zwischen denen mit Tastendruck gewechselt werden konnte. Die Korrelation beider Versionen betrug  $r=.83$ .

---

### 5.1.3 Testhalbierung und Interne Konsistenz

In der klassischen Testtheorie wird die Halbierungs-Reliabilität bestimmt, indem die Items eines Tests in zwei Gruppen aufgeteilt werden und das Ergebnis beider Gruppen korreliert wird. Die Interne Konsistenz ist die korrigierte durchschnittliche Korrelation aller Items untereinander. Es ist deshalb jeweils nur eine Messung nötig.

Das Fehlen unabhängiger Items verhindert die Anwendung dieser Technik auf komplexe Problemlöseszenarien. Jeder Zeittakt und damit auch die erzielbare Leistung in jedem Zeittakt hängt von allen vorhergehenden Zeittakten ab, es gibt mit dem Problemlösegütemaß letztlich nur eine einzige unabhängige Messung.

Wagener (2001) versucht, dieses Problem bei FSYS 2.0 zu lösen, indem er die durch Abhängigkeit der Zeittakte entstehende gemeinsame Varianz herausrechnet. Dazu korreliert er einerseits die ungeraden und geraden Zeittakte von 161 Studenten und andererseits die von 500 computersimulierten Problemlösern, die sich völlig unsystematisch verhielten. Als Schätzung der internen Konsistenz verwendet er das Quadrat der Differenz der beiden Determinationskoeffizienten und kommt auf  $r=.80$ .

Kröner (2001) teilt die Bearbeitung von MULTIFLUX in eine nicht bewertete Explorationsphase, in der die Zeittakte voneinander abhängig sind und eine Testphase, in der die Zeittakte nicht mehr voneinander abhängig sind. In der Explorationsphase soll der Proband Wissen über das System sammeln. In der Testphase wird ihm dann mehrfach jeweils ein Systemzustand und eine Zielvorgabe präsentiert. Er erhält keine Rückmeldung mehr über seine Eingriffe, so dass in dieser Phase auch nicht von „Item“ zu „Item“ dazugelernt wird. Um die Reliabilität zu berechnen wendet Kröner die Methode der Internen Konsistenz auf die „Items“ der Testphase an. Das Problem an diesem Vorgehen ist, dass die „Items“ alle vom einmaligen falschen oder richtigen Vorgehen in der Explorationsphase abhängig sind und alle ein und dasselbe Szenario betreffen. Von Unabhängigkeit kann also nicht gesprochen werden.

---

## 5.2 Das Primat der Validität

Keine der dargestellten Methoden ist ein adäquater Weg, die Reliabilität eines Szenarios zu bestimmen. In der angewandten Eignungsdiagnostik ist es aber letztlich sowieso nur die Kriteriumsvalidität, die wirklich zählt. So werden auch für Assessment Center regelmäßig geringe Werte für Reliabilität und Konstruktvalidität ermittelt, trotzdem werden sie aufgrund ihrer hohen Akzeptanz und Validität eingesetzt.

Die in Abschnitt 4.3, 4.4 und 4.7 geschilderten Studien haben gezeigt, was passiert, wenn ein Szenario die Problemlösefähigkeit nicht reliabel misst: Die Korrelation mit Außenkriterien, in diesem Fall Intelligenz, bricht auf Null ein. Der Satz der klassischen Testtheorie, dass ein Instrument nie höher mit einem Außenkriterium korrelieren kann als mit seinem eigenen wahren Wert, die Validität demnach nie höher sein kann als die Reliabilität, kann als Erklärung solcher Befunde dienen. Reliabilität ist kein Wert an sich, sondern nur eine notwendige Voraussetzung für Validität.

Soll ein Problemlöseszenario explizit für den Einsatz in der Eignungsdiagnostik eingesetzt werden, macht es demnach Sinn, keine Zeit für die empirische Bestimmung der Reliabilität zu verschwenden, sondern das Szenario direkt an Außenkriterien zu validieren. Sind die ermittelten Zusammenhänge in einer großen Stichprobe befriedigend, kann daraus auch auf eine hinreichende Reliabilität des Szenarios geschlossen werden.

## 5.3 Optimierung der Reliabilität

Ein Verzicht auf die empirische Bestimmung der Reliabilität heißt natürlich nicht, dass sie ignoriert werden darf. Bei der Konstruktion eines Szenarios ist es sehr wichtig, sich darüber Gedanken zu machen, wie sie optimiert werden kann.

Wagener (2001) stellt eine umfangreiche Taxonomie komplexer Szenarios (TAKS) vor, aus der sich zahlreiche Punkte ableiten lassen, die für eine Optimierung der Reliabilität entscheidend sind:

- 
- *Akzeptanz von Inhaltsbereich und Aufgabe*: Die inhaltliche Einkleidung des Szenarios sollte so gewählt werden, dass die Akzeptanz bei allen Teilnehmern gleichmäßig hoch ist, um eine Konfundierung des Gütemaßes mit motivationalen Einflüsse zu verhindern.
  - *Impact inhaltlichen Vorwissens*: Da eine generelle Problemlösefähigkeit gemessen werden soll, sollte der Einfluss spezifischen Vorwissens gering sein.
  - *Modus der Kommandoübermittlung und Informationsbeschaffung*: Der Proband sollte direkt mit dem Computer arbeiten. Die früher übliche Bereitstellung von Informationen und Eingabe von Maßnahmen durch den Versuchsleiter kann zu Versuchsleitereffekten und dem Einfluss von Persönlichkeitseigenschaften wie Extraversion führen.
  - *Ergonomie der Benutzeroberfläche*: Komplexe Problemlöseszenarien sollten möglichst einfach zu bedienen sein. Je höher die kognitive Beanspruchung durch die Programmbedienung, desto weniger Ressourcen sind für die Problemlösung frei. Dies kann dazu führen, dass der Problemlösescore stark von der Computererfahrung abhängig wird.
  - *Notizen*: Dem Akteur sollte erlaubt werden, Notizen zu machen, um einen zu hohen Einfluss der Gedächtnisleistung auf das Problemlöseergebnis zu verhindern.
  - *Umfang und Reduzierbarkeit der Regeln*: Lässt sich die Systemsteuerung auf wenige einfache „Faustregeln“ reduzieren, ist die Reliabilität und Validität des Instruments auf Dauer gefährdet. Denn einfachen Regeln lassen sich leicht erinnern und an spätere Probanden weitergeben.
  - *Reversibilität*: Reversibilität beschreibt die Gutmütigkeit des Systems nach Fehlsteuerungen. Ein System sollte sich in den Takten nach einer massiven Fehlsteuerung wieder in eine Aufwärtsphase steuern lassen. Sonst kann es passieren, dass ein Proband aufgrund des durch eine Fehlsteuerung erzeugten irreversiblen Zustands nicht mehr in der Lage ist, überhaupt Problemlösefähigkeit zu zeigen.
  - *Übermächtige Entscheidungen*: Es sollte keine Einzelentscheidungen geben, die allein einen übermächtigen Einfluss auf das System haben. So ist es z.B. beim TAILORSHOP sehr wichtig, ausschließlich 100-Hemden-Maschinen zu verwenden, sonst fällt die Rendite gering aus.
  - *Zufallseinflüsse*: Es ist zwischen Echten Zufällen und Pseudo-Zufällen zu unterscheiden. Echte Zufälle erzeugen bei jedem Szenariolauf unterschiedliche Werte. Pseudo-Zufälle erzeugen zwar bei jedem Szenariolauf die gleichen Werte, diese sind aber für den Akteur

---

nicht vorhersagbar. Auf echte Zufälle sollte aus naheliegenden Gründen verzichtet werden. Auch Pseudo-Zufälle können dann ein Problem sein, wenn sie je nach aktuellem Wert der Systemvariablen von Proband zu Proband einen unterschiedlichen Einfluss haben.

- *Information über Erfolgskriterien:* Das Erfolgskriterium der Problemlöseleistung sollte offengelegt werden. Wenn sich Probanden eigene Hauptziele setzen besteht die Gefahr, dass die Auswertung nicht mehr auf der Problemlösekompetenz beruht sondern vielmehr darauf, ob die Ziele des Probanden mit denen des Problemkonstruktors übereinstimmen.

Ein weiterer sehr wichtiger und sehr allgemeiner Punkt ist die Problemschwierigkeit. Die Studien von Süß, Kersting und Oberauer (1991) und Kersting (1999) zeigen sehr deutlich, dass eine zu hohe Problemschwierigkeit dazu führen kann, dass gute und schlechte Problemlöser nicht mehr differenziert werden. Auf der anderen Seite ist denkbar, dass ein zu leichtes Szenario aufgrund von Deckeneffekten ebenfalls nicht mehr differenziert.

Werden bei der Neukonstruktion eines Szenarios diese Punkte beachtet, sollte eine hinreichende Reliabilität resultieren. Ob dies wirklich der Fall ist, zeigt die Validierung.



---

## 6. INTELLIGENZ, WISSEN UND PROBLEMLÖSEN

„Intelligenz, Wissen und Problemlösen“ ist der Titel eines Buches von Heinz Martin Süß (1995), der die Bedeutung dieser beiden Konstrukte für die Problemlöseforschung hervorhebt. Die Studien von Süß, Kersting und Oberauer (1991) und Wittmann und Süß (1999) zeigen eindrucksvoll hohe Zusammenhänge zwischen diesen drei Konstrukten. Funke (1988), Kersting (1999) und Wagener (2001) können diese Befunde ebenfalls bestätigen.

### 6.1 Der Zusammenhang mit Intelligenz

„Gemeinsamer Kern praktisch aller Intelligenzkonzepte ist die Qualität und Geschwindigkeit der Lösung neuer (also nicht routinebestimmter) Aufgaben.“ (Schuler, 2005). Wird das Konstrukt Intelligenz dieser breiten Gültigkeit gerecht, müsste sich ein Zusammenhang mit der komplexen Problemlöseleistung nachweisen lassen – komplexe Problemlöseszenarien sind schließlich nichts anderes als neue, anspruchsvolle Aufgaben. Die in den Abschnitten 4.2 bis 4.7 vorgestellten Studien zeigen, wie sich die Meinung zum Zusammenhang von komplexer Problemlöseleistung und Intelligenz gewandelt hat. Wurden zunächst mehrheitlich Nullkorrelationen gefunden und durch die unterschiedlichen Anforderungen von Simulationen und Intelligenztests erklärt, so fanden sich in späteren Studien vermehrt substantielle Zusammenhänge. Doch was ist Intelligenz überhaupt? Eine häufig zitierte Definition von Stern (1911) konzeptualisiert Intelligenz als allgemeine Fähigkeit eines Individuums, sein Denken bewusst auf neue Situationen einzustellen. Verschiedene Konzepte dieser allgemeinen Fähigkeit unterscheiden sich vor allem in der postulierten Intelligenzstruktur.

#### 6.1.1 Strukturmodelle der Intelligenz

Die Diskussion um die Dimensionalität der Intelligenz reicht weit zurück. Die „Zweifaktorentheorie“ von Spearman (1904) unterscheidet nur die Allgemeine Intelligenz als Generalfaktor „g“ und spezifische Faktoren (s-Faktoren). Letztere sind als aufgabenspezifische Fähigkeiten

---

nicht der Intelligenz zuzurechnen, die als eindimensionale, nicht in Komponenten zerlegbare Fähigkeit angesehen wird. Die Gegenposition vertrat Thurstone (1938) in der „Primärfaktorentheorie“. Sie unterscheidet sieben voneinander unabhängige, spezifische Faktoren: Sprachverständnis, Wortflüssigkeit, Raumvorstellung, Wahrnehmungsgeschwindigkeit, Rechenfertigkeit, schlussfolgerndes Denken und Merkfähigkeit. Gestützt auf den Befund, dass Untertests von Intelligenztests stets positiv miteinander korrelieren, wurden beide Modelle später zu hierarchischen Modellen weiterentwickelt, in denen die Allgemeine Intelligenz als generellstes Konstrukt auf der obersten Ebene angesiedelt ist und voneinander nicht unabhängige spezifische Faktoren auf darunterliegenden Ebenen. Heute dominieren in der Intelligenzforschung Versuche, verschiedene Intelligenzkonzepte im Rahmen solcher hierarchischer Modelle zu integrieren. Die Ansätze sind auch für diese Arbeit relevant und sollen deshalb im Folgenden vorgestellt werden.

### **Die Theorie der Fluiden und Kristallinen Intelligenz**

Die Theorie von Horn und Cattell (1966) unterscheidet zwei Faktoren der Intelligenz. Fluide Intelligenz beschreibt Basisprozesse des Denkens und anderer mentaler Aktivitäten, deren Leistungsfähigkeit primär genetisch determiniert ist. Kristalline Intelligenz dagegen ist die Fähigkeit, erworbenes (kristallisiertes) Wissen zur Lösung von Problemen anzuwenden. Zur Messung der fluiden Intelligenz werden figural-bildhafte Aufgaben zum induktiven und deduktiven Denken verwendet. Cattell bezeichnet diese Aufgaben als *culture fair*, da sie im Gegensatz zu den zur Messung der kristallinen Intelligenz verwendeten numerischen und verbalen Aufgaben nicht kulturabhängig seien.

### **Das Berliner Intelligenzstrukturmodell**

Das Berliner Intelligenzstrukturmodell wurde von Jäger (1982) erstmals vorgestellt und soll unterschiedliche Intelligenzmodelle integrieren. Zur Konstruktion wurde die Struktur einer repräsentativen Stichprobe aller in der Forschungsgeschichte publizierten Intelligenztestaufgaben untersucht und ein Mehrebenenmodell gefunden. In diesem Modell wird auf der höchsten Ebene die Allgemeine Intelligenz als Integral aller Fähigkeitskomponenten angenommen. Auf der zweiten Ebene werden sieben hochgradig generelle Strukturkomponenten postuliert, die in zwei Modalitäten aufgeteilt sind, Operationen und Inhalte.

---

- *Operationen*

Verarbeitungskapazität, Bearbeitungsgeschwindigkeit, Merkfähigkeit, Einfallsreichtum

- *Inhalte*

Figural-bildhaft, Numerisch, Verbal

Die Fähigkeitskomponenten stehen nicht orthogonal zueinander, sondern hängen zusammen. Als Kombination von Operationen und Inhalten ergeben sich auf der dritten Stufe zwölf spezifische Teilfähigkeiten, die allerdings im Modell keine Rolle spielen. Interessant sind sie dennoch, da sie die Einordnung anderer Intelligenztests in das Modell ermöglichen. So stellen sich die Advanced Progressive Matrices von Raven als Maß der Figural-bildhaften Verarbeitungskapazität und nicht der Allgemeinen Intelligenz dar.

Eine der Kernannahmen des Modells ist, dass an jeder Intelligenzleistung alle Komponenten beteiligt sind, jedoch mit deutlich unterschiedlicher Gewichtung. Eine Besonderheit ist die Integration der Kreativität in Form der Facette Einfallsreichtum.

### **Die Three-Stratum-Theorie**

Die Three-Stratum-Theorie von Carroll (1993) resultierte aus einer Reanalyse aller publizierten Korrelationsmatrizen von Intelligenzleistungen. Ergebnis war ein hierarchisches Strukturmodell mit drei Ebenen. Auf der obersten Ebene findet sich die Allgemeine Intelligenz, auf der zweiten Ebene acht breite Konstrukte: Fluid Intelligence, Crystallized Intelligence, General Memory and Learning, Broad Visual Perception, Broad Auditory Perception, Broad Retrieval Ability, Broad Cognitive Speediness und Processing Speed. Auf der dritten Ebene finden sich 68 spezifische Fähigkeitskonstrukte – fast alle, die je publiziert wurden.

Im Vergleich zum Berliner Intelligenzstrukturmodell fällt zunächst auf, dass operative und inhaltsgebundene Fähigkeiten auf der zweiten Ebene nebeneinander existieren. Sie überlappen sich zum Teil stark mit den Facetten des BIS, es gibt aber auch bedeutende Unterschiede. So fehlen beispielsweise auditive Aufgaben im BIS völlig, die Three-Stratum-Theorie wiederum verzichtet auf die Einbindung der Kreativität.

---

## 6.1.2 Verarbeitungskapazität

*"Verarbeitung komplexer Information bei Aufgaben, die nicht auf Antriebe zu lösen sind, sondern Heranziehen, Verfügbarhalten, vielfältiges Beziehungsstiften, formallogisch exaktes Denken und sachgerechtes Beurteilen von Information erfordern"* (Jäger, 1982)

Neuere empirische Studien haben den Zusammenhang verschiedener Intelligenzfacetten mit der komplexen Problemlöseleistung untersucht. Der höchste Zusammenhang ergab sich dabei fast durchgängig mit der Verarbeitungskapazität. Verarbeitungskapazität ist eine Facette des im vorigen Abschnitt vorgestellten Berliner Intelligenzstrukturmodells. Wie Intelligenz ist sie ein hochgradig generell definiertes Konstrukt mit dem Anspruch, bei jeder kognitiven Leistung beteiligt zu sein. Wird sie diesem Anspruch gerecht, ist ein substantieller Zusammenhang auch zur Problemlöseleistung zu erwarten, die ja eine kognitive Leistung darstellt.

Zur Messung der Verarbeitungskapazität werden klassische Aufgaben wie Zahlenreihen, Analogien und Schlüsse eingesetzt. Sie findet ihre Entsprechungen unter anderem Namen auch in den anderen Intelligenzmodellen: Bei Thurstone heißt sie reasoning, bei Carroll fluid intelligence. Auch die von Cattell verwendeten Aufgaben zur Messung der fluiden Intelligenz messen das gleiche wie Verarbeitungskapazität.

## 6.1.3 Arbeitsgedächtniskapazität

Das Berliner Intelligenzstrukturmodell ergab sich aus statistischen Analysen, nicht aus einer Theorie kognitiver Prozesse. Dies trifft auch auf die meisten anderen Intelligenzstrukturmodelle zu. Das Konstrukt der Intelligenz im Allgemeinen ist theoretisch leider noch immer nur schwach unterfüttert.

Neuere Studien weisen darauf hin, dass sich die gemeinsame Varianz der unterschiedlichen Aufgaben zur Messung der Verarbeitungskapazität über die Kapazität des Arbeitsgedächtnisses aufklären lässt. Das Arbeitsgedächtnis ist in der Kognitionspsychologie als zentrale kognitive Ressource für die gleichzeitige Kurzzeitspeicherung und Verarbeitung von Information konzipiert. Je größer die Kapazität des Arbeitsgedächtnisses, desto mehr Information kann gleichzeitig gespeichert und verarbeitet werden. Baddeley & Hitch (1974) differenzieren das

---

Arbeitsgedächtnis in zwei unselbständige „slave systems“ und eine „zentrale Exekutive“. Der „articulatory loop“ speichert und verarbeitet sprachliches, der „visuo-spatial scratch pad“ visuell-räumliches Material. Die „zentrale Exekutive“ fungiert als „supervisor“, der die Prozesse in den „slave systems“ steuert und überwacht. Außerdem speichert und verarbeitet sie Material, das nicht von den „slave systems“ bearbeitet wird.

Wittmann, Süß, Oberauer, Schulze & Wilhelm (1995) finden einen Zusammenhang von  $r=.80$  zwischen Aufgaben zur Erfassung der Arbeitsgedächtniskapazität und der Verarbeitungskapazität aus dem BIS-4. Die Aufgaben zur Messung der Arbeitsgedächtniskapazität waren dabei von denen zur Messung der Verarbeitungskapazität deutlich verschieden.

Sollten sich diese Befunde bestätigen darf angenommen werden, dass Verarbeitungskapazität und Arbeitsgedächtniskapazität das gleiche sind. Der zentrale Anspruch des Arbeitsgedächtnisses als Grenze kognitiver Leistungen wäre bestätigt und ein großer Schritt hin zu einer theoretischen Fundierung des Intelligenzkonstrukts getan.

Komplexe Problemlöseszenarien erfordern die gleichzeitige Kurzzeitspeicherung, Verarbeitung und Integration unterschiedlicher Systemzustände um ein mentales Kausalmodell der Simulation zu erstellen. Das Arbeitsgedächtnis als entscheidend für den Problemlöseerfolg anzunehmen liegt also nahe und wäre eine Erklärung für den Zusammenhang von Verarbeitungskapazität und komplexer Problemlöseleistung.

#### **6.1.4 Der Zusammenhang von Intelligenz und Problemlöseleistung**

*„Reasoning, problem solving and intelligence are so closely interrelated that it is often difficult to tell them apart. ... Whatever intelligence may be, reasoning and problem solving have traditionally been viewed as important subsets of it. Almost without regard to how intelligence has been defined, reasoning and problem solving have been part of the definition.”*  
(Sternberg, 1982, S. 225)

*“Bei derart hohen Übereinstimmungen auf der Ebene der Theoriesprache ist zu erwarten, dass die jeweiligen Indikatoren der Konstrukte (Ebene der Beobachtungssprache) miteinander*

---

*der korrelieren, und dass z.B. die Bearbeitung von Problemlöseszenarien Anforderungen an die Intelligenz der Versuchspersonen stellt.“ (Kersting, 2001)*

So oder so ähnlich wird die Annahme eines Zusammenhangs zwischen Intelligenz und Problemlöseleistung häufig begründet. Die vorigen Abschnitte machen da keine Ausnahme: Intelligenz als Qualität und Geschwindigkeit der Lösung neuer Aufgaben, Verarbeitungskapazität als hochgradig generell definiertes Konstrukt mit dem Anspruch, bei jeder kognitiven Leistung beteiligt zu sein, Arbeitsgedächtniskapazität als zentrale kognitive Ressource für die gleichzeitige Kurzzeitspeicherung und Verarbeitung von Information – die Konstrukte sind so allgemein definiert, dass sie mit Problemlöseleistung korrelieren *müssen*.

Zu einer theoretischen Fundierung des Zusammenhangs von Intelligenzkonstrukten und Problemlöseleistung gehört mehr als der Verweis auf Definitionen. Die Suche nach einer wirklichen theoretischen Fundierung fällt jedoch ernüchternd aus.

In der Anfangszeit der Forschung zum komplexen Problemlösen überwogen Begründungen für ein Fehlen dieses Zusammenhangs. Diese basierten alle auf einem simplen Vergleich der Oberflächenanforderungen von Intelligenztests und komplexen Problemlöseszenarien, der verschiedene Unterschiede ergab. Nach Dörner und Kreuzig (1983) sind Intelligenztestaufgaben wenig komplex, wenig vernetzt, transparent, statisch, monotetisch und gut strukturiert. Dies steht ihrer Definition komplexer Probleme diametral entgegen. Nach Misiak, Haider und Kluwe (1988) ist das Zeitverhalten komplexer Systeme dynamisch (Intelligenztests: statisch), ihre Reliabilität nicht berechenbar (berechenbar), ihr Lösungsraum weit (eng), der Lösungsweg heuristisch (algorithmisch), Exploration nötig (verboten) und die Transparenz niedrig (hoch) (zitiert nach Süß, 1996). Auch die in Abschnitt 4.2 beschriebene Gegenüberstellung von Intelligenztests und Problemlöseszenarien durch Putz-Osterloh (1981) folgt diesem Argumentationsschema.

Die Argumentation verkennt, dass es bei der Konstruktion von Intelligenztests nicht das Ziel ist, kognitive Anforderungen realer Aufgaben repräsentativ abzubilden. Intelligenztests beanspruchen keine Kontentvalidität, sondern Konstruktvalidität – sie sollen eine basale kognitive Ressource messen, die für den erfolgreichen Umgang mit realen Problemen benötigt wird. Auch die Schule, das Studium und der Beruf stellen Anforderungen, die Intelligenztests nicht

---

abbilden – trotzdem sind Schul-, Studiums- und Berufsleistung gut durch Intelligenztests vorhersagbar.

Wie ist also der empirisch nachweisbare Zusammenhang zwischen Intelligenz und Problemlösefertigkeit begründbar?

Nach Süß et al (1991, S. 337) sind beim komplexen Problemlösen zunächst induktive Denkprozesse nötig, um aus den Veränderungen der Systemvariablen Regelmäßigkeiten zu extrahieren und daraus Wenn-Dann-Regeln abzuleiten. Mit Hilfe deduktiver Denkprozesse wird aus diesen Regeln dann abgeleitet, wie das System unter gegebenen Umständen optimal zu steuern ist. Induktives und deduktives Denken ist der Verarbeitungskapazität zugeordnet. Wie in den vorigen Abschnitten dargelegt wurde, ist es tatsächlich vor allem die Intelligenzfacette Verarbeitungskapazität, die den Zusammenhang von Intelligenz und Problemlösen ausmacht. Süß (1996) argumentiert außerdem damit, dass Problemlösen in hohem Maße das Arbeitsgedächtnis belastet und Problemlöseleistung deshalb durch die Kapazität des Arbeitsgedächtnisses begrenzt ist. Er zitiert dazu die ACT\*-Theorie von Anderson, die den Erwerb komplexer kognitiver Fertigkeiten als Prozess der Prozeduralisierung des Wissens durch Handeln und Üben beschreibt. In der ersten Phase dieses Lernprozesses, der *Interpretation deklarativer Wissensbestände*, wird neues deklaratives Wissen durch Erfahrung aus erfolgreichen und erfolglosen Versuchen und aus Fehlern erworben. Das Wissen ist jedoch nicht direkt in Handlungen umzusetzen, sondern muss durch bereichsunspezifische Problemlöseprozeduren wie Analogiebildung, Mittel-Ziel-Analyse und Teilzielbildung interpretiert werden. Im Rahmen dieses Prozesses muss das deklarative Wissen permanent im Arbeitsgedächtnis präsent gehalten werden. Aus der hohen Belastung des Arbeitsgedächtnis kann abgeleitet werden, dass dessen Kapazität für komplexes Problemlösen entscheidend ist. Auch das passt zu den hohen Korrelationen der Problemlöseleistung mit Verarbeitungskapazität, die ja wie in Abschnitt 6.1.2 dargelegt primär ein Maß der Arbeitsgedächtniskapazität ist.

Die zweite und dritte Phase der ACT\*-Theorie, in der das deklarative Wissen prozeduralisiert wird und dadurch nicht mehr im Arbeitsgedächtnis gehalten werden muss, wird bei der Bearbeitung komplexer Problemlöseszenarien nicht erreicht – dafür werden sie nicht lange genug bearbeitet.

Solche Ansätze einer theoretischen Fundierung des Zusammenhangs von Intelligenz und Problemlösen wurden aber leider kaum weiter ausgearbeitet.

### 6.1.5 Empirische Studien

Um den Zusammenhang von Intelligenz und komplexer Problemlöseleistung empirisch zu prüfen sollen die vorhandenen Studien, die diesen Zusammenhang untersuchen, gesammelt und ausgewertet werden. Zur Sammlung von Studien wurde die dem Autor bekannte Literatur zum komplexen Problemlösen nach Zitaten solcher Studien durchsucht. Alle gefundenen Studien wurden dann ebenfalls wieder durchsucht, bis keine Zitate mehr gefunden wurden. Zeitgleich wurde in gängigen Literaturdatenbanken und Suchmaschinen über die Stichworte Intelligenz und Problemlösen gesucht. Die Liste auf den folgenden Seiten ist das Ergebnis. Angegeben wird jeweils das verwendete Szenario, der verwendete Intelligenztest, die Stichprobengröße und der Zusammenhang. In vielen Studien wurden mehrere Zusammenhänge untersucht, diese sind getrennt unter der jeweiligen Studie aufgeführt.

Szenario	Intelligenztest	N	R
Beckmann, J.F. & Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. In P.A. Frensch & J. Funke (Eds.). <i>Complex problem solving. The European perspective</i>			
CHERRY TREE Kleines lineares System	Verarbeitungskapazität (IST und LST-3 Subskalen)	N=52	r=.09
MACHINE Keine semantische Einkleidung	Verarbeitungskapazität (IST und LST-3 Subskalen)	N=40	r=.36
Dörner, D., Kreuzig, H.W., Reither, F. & Stäudel, T. (Hrsg.): Lohhausen: Vom Umgang mit Komplexität. Huber, Bern 1983.			
LOHHAUSEN Zieloffen, Eingaben über Versuchsleiter vermittelt	Raven-APM	N=48	r=-.29
Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? <i>Diagnostica, 14</i> , 283-302.			
TAILORSHOP Gütemaß TRENDFU	Raven-APM	N=48	r=.28



Funke, J. (1985). Steuerung dynamischer System durch Aufbau und Anwendung subjektiver Kausalmodelle. <i>Zeitschrift für Psychologie</i> , 193, 443-465.			
DYNAMIS Kleinsystem mit sechs Variablen, Gütemaß Rangplatz	BIS Experimentalversion, angegeben ist nur die Korrelation mit den Zellen und nur wenn $>r=.10$ (KV: .17, KN und KF n.a.)	N=43	n. a.
Funke, J. (1988). Bedingungen und Auswirkungen der Informationssuche und -aufnahme beim Bearbeiten des komplexen Simulationssystems „TAILORSHOP“, <i>Berichte aus dem Psychologischen Institut der Universität Bonn</i> , 14, 5.			
TAILORSHOP 8032 Gütemaß Gesamtkapital	Verarbeitungskapazität BIS-Test	N=68	r=.21
Funke, U. (1992). <i>Diagnostisches interaktives System zur Komplexitätssimulation „DISKo/c“</i> . Handbuch. Fil-derstadt: Care applications Hoffmann KG (Vertrieb).			
DISKO	BIS-Test	N=17	r=.40
Gediga, G., Schöttke, H. & Tücke-Bressler, M. (1984). Problemlösen und Intelligenz. <i>Psychologische Forschungsberichte FB 8 der Universität Osnabrück</i> , Nr. 34.			
HAMURABI Gütemaß Ackerfläche	IST-70	N=14	r=.19
SERUM Gütemaß Jahre ohne Konkurs, Simulation zu schwer eingestellt	IST-70	N=12	r=.05
Hartung, S. & Schneider, I. (1995). Entwicklung und Anwendung computersimulierter Szenarien. In B. Strauß & M. Kleinmann (Hrsg.), <i>Computersimierte Szenarios in der Personalarbeit (S.219-236)</i> . Göttingen: Verlag für Angewandte Psychologie.			
UTOPIA Gütemaß Strategieauswertung	IST-70	N=54	r=.52
Hasselmann, D. (1993). <i>Computersimierte komplexe Problemstellungen in der Management-Diagnostik</i> . Hamburg: Windmühle.			
TEXTILFABRIK Gütemaß Gesamtkapital	IST-70 In der Studie werden nur Korrela- tionen mit fünf Subtests angegeben, hier wird der Mittelwert verwendet	N=21	r=.33
TEXTILFABRIK Gütemaß Gesamtkapital	IST-70 In der Studie werden nur Korrela- tionen mit vier Subtests angegeben, hier wird der Mittelwert verwendet	N=41	r=.12
Hesse, F.W. (1982). Effekte des semantischen Kontextes auf die Bearbeitung komplexer Probleme. <i>Zeitschrift für experimentelle und angewandte Psychologie</i> , 29, 62-91.			
DORI Semantisch eingekleidet, transparent	Raven-APM	N=30	r=.06
DORI Semantisch eingekleidet, intransparent	Raven-APM	N=30	r=-.17
DORI Nicht semantisch, transparent	Raven-APM	N=30	r=.38
DORI Nicht semantisch, intransparent	Raven-APM	N=30	r=.46

Hörmann, H.J. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexen Problemlösen. <i>Sprache &amp; Kognition</i> , 1, 23-31.			
TAILORSHOP Gütemaß TRENDFU, transparent	Verarbeitungskapazität BIS-Test	N=19	r=.45
TAILORSHOP Gütemaß TRENDFU, intransparent	Verarbeitungskapazität BIS-Test	N=22	r=.25
Hussy, W. (1984). Zum Begriff der Problemschwierigkeit beim komplexen Problemlösen. <i>Trierer Psychologische Berichte</i> , 11, 4.			
MONDLANDUNG transparent, 3 Eingabevariablen (leicht)	CFT-3	N=30	r=.50
MONDLANDUNG intransparent, 3 Eingabevariablen (leicht)	CFT-3	N=30	r=.35
MONDLANDUNG transparent, 5 Eingabevariablen (schwer)	CFT-3	N=30	r=-.30
MONDLANDUNG intransparent, 5 Eingabevariablen (schwer)	CFT-3	N=30	r=.25
Hussy, W. (1989). Intelligenz und komplexes Problemlösen. <i>Diagnostica</i> , 35, 1-16.			
ZAP-85 Transparent	BIS-4	N=75	r=.36
ZAP-85 Intransparent	BIS-4	N=79	r=.44
Hussy, W. (1991). Komplexes Problemlösen und Verarbeitungskapazität. <i>Sprache &amp; Kognition</i> , 10, 208-220.			
TAILORSHOP In der Studie werden Korrelationen für fünf Durchgänge angegeben, hier wird der Mittelwert verwendet	Verarbeitungskapazität BIS-Test	N=20	r=.56
Kersting, M. (1999). Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarios? Zur Kriteriumsvalidität von Problemlöseszenarios und Intelligenztests. Göttingen: Hogrefe.			
TAILORSHOP Gütemaß Gesamtkapital	Verarbeitungskapazität BIS-Test und ISA verrechnet	N=103	r=.21
DISKO Modifiziertes Gütemaß	Verarbeitungskapazität BIS-Test und ISA verrechnet	N=99	r=.26
Kröner, S. (2001). Intelligenzdiagnostik per Computersimulation. Berlin: Waxmann.			
MULTIFLUX	Raven-APM	N=28	r=.58
MULTIFLUX Simulation zu schwer eingestellt	Raven-APM	N=96	r=.30
MULTIFLUX	Verarbeitungskapazität BIS-Test	N=101	r=.65
Kühle, H.J. & Badke, P. (1986). Die Entwicklung von Lösungsvorstellungen in komplexen Problemsituationen und die Gedächtnisstruktur. <i>Sprache &amp; Kognition</i> , 5, 95-105.			
MORO Steuerung über Versuchsleiter	CFT	N=20	n. s.

Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. <i>Computers in Human Behavior</i> , 18, 685-697.			
HUNGER IN THE SAHEL Zwei Durchläufe, zweiter auf vier Experimentalgruppen aufgeteilt. Es wird die mittlere Korrelation der Experimentalgruppen und danach der Mittelwert beider Durchgänge berechnet, dieser wird verwendet. Korrelationskoeffizienten mussten dazu aus einer Grafik abgelesen werden.	Verarbeitungskapazität BIS-Test	N=200	r=.42
Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. <i>Zeitschrift für Psychologie</i> , 189, 79-100.			
TAILORSHOP Gütemaß TRENDPO, intransparent	Raven-APM	N=35	r=-.22
TAILORSHOP Gütemaß TRENDPO, transparent	Raven-APM	N=12	r=-.13
Putz-Osterloh, W., Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. <i>Zeitschrift für experimentelle und angewandte Psychologie</i> , 28, 309-334.			
TAILORSHOP Gütemaß TRENDPO, intransparent	Raven-APM	N=35	r=.00
TAILORSHOP Gütemaß TRENDPO, transparent	Raven-APM	N=33	r=.31
URGOS	Raven-APM	N=34	r=.40
Putz-Osterloh, W. (1985). Selbstreflexion, Testintelligenz und interindividuelle Unterschiede bei der Bewältigung komplexer Probleme. <i>Sprache &amp; Kognition</i> , 4, 203-216.			
MORO Zieloffen, Selbstreflexionstraining	BIS-Test	N=25	r=.36
MORO Zieloffen	BIS-Test	N=25	n. s.
Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme. <i>Zeitschrift für Psychologie</i> , 195, 63-84.			
TAILORSHOP Gütemaß Flüssigkapital	BIS-Test	N=30	r=.33
MORO Zielvorgabe	BIS-Test	N=30	n.s.
Putz-Osterloh, W., Köster, K. (1988). Diagnostik komplexer Denk- und Entscheidungsstrategien bei einem computersimulierten Planspiel. In Bundesminister für Verteidigung (Hrsg.), <i>Untersuchungen des Psychologischen Dienstes der Bundeswehr</i> , 223-255.			
TAILORSHOP Kombiniertes Gütemaß	IST	N=100	r=-.16
Reichert, U. & Dörner, D. (1988). Heuristiken beim Umgang mit einem „einfachen“ dynamischen System. <i>Sprache &amp; Kognition</i> , 7, 12-24.			
KÜHLHAUS	Verarbeitungskapazität BIS-Test	N=54	r=.31

Schoppek, W. (1996). Kompetenz, Kontrollmeinung und komplexe Probleme. Zur Vorhersage individueller Unterschiede bei der Systemsteuerung. Bonn: Holos.			
JOGI'92 Ohne Training	Verarbeitungskapazität BIS-Test	N=24	r=.74
JOGI'92 Mit Labortraining	Verarbeitungskapazität BIS-Test	N=24	r=.09
JOGI'93 Ohne Training	Verarbeitungskapazität BIS-Test	N=24	r=.63
JOGI'93 Mit Training operativer Fertigkeiten	Verarbeitungskapazität BIS-Test	N=24	r=.43
Stäudel, T. (1987). Problemlösen, Emotionen und Kompetenz. Regensburg: Roderer.			
MORO Über Versuchsleiter vermittelt, zieloffen, lautes Denken, 40 simulierte Jahre	CFT3	N=43	n. s.
Strohschneider, St. (1991). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. <i>Diagnostica</i> , 4, 353-371.			
MORO Zieloffen	Verarbeitungskapazität BIS-Test	N=25	r=.03
MORO Mit Zielvorgabe	Verarbeitungskapazität BIS-Test	N=20	r=.77
VEKTOR Mit Zielvorgabe	Verarbeitungskapazität BIS-Test	N=20	r=.60
Süß, H.M. (1996). Intelligenz, Wissen und Problemlösen. Göttingen: Hogrefe.			
TAILORSHOP Neues Problemlösegütemaß, 3 aggregierte Durchgänge	Verarbeitungskapazität BIS-Test	N=214	r=.40
TAILORSHOP Neues Problemlösegütemaß, 2 aggregierte Durchgänge	Verarbeitungskapazität BIS-Test	N=137	r=.34
TOMATEN 3 aggregierte Durchgänge	Verarbeitungskapazität BIS-Test	N=39	r=.33
Wagener, D. (2001). Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation. Lengerich; Berlin; Riga; Rom; Wien; Zagreb: Pabst Science Publishers.			
FSYS 2.0 Gütemaß korrigiertes Gesamtkapital	Verarbeitungskapazität BIS-Test	N=139	r=.36
FSYS 2.0 Gütemaß korrigiertes Gesamtkapital	ISA	N=35	r=.64
M3	Verarbeitungskapazität BIS-Test	N=90	r=.45
K4 Gütemaß Kapital alle drei Stufen	Verarbeitungskapazität BIS-Test	N=71	r=.29

---

Wittmann, W. W., Süß, H.-M., Oberauer, K. (1996). *Determinanten komplexen Problemlösens*. (Bericht 9). Mannheim: Universität Mannheim, Lehrstuhl Psychologie II.

TAILORSHOP	Verarbeitungskapazität BIS-Test	N=92	r=.48
POWERPLANT	Verarbeitungskapazität BIS-Test	N=92	r=.39
LEARN	Verarbeitungskapazität BIS-Test	N=57	r=.44

### 6.1.6 Auswertung der Studien

Insgesamt konnten 56 Korrelationskoeffizienten aus 28 Studien gefunden werden. Der Mittelwert der Korrelationskoeffizienten beträgt  $r=.31$ . Werden die Koeffizienten mit der Stichprobengröße gewichtet steigt er auf  $r=.32$ . Im Mittel zeigt sich also ein Zusammenhang zwischen Intelligenz und Problemlösen.

Die breite Varianz der Korrelationskoeffizienten von  $r=-.29$  bis  $r=.77$  lässt sich nicht allein durch Zufallsschwankungen erklären. Verschiedene Moderatoren des Zusammenhangs wurden schon vorgeschlagen und untersucht (siehe dazu auch Beckmann & Guthke, 1995):

- *Transparenz*: Transparenz soll die Simulation einem Intelligenztest annähern und so zu höheren Zusammenhängen führen. Allerdings ist Transparenz in unterschiedlichen Studien sehr unterschiedlich definiert.
- *Semantische Einkleidung*: Die Aktivierung und Nutzung von Vorwissen aufgrund der semantischen Einkleidung soll den Problemlöseprozess beeinflussen. Vorwissen tritt als weiterer wichtiger Faktor neben die reine Intelligenz. Das kann bei einer Inkompatibilität des Szenarios mit dem eigenen Vorwissen auch hinderlich sein.
- *Vorwissen*: Nach Elshout-Raaheim-Hypothese (Elshout, 1987; Raaheim, 1988) ist der Zusammenhang des Problemlöseerfolgs mit Intelligenz bei mittlerem Vorwissen maximal, bei gar keinem und sehr hohem Vorwissen dagegen nahe Null.
- *Zielspezifität*: Der Wegfall der Zielelaboration soll bei Vorgabe klarer Ziele die Aufgabe einer Intelligenztestsituation annähern und so zu höheren Zusammenhängen führen. Eine andere Erklärung ist, dass es bei Zieloffenheit einen Einfluss hat, ob die Ziele des Probanden mit denen des Entwicklers übereinstimmen, und dass das eher eine Frage der Weltanschauung als der Problemlösefähigkeit sei.

- 
- *Versuchsleitereffekte*: Wenn Informationsbeschaffung und Maßnahmeneingabe über einen Versuchsleiter geschieht kann dies zu einer Verfälschung der Messung und zu einer Konfundierung mit Persönlichkeitsmerkmalen wie Extraversion führen.
  - *Multidimensionalität*: Nicht die Allgemeine Intelligenz, sondern vor allem die Facette Verarbeitungskapazität zeigt Zusammenhänge mit der Problemlösefähigkeit. Wird in Studien nur ein Maß der Allgemeinen Intelligenz untersucht, kann das den Zusammenhang verringern.
  - *Reliabilität*: Misst das Gütemaß des Szenarios die Problemlösefähigkeit nicht reliabel, so kann auch kein Zusammenhang mit einem Außenkriterium auftreten.
  - *Schwierigkeit*: Wie schon in Abschnitt 5.3 beschrieben hat sich gezeigt, dass eine zu große Problemschwierigkeit dazu führen kann, dass nicht mehr zwischen guten und schlechten Problemlösern differenziert wird. Ein Zusammenhang mit Intelligenz kann dann auch nicht mehr auftreten.

Um diese Moderatoren statistisch zu untersuchen ist die Studienzahl zu klein. Stattdessen sollen hier die Studien im Einzelfall betrachtet werden, die der Hypothese eines Zusammenhangs zwischen Problemlösen und Intelligenz widersprechen.

### **Die LOHHAUSEN-Studie**

LOHHAUSEN simulierte auf einem Uni-Großrechner eine Kleinstadt mit mehr als 2000 Variablen. Das sehr komplexe Variablengefüge, dessen Realitätsnähe angezweifelt werden darf, war für die Teilnehmer intransparent. Von einer hohen Schwierigkeit kann also ausgegangen werden. Ein klares Ziel wurde nicht vorgegeben. Informationen mussten vom Versuchsleiter eingeholt werden, Maßnahmen wurden über ihn vermittelt.

Dazu kommen weitere Besonderheiten dieser Studie. Am Ende jeder Sitzung konnte der Proband frei entscheiden, wie viele Zeittakte er Lohhausen ohne Eingriffe laufen lassen wollte. Am Ende sollten so insgesamt zehn Jahre simuliert werden. Während der Programmsteuerung wurden zahlreiche neue Variablen einprogrammiert, um auf Maßnahmen von Probanden zu reagieren.

Alles zusammengenommen zeigt recht deutlich, dass die Reliabilität des Szenarios bezweifelt werden darf. Dafür spricht auch, dass ein Versuchsleiter-Urteil über die Intelligenz des Pro-

---

banden am Ende der Programmsteuerung sehr wohl positiv mit der Intelligenz korrelierte ( $r=.25$  mit den Raven-APM,  $r=.31$  mit dem CFT, zitiert nach Hasselmann, 1993).

### **Die Studien von Putz-Osterloh**

Die Studien von Putz-Osterloh nutzten nahezu alle die Simulation TAILORSHOP. Die Bearbeitung wurde über einen Versuchsleiter vermittelt. Erschwerend kommt hinzu, dass die Versuchspersonen zu lautem Denken aufgefordert wurden.

Funke (1983) erklärt die ausbleibenden Korrelationen in den beiden Studien von 1981 außerdem damit, dass das verwendete Gütemaß TRENDPO im Gegensatz zu dem von ihm vorgeschlagenen Gütemaß TRENDFU nicht reliabel ist.

Süß, Kersting & Oberauer (1991) erklären ausbleibende Korrelationen dagegen mit der zu hohen Schwierigkeit des Programms. Sie hatten in ihrer Studie auch mit TRENDFU keinen Zusammenhang finden können. Das lässt sich damit erklären, dass die Oberstufenschüler aus ihrer Studie stärker überfordert waren als die Studenten aus Funkes Untersuchung. Der Befund von Putz-Osterloh, dass sich unter Transparenz sehr wohl ein Zusammenhang zwischen TRENDPO und Intelligenz ergibt, erklärt sich vor diesem Hintergrund damit, dass die Vorgabe einer Vernetzungsgrafik die Schwierigkeit der Programmsteuerung senkt.

Unklar bleibt, warum in der Studie von 1987 ein Zusammenhang auch unter Intransparenz gefunden werden konnte.

In den Studien zum System MORO (Putz Osterloh, 1985; Putz Osterloh, 1987) wird leider nur angegeben, dass der Zusammenhang zu Intelligenz nicht signifikant war, über die Höhe wird keine Aussage gemacht. Das ist aufgrund der geringen Teststärke der Studien ärgerlich. Es wäre durchaus denkbar, dass so ein mittlerer Effekt verdeckt wird.

### **Die Studie von Strohschneider**

Strohschneider (1991) fand für das System MORO keinen Zusammenhang mit Intelligenz, wenn kein klares Ziel vorgegeben wurde. Bei klarer Zielvorgabe fand er einen hohen Zusammenhang. Dies unterstützt die Annahme der Zielspezifität als Moderator des Zusammenhangs zwischen Problemlöseleistung und Intelligenz. Dafür spricht auch, dass Stäudel (1987) mit einer zieloffenen Version von MORO ebenfalls keinen signifikanten Zusammenhang zwischen Problemlöseleistung und Intelligenz findet.

---

Dagegen spricht, dass Putz-Osterloh (1985) MORO zieloffen verwendet und einen Zusammenhang von  $r=.36$  mit Intelligenz findet. Dies könnte wiederum auf die Effekte des vorher durchgeführten Selbstreflexionstrainings zurückzuführen sein, bei dem die Versuchspersonen explizit angewiesen wurden, sich jede Runde die Frage „Habe ich mir ein klares Ziel gesetzt?“ zu stellen. Interessant wäre die Höhe der Korrelation ohne Selbstreflexionstraining, die von Putz-Osterloh aber nur als nicht signifikant angegeben wird. Nicht signifikant war auch die Korrelation beim zielspezifischen Einsatz von MORO (Putz-Osterloh, 1987).

Eine andere mögliche Erklärung des Befundes von Strohschneider könnte in der Struktur des MORO-Systems liegen. MORO ist so programmiert, dass es sich viele Monate scheinbar stabil verhält um dann relativ unvermittelt in schwere Krisen (Rinderkatastrophe, Grundwasserkatastrophe, Bevölkerungskatastrophe) abzurutschen, die schwer vorherzusehen sind. Diese Krisen werden durch zu starkes Wachstum ausgelöst. In der zielspezifischen Version von Strohschneider werden konkrete Zielwerte für sechs wichtige Variablen vorgegeben, die weder unter- noch überschritten werden sollen. Katastrophen werden bei Einhaltung dieser Ziele verhindert. Denkbar wäre, dass diese Katastrophen den Unterschied zwischen guten und schlechten Problemlösern überdecken. In der Studie von Strohschneider würde dieser Effekt noch dadurch verstärkt, dass die Simulation nach 20 Monaten noch einmal für 10 Monate weiterläuft, ohne dass der Problemlöser gegensteuern kann. In der Studie von Stäudel läuft MORO sogar für 40 simulierte Jahre. In der Studie von Putz-Osterloh werden dagegen nur 20 Jahre gespielt, wodurch später auftretende Katastrophen nicht ausgelöst werden.

Welche der beiden Erklärungen zutrifft, kann auf Basis der vorliegenden Daten nicht entschieden werden.

### **Die Studien von Hesse und Beckmann & Guthke**

In der Studie von Hesse (1982) bearbeiten insgesamt 60 Probanden das Szenario DORI. 60 Probanden bearbeiten das gleiche System, allerdings sind die Variablen nur mit Buchstaben bezeichnet. Beckmann & Guthke (1995) ließen 52 Personen das Szenario CHERRY TREE bearbeiten, ein semantisch eingekleidetes Kleinsystem mit sechs Variablen. 40 Personen bearbeiteten das gleiche System, allerdings ohne semantische Einkleidung als drei Regler und drei Displays einer Maschine. In beiden Studien ergibt sich nur für das nicht semantisch eingekleidete System eine Korrelation zwischen Intelligenz und Steuerungsleistung.



---

Diese Befunde sind schwer erklärbar. An der semantischen Einkleidung an sich kann es nicht liegen, dafür liegen mittlerweile zu viele Studien mit semantisch eingekleideten Systemen vor, die einen Zusammenhang zeigen.

Bei der Studie von Hesse könnte der Grund wieder im System liegen. DORI ist ein Subsystem von MORO und ähnlich katastrophengefährdet. Die Auswertungen von Hesse zeigen, dass die Probanden in der nichtsemantischen Gruppe viel vorsichtiger agierten als die Probanden der semantischen Gruppe, und dementsprechend ein geringeres Wachstum erzielten. Dadurch wurde ihre Leistung geringer eingeschätzt, sie lösten aber womöglich auch nicht die Katastrophen aus, die die Reliabilität beeinträchtigen.

Beckmann und Guthke berichten, dass beide Gruppen kein systemspezifisches Vorwissen hatten. Die semantische Gruppe erwarb auch während der Steuerung kein relevantes Wissen über das Szenario, die nichtsemantische hingegen schon. Trotzdem ist die Leistung beider Gruppen gleich.

Auf der Basis der vorhandenen Daten können auch diese Befunde nicht erklärt werden.

Neben diesen Studien ist noch auf die von Gediga, Schöttke und Tücke-Bressler (1984) hinzuweisen, die erneut zeigen, wie wichtig eine adäquate Problemschwierigkeit für den Zusammenhang zu Intelligenz ist. Das von ihnen eingesetzte Szenario SERUM war nach eigenen Angaben deutlich zu schwer für die Probanden. Das Resultat ist eine Korrelation mit Intelligenz von nur  $r=.05$ .

Werden die TAILORSHOP-Studien von Putz-Osterloh, die Studien mit dem zieloffenen MORO-System, LOHHAUSEN und SERUM nicht in die Mittelwertberechnung einbezogen ergibt sich eine mittlere Korrelation zwischen Intelligenz und Problemlöseleistung von  $r=.37$  (gewichtet  $r=.37$ ). Die ausgeschlossenen Studien ergeben ungewichtet und gewichtet einen gemittelten Zusammenhang von  $r=.00$ .

---

## 6.2 Der Zusammenhang mit Wissen

Die in den Abschnitten 4.4 und 4.6 beschriebenen Befunde zum Zusammenhang von Wissen und Problemlösen bestätigen die Theorien, die komplexes Problemlösen relativ übereinstimmend als Wissenserwerb konzeptualisieren. So enthält beispielsweise Dörners Prozessmodell des komplexen Problemlösens eine „Phase der Modellbildung und Informationssammlung“ (Dörner, 1989, S.67ff). Für Funke (1992) ist komplexes Problemlösen der Erwerb subjektiver Kausalmodelle, bei dem nacheinander die Existenz, Richtung und Stärke von Relationen zwischen Variablen entdeckt werden. Der Zusammenhang von Wissen und Problemlösen wird aber auch ins Feld geführt, um die Existenz einer komplexen Problemlösefähigkeit anzuzweifeln, da der Problemlöseerfolg vollständig durch Intelligenz und Wissen erklärbar sei.

Süß (1996) unterscheidet dabei Sachwissen und Handlungswissen. Sachwissen definiert er als Wissen über Sachverhalte in einem Realitätsausschnitt, Handlungswissen als Wissen über adäquate Handlungen bzw. kognitive Operationen in einer Situation im Hinblick auf ein Ziel. Sachwissen muss zur erfolgreichen Steuerung eines Systems zuerst in Handlungswissen transformiert werden. Zum Sachwissen über ein komplexes Problemlöseszenario gehört nach Süß Wissen über die Variablen (Variablenwissen), die Verknüpfungen zwischen Variablen (Relationenwissen) sowie deren Richtung und Stärke, die zusammengenommen das Kausalmodell eines Systems ausmachen (Funke, 1992). Sachwissen über ein komplexes Problemlöseszenario wird in dieser Arbeit Systemwissen genannt.

### 6.2.1 Vorwissen und erworbenes Wissen

Bei einer Interpretation der Studienergebnisse aus den Abschnitten 4.4 und 4.6 ist es zunächst wichtig, zwischen Vorwissen und erworbenem Systemwissen zu unterscheiden. Vorwissen besitzt der Proband, bevor er mit dem Szenario in Berührung kommt. Es wird bei Kontakt mit dem Szenario durch dessen semantische Einkleidung aktiviert und „löst beim Problemlöser eine Reihe von Vorstellungen, Erwartungen, Hypothesen, Bewertungen und Prozessen aus, die einen Einfluss auf seine Vorgehensweise bei der Problembearbeitung bewirken“ (Hesse, 1982). Nach den Theorien zum komplexen Problemlösen erwirbt der Problemlöser dann im

---

Laufe der Systembearbeitung *neues* systemspezifisches Wissen, das ihm dabei hilft, das Szenario erfolgreich zu steuern. Der Erwerb dieses neuen Wissens ist abhängig von einer Kompetenz zum komplexen Problemlösen. Differenzen gibt es bei der Frage, von welcher Kompetenz der Wissenserwerb abhängt. Die einen sehen hier ein neues Konstrukt „Komplexe Problemlösefähigkeit“ oder „Operative Intelligenz“, die anderen sehen den Wissenserwerb vor allem von Intelligenz abhängig.

Kersting (1999) ließ seine Probanden zunächst den TAILORSHOP bearbeiten, danach bearbeiteten sie einen systemspezifischen Wissenstest. Eine solche Abfolge ist zweifelhaft, wenn man untersuchen will, inwiefern Intelligenz und Wissen die Problemlöseleistung *vorhersagen*. Das gemessene Wissen könnte durchaus abhängig von einer „Komplexen Problemlösefähigkeit“ während der Programmsteuerung erworben worden sein.

Wittmann, Süß und Oberauer (1991) sowie Wittmann und Süß (1999) setzen den systemspezifischen Wissenstest deshalb *vor* der Programmsteuerung ein. Es ist unwahrscheinlich, dass die Probanden in den zwei Übungsmonaten, die vor dem Wissenstest stattfanden, relevantes Wissen gesammelt haben. Die Ergebnisse dienen als Beleg, dass Intelligenz und systemspezifisches Vorwissen die Problemlöseleistung sehr gut vorhersagen. Das systemspezifische Vorwissen klärt dabei inkrementell Varianz auf.

Die Schlussfolgerung, dass außer Intelligenz und Vorwissen keine Kompetenz am Problemlöseprozess beteiligt ist, muss trotzdem als vorschnell kritisiert werden. Auch Vorwissen wird erworben, und auch dabei spielt mit großer Wahrscheinlichkeit eine kognitive Kompetenz eine Rolle. Der Zusammenhang von Vorwissen und Problemlöseleistung ließe sich auch über diese kognitive Kompetenz erklären. Plausibilität gewinnt diese Annahme durch die Befunde von Wittmann, Süß und Oberauer (1996) und Kersting (1999). In ihren Studien korrelierte die Problemlösefähigkeit in einem Szenario zwar mit dem zugehörigen systemspezifischen Wissenstest, aber in nahezu gleicher Höhe auch mit den szenariospezifischen Wissenstests anderer Problemlöseszenarien. Das in den Wissenstests zu diesen anderen Szenarien erfasste Wissen war für die Steuerung jedoch gar nicht erforderlich. Auch die Befunde von Wagener (2001) sind hier aufschlussreich. Er setzte neben einem zur semantischen Einkleidung seines Szenarios FSYS 2.0 passenden Vorwissenstests auch einen Geschichtswissenstest ein. Dieser erwies sich in der Auswertung als bester Prädiktor des Problemlöseerfolgs. Und schließlich

---

weist Süß (1996) selbst darauf hin, dass die Induktion von Vorwissen die Steuerungsleistung nicht verbessert. In der Wiederholungsuntersuchung seiner Studie erhielten N=48 Probanden die Möglichkeit, die Vernetzungsgrafik des TAILORSHOP für 30 Minuten zu studieren. Die N=89 restlichen Probanden hatten diese Möglichkeit nicht. Der direkt nach diesem Treatment durchgeführte systemspezifische Wissenstest zeigte, dass die 48 Probanden dadurch tatsächlich über hochsignifikant mehr deklarierbares Systemwissen verfügten. Sie schnitten in der anschließenden Steuerung des Systems trotzdem nicht besser ab. Dieser Befund zeigt, dass es nicht direkt auf das Vorwissen ankommt, sondern auf die Kompetenz, die hinter dem Vorwissenwerb steht.

Anstatt einen direkten Einfluss spezifischen Systemvorwissens anzunehmen kann also von einer generellen Fähigkeit ausgegangen werden, die

1. Den Erwerb von (Vor-)Wissen in allen Bereichen beeinflusst und
2. Vermittelt über den Erwerb von Systemwissen während der Problembearbeitung auch auf den Problemlöseerfolg wirkt.

Der Zusammenhang von Vorwissen und Steuerungsleistung stellt sich dann als „Scheinkorrelation“ dar, da beide Variablen auf eine Drittvariable zurückzuführen sind. Dies muss nicht unbedingt die postulierte „Operative Intelligenz“ sein. Denkbar wäre auch, dass die klassische Intelligenz diese Rolle einnimmt. Die zusätzliche Varianzaufklärung des Wissens kann in diesem Fall entweder auf nicht kognitive Merkmale wie Interesse und Motivation zurückgeführt werden oder aber auf szenariospezifische Faktoren, die nichts mit einem generellen Konstrukt zu tun haben. Die Forschung ist leider noch nicht eindeutig.

## **6.2.2 Implizites Wissen**

Die Oxforder Arbeitsgruppe um Donald Broadbent (z.B. Berry & Broadbent 1984, 1987, 1988; Broadbent et al, 1986) fand in ihren Studien keinen Zusammenhang zwischen der in Kleinsystemen mit zwei bis vier Variablen gemessenen Problemlöseleistung und dem Systemwissen über diese Szenarien. Außerdem verbesserte die Vermittlung von Systemwissen die Steuerungsleistung nicht. Die Autoren zogen daraus den Schluss, dass die Probanden implizites Wissen erwerben, also nicht verbalisierbares und nicht bewusstseinsfähiges Wissen.

---

Implizites Wissen wird passiv erworben: Das implizite Lernen ist unbewusst, der Lernprozess nicht gerichtet oder selektiv. Als Resultat kann der Proband praktisch zeigen, was er weiß, das solcherart Gekonnte aber nicht verbalisieren.

Replikationsstudien sprechen eher gegen diese Schlussfolgerung. Insbesondere zeigte sich, dass das von Broadbent über Multiple-Choice-Fragen erhobene Systemwissen gar nicht vollständig zur Steuerung des Systems benötigt wurde (Haider, 1991). Es reichte ein einfacher Algorithmus. Weiterhin liegen mittlerweile verschiedene Studien vor, die sehr wohl einen Zusammenhang zwischen deklarierbarem Systemwissen und Steuerungsleistung zeigen.

Ein fehlender Zusammenhang zwischen Problemlöseleistung und Systemwissen liegt demnach wohl eher an unreliaiblen Maßen der Problemlöseleistung oder des Systemwissens als an impliziten Lernprozessen.



---

## 7. DER ZUSAMMENHANG MIT BERUFSLEISTUNG

Soll ein komplexes Problemlöseszenario in der Eignungsdiagnostik eingesetzt werden, reicht es nicht, Zusammenhänge mit Intelligenz und Wissen nachzuweisen. Wie schon in Abschnitt 4.1.2 dargelegt wurde muss gezeigt werden, dass die im Problemlöseszenario gezeigte Fähigkeit auch mit Berufsleistung zusammenhängt. Dieses Thema wurde in der Vergangenheit leider stark vernachlässigt.

Das größte Problem bei einer solchen Validierung ist, das Kriterium Berufsleistung reliabel zu messen. „Harte“ Kriterien wie Gehalt und Position können als Indikator der generellen Berufsleistung verwendet werden. Ob sie diese allerdings wirklich messen hängt davon ab, ob der interne Aufstieg etwas mit Leistung zu tun hat. Dazu müssen die internen Personalauswahlverfahren, die zur Laufbahnplanung verwendet werden, valide sein. Letztlich sind Gehalt und Position also wieder von „weichen“ Faktoren wie Vorgesetztenbeurteilung und interner Personalauswahl abhängig. Zusätzliche Probleme bereitet es, wenn das zu validierende Verfahren selbst Teil der Laufbahnplanung ist. Wie bei einer sich selbst erfüllenden Prophezeiung ist das Kriterium dann direkt vom eigentlich zu validierenden Prädiktor abhängig.

Neben „harten“ Daten werden oft Vorgesetztenbeurteilungen als Kriterium verwendet. Dies hat den Vorteil, dass nicht nur die generelle Berufsleistung bewertet werden kann, sondern auch Teilbereiche wie die berufliche Problemlöseleistung. Das Problem von Vorgesetztenbeurteilungen ist die zweifelhafte Reliabilität. Die Vergleichbarkeit der Bewertungen ist dadurch eingeschränkt, dass die Personen unterschiedliche Vorgesetzte und damit unterschiedliche Beurteiler haben. Beurteilungsfehler wie Sympathie, Milde und Strenge sind kaum auszuschließen, Mehrfachbeurteilung findet selten statt.

Die dritte Möglichkeit ist, Instrumente an Assessment Centern zu validieren. Diese Verfahren sind als realitätsnahe und damit inhaltsvalide Simulationen des Berufsalltags konzipiert. Der Vorteil ist, dass alle Probanden auf Basis der gleichen Situationen bewertet werden. Außerdem sind die Beobachter geschult, um Beurteilungsfehler zu minimieren, und es findet eine Mehrfachbeurteilung statt. Es kann also von einer im Vergleich zu Vorgesetztenbeurteilungen höheren Reliabilität und Vergleichbarkeit ausgegangen werden. Zudem besteht die Chance, in passenden Übungen explizit das Kriterium berufliche Problemlöseleistung zu erfassen. Dieser

---

Ansatz steht und fällt allerdings mit der Vergleichbarkeit von Leistungen im Assessment Center und realer Berufsleistung, die schwer zu überprüfen ist. Denn zur Prüfung muss wiederum auf die weiter oben angezweifelten „harten“ Daten oder Vorgesetztenbeurteilungen zurückgegriffen werden.

Im folgenden sollen Studien besprochen werden, die den Zusammenhang komplexer Problemlöseszenarien mit Maßen des Berufserfolgs untersuchen.

### **Der Zusammenhang mit Vorgesetztenbeurteilungen**

- Die größte und methodisch beste Studie zum Zusammenhang mit Vorgesetztenbeurteilungen stammt von Kersting (1999) und wurde in Abschnitt 4.7 beschrieben. N=104 Polizisten bearbeiteten den TAILORSHOP und das Szenario DISKO. Für N=73 Polizisten wurden durchschnittlich ein Jahr und siebeneinhalb Monate später Vorgesetztenbeurteilungen der Berufsleistung erhoben. Der Zusammenhang mit dem Vorgesetztenurteil zur beruflichen Problemlöseleistung betrug  $r=.37$  für den TAILORSHOP und  $r=.33$  für DISKO. Es ergab sich außerdem inkrementelle Validität über Intelligenz hinaus. Die vom Vorgesetzten eingeschätzte Kooperationsfähigkeit konnte dagegen nicht vorhergesagt werden, und ein Gesamturteil der Berufsleistung wurde nicht untersucht.
- U. Funke validierte das Problemlöseszenario DISKO im Rahmen einer Studie zur Personalauswahl in Forschung und Entwicklung, die N=117 Personen aus unterschiedlichen Unternehmen untersuchte (Schuler et al., 1995). Die Korrelation mit der konkurrent durch Vorgesetzte eingeschätzten allgemeinen Berufsleistung betrug  $r=.09$ , die Korrelation mit der beruflichen Problemlöseleistung  $r=.12$ . Für eine selektierte Subgruppe mit hohen kognitiven Anforderungen, die aus N=59 Industrieforschern und Ingenieuren bestand, ergab sich ein Zusammenhang von  $r=.44$  mit der beruflichen Problemlöseleistung (U. Funke, 1995). Der Zusammenhang mit allgemeiner Berufsleistung wurde für diese Subgruppe nicht angegeben.
- Hartung & Schneider (1995) berichten über einen Zusammenhang von  $r=.48$  zwischen der Strategiewertung des Szenarios UTOPIA und dem Vorgesetztenurteil von N=54 Personen. Mehr Informationen werden leider nicht gegeben.
- In Abschnitt 4.6 wurde die Studie von Hasselmann (1993) schon vorgestellt. Bei N=21 Führungsnachwuchskräften ergab sich ein Zusammenhang der TEXTILFABRIK mit der



---

Position zwei Jahre nach der Bearbeitung von  $r=.49$ , mit dem Gehalt von  $r=.53$ . Es ergab sich ein hoher Zusammenhang mit dem vom Personalbetreuer eingeschätzten Potenzial von  $r=.49$ . Der Zusammenhang mit der eingeschätzten tatsächlichen Arbeitsleistung ( $r=.16$ ) und dem eingeschätzten aktuellen Nutzen des Mitarbeiters ( $r=.09$ ) war dagegen gering. Das Problem der Studie ist die geringe Stichprobengröße.

- Hasselmann (1995) berichtet außerdem über die prädiktive Validierung des Szenarios BRENNSTOFFVERTRIEB an  $N=25$  Führungsnachwuchskräften. Der Zusammenhang mit der Position zwei Jahre nach der Bearbeitung betrug hier  $r=.12$ , der mit dem Gehalt  $r=.30$ . Mit der Vorgesetztenbeurteilung des Potenzials hing die Steuerungsleistung zu  $r=.41$  zusammen, mit der Beurteilung der aktuellen Leistung zu  $r=.30$  und des Nutzens zu  $r=.37$ . Auch hier ist das Problem die kleine Stichprobe.
- Obermann (1991) ließ  $N=23$  Gruppenleiter das Szenario AIRPORT bearbeiten. Konkurrent wurde über ein einzelnes Item die Ausprägung des beruflichen Problemlöseverhaltens im Vergleich zu anderen Mitarbeitern erhoben. Es ergab sich ein Zusammenhang von  $r=.55$ . Ein Maß der allgemeinen Berufsleistung wurde dagegen nicht erfasst. Auch hier lässt sich die geringe Stichprobengröße kritisieren.

### **Der Zusammenhang mit Assessment Center Leistungen**

- Hasselmann (1993) berichtet auch Zusammenhänge des Szenarios TEXTILFABRIK mit Assessment Center Leistungen von  $N=21$  Führungsnachwuchskräften. Zusammenhänge ergaben sich für die Kriterien Problemlösefähigkeit ( $r=.45$ ), Entscheidungsfähigkeit ( $r=.39$ ), Überzeugungsfähigkeit ( $r=.26$ ) und Belastbarkeit ( $r=.38$ ). Korrelationen um Null ergaben sich für die Kriterien Teamgeist, Kontaktfähigkeit und Verbaler Ausdruck. Ein Zusammenhang mit dem Gesamtergebnis wird nicht berichtet.
- Hartung & Schneider (1995) berichten über einen Zusammenhang von  $r=.28$  zwischen der Strategieauswertung des Szenarios UTOPIA und dem Assessment Center Gesamtergebnis von  $N=54$  Teilnehmern von neun Assessment Centern, die in einem Elektrokonzern durchgeführt wurden.
- Wagener (2001) berichtet über Zusammenhänge der Leistung im Szenario FSYS 2.0 mit kognitiv geprägten Assessment Center Übungen.  $N=69$  Vertriebsmitarbeiter bearbeiteten im Rahmen einer Personalentwicklungsmaßnahme das Szenario und eine komplexe Fall-

---

studie in schriftlicher Einzelarbeit. Es ergaben sich Korrelationen mit den vier Auswertungsskalen von  $r=.27$ ,  $r=.19$ ,  $r=.28$  und  $r=.19$ .  $N=35$  Teilnehmer einer Personalentwicklungsmaßnahme in einem Anlagenbau-Unternehmen bearbeiteten in Kleingruppen eine komplexe Fallstudie und präsentierten sie dann. Es ergab sich eine Korrelation zwischen FSYS 2.0 und dem Gesamtscore von  $r=.27$ , mit dem Kriterium Methodenkompetenz ergab sich sogar eine Korrelation von  $r=.41$ . Diese Stichprobe bearbeitete außerdem eine Postkorbübung, deren Gesamtscore zu  $r=.28$  mit der FSYS-Leistung korrelierte. Auch  $N=51$  Bewerber um eine kaufmännische Ausbildung bearbeiteten FSYS und eine Postkorbübung. Der Zusammenhang zwischen Gesamtscore und FSYS-Leistung betrug hier  $r=.39$ .

Es muss darauf hingewiesen werden, dass vielfach der Zusammenhang mit dem Kriterium berufliche Problemlöseleistung berichtet wird, nicht mit Berufsleistung im Allgemeinen. Validitäten von  $r=.30$  bis  $r=.40$  müssen vor diesem Hintergrund interpretiert werden.

Die geringe Studienzahl zur Kriteriumsvalidität von komplexen Problemlöseszenarien mit zum Teil deutlich zu kleinen Stichproben macht jedenfalls deutlich, dass in diesem Bereich großer Bedarf an weiteren Validierungsstudien besteht.

---

## 8. INTEGRATION DER BEFUNDE UND ABLEITUNG VON HYPOTHESEN

Als Abschluss des theoretischen Teils sollen die geschilderten Befunde in einem sehr einfachen Arbeitsmodell integriert werden, auf dessen Basis dann inhaltliche Hypothesen aufgestellt werden.

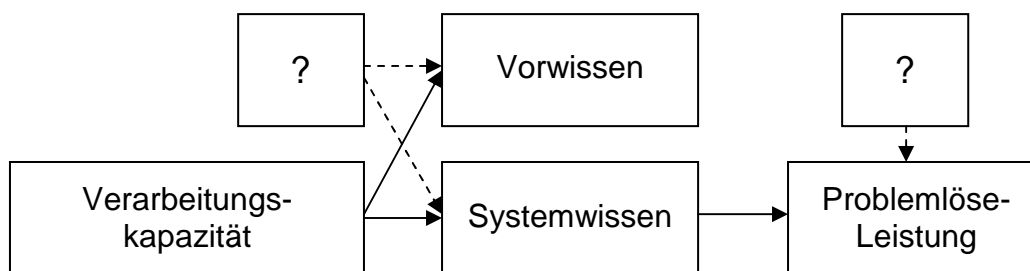


Abbildung 8.1. Arbeitsmodell zum Zusammenhang von Intelligenz, Wissen und Problemlösen

Die zentrale Wichtigkeit der Konstrukte Intelligenz und Wissen für die Problemlöseleistung wurde in den vorangegangenen Abschnitten immer wieder betont und belegt. Es kann davon ausgegangen werden, dass es eine Fähigkeit zum Erwerb von deklarativem Wissen in komplexen Situationen gibt. Die Befunde zeigen weiterhin, dass es sich dabei vor allem um die Arbeitsgedächtniskapazität handelt, die die simultane Kurzzeitspeicherung und Verarbeitung von Information ermöglicht. Diese wiederum kann mit der in klassischen Intelligenztests gemessenen Verarbeitungskapazität gleichgesetzt werden. Im Arbeitsmodell wird also davon ausgegangen, dass das während der Systemsteuerung erworbene Systemwissen von der Verarbeitungskapazität abhängt. Weiterhin wird davon ausgegangen, dass dieses erworbene Systemwissen die Problemlöseleistung determiniert.

Das Arbeitsmodell entspricht damit in wesentlichen Punkten der klassischen Theorie von E. L. Thorndike, die Hunter (1986) auch als Basis seines Kausalmodells der Berufsleistung vorstellt. Im Modell von Hunter (1986) wird davon ausgegangen, dass Intelligenz auch über das erworbene berufsspezifische Wissen hinaus auf die Arbeitsleistung wirkt. Als Grund dafür wird vor allem genannt, dass erworbenes Wissen flexibel an neue Situationen angepasst werden muss. Diese Kompetenz ist in Problemlöseszenarien weniger wichtig: Neue Situationen entstehen praktisch nur aufgrund der Schwankung bekannter Variablen in bekannten Grenzen. Anders

---

als in realen komplexen Situationen ist es in einem Problemlöseszenario möglich, das System komplett zu durchschauen. Deshalb wird in dieser Arbeit davon ausgegangen, dass die Verarbeitungskapazität nur über das Systemwissen vermittelt auf die Problemlöseleistung wirkt. Und schließlich wird davon ausgegangen, dass es keinen direkten kausalen Einfluss allgemeinen Vorwissens auf die Problemlöseleistung gibt. Die Gründe dafür wurden in Abschnitt 6.2.1 ausgeführt. Stattdessen wird angenommen, dass auch der Erwerb allgemeinen Vorwissens durch die Verarbeitungskapazität beeinflusst wird, der Zusammenhang von Vorwissen und Problemlöseleistung also eine „Scheinkorrelation“ ist.

Nach dieser Grobkonzeption ist interessant, ob es sich bei den postulierten Verknüpfungen um monokausale Beziehungen handelt. Dann würde die Verarbeitungskapazität sämtliche systematische Varianz im erworbenen systemspezifischen Wissen aufklären, das wiederum sämtliche systematische Varianz der Problemlöseleistung aufklären sollte. Bleibt systematische Varianz unaufgeklärt, muss an zwei Stellen über weitere Konstrukte nachgedacht werden. Zum einen könnte es ein Konstrukt geben, das neben der Verarbeitungskapazität zum Wissenserwerb beiträgt. Zum anderen ist denkbar, dass Systemwissen und die zum Erwerb von Systemwissen nötige Verarbeitungskapazität allein nicht reichen, um ein System erfolgreich zu steuern, sondern dass es eine davon unabhängige Fähigkeit zur Anwendung von Systemwissen gibt. An beiden Stellen könnte ein neues Konstrukt „Operative Intelligenz“ ins Spiel gebracht werden. Es könnte sich aber auch um nicht kognitive Merkmale wie Motivation und Interesse handeln.

Über das Arbeitsmodell hinaus wird davon ausgegangen, dass die Problemlöseleistung mit Berufsleistung zusammenhängt. In Abschnitt 3.3 wurde ausgeführt, dass es auch im Beruf von zentraler Bedeutung ist, berufsspezifisches Wissen zu erwerben und erfolgreich einzusetzen. Auch der Erwerb berufsspezifischen Wissens hängt letztlich vor allem von der Intelligenz bzw. genauer von der Arbeitsgedächtniskapazität ab. Abbildung 8.2 verdeutlicht diesen Zusammenhang.

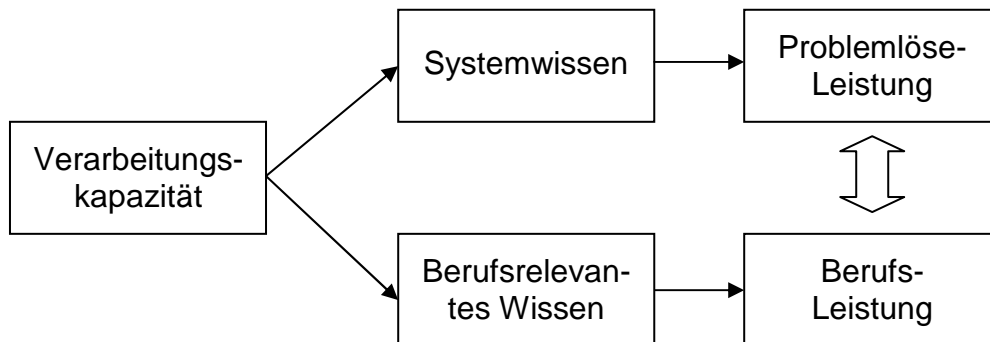


Abbildung 8.2. Zusammenhang von Problemlöseleistung und Berufsleistung

Der Zusammenhang von Problemlöseleistung und Berufsleistung resultiert demnach daraus, dass für beide Leistungen Verarbeitungskapazität eine wichtige Rolle spielt und der Prozess, nämlich Leistung durch Wissenserwerb, sehr ähnlich ist.

Natürlich ist dieser theoretische Hintergrund eher schwach. Ein sauberer empirischer Nachweis ist deshalb sehr wichtig und zentraler Bestandteil dieser Validierungsstudie.

Die zentralen Hypothesen, die sich aus dem Arbeitsmodell ergeben, sind:

- H1: Verarbeitungskapazität hat einen Einfluss auf erworbenes Systemwissen.
- H2: Erworbenes Systemwissen hat einen Einfluss auf die komplexe Problemlöseleistung.
- H3: Verarbeitungskapazität wirkt nur vermittelt über das erworbene Systemwissen auf die Problemlöseleistung.
- H4: Problemlöseleistung und Berufsleistung hängen zusammen.

Weiterhin ist für diese Arbeit interessant, ob es neben der Verarbeitungskapazität weitere Konstrukte gibt, die einen unabhängigen Einfluss auf die Problemlöseleistung haben. In früheren Studien wurden in diesem Zusammenhang vor allem Persönlichkeitsmerkmale, Motivation und Interessen untersucht.

Unter den Persönlichkeitsmerkmalen hat die Gewissenhaftigkeit eine herausragende Stellung, da sie wie in Abschnitt 3.1 berichtet den höchsten Zusammenhang mit Berufsleistung aufweist. Zudem scheint sie, wie in Abschnitt 3.3 beschrieben, auch einen Einfluss auf das erworbene berufsspezifische Wissen zu haben. Schmidt & Hunter (1998) begründen diesen Zu-

---

sammenhang damit, dass gewissenhafte Menschen mehr Zeit mit ihrer eigentlichen Aufgabe verbringen und deshalb mehr lernen. Trifft diese Begründung zu, ist kein Zusammenhang von Gewissenhaftigkeit mit komplexer Problemlöseleistung zu erwarten: Komplexe Problemlöseszenarien erfassen maximale Leistung, nicht typische Leistung. Es ist davon auszugehen, dass auch wenig gewissenhafte Menschen sich die gesamte Bearbeitungszeit dem Problemlöseszenario widmen. Ganz ähnlich lässt sich bei anderen Persönlichkeitsmerkmalen, Motivation und Interessen argumentieren. Sie alle beeinflussen eher die typische Leistung als die maximale Leistung. Selbst einem unmotivierten und uninteressierten Teilnehmer wird es möglich sein, sich für wenige Stunden zusammenzureißen. Ein Zusammenhang über die allgemeine Intelligenz hinaus ist deshalb kaum zu erwarten.

---

## 9. DIE UNTERSUCHUNG

Ziel dieser Untersuchung ist es, den Zusammenhang der mit AGRIMAN gemessenen Problemlöseleistung mit Berufsleistung und beruflicher Problemlöseleistung zu untersuchen. Operationalisiert wird Berufsleistung und berufliche Problemlöseleistung über Assessment Center Ergebnisse. N=185 Personen bearbeiteten im Rahmen von Assessment Centern das Szenario AGRIMAN.

### 9.1 Stichproben

Es gab drei unterschiedliche Arten von Assessment Center, die sich vor allem in den eingesetzten Übungen unterschieden. Die unterschiedlichen Verfahren definieren drei unterschiedliche Stichproben.

**Stichprobe I** besteht aus N=53 Personen (24 männlich, 29 weiblich), die an drei dreitägigen Assessment Centern eines Großunternehmens teilnahmen. Die Stichprobe umfasste sowohl Teilnehmer mit Ausbildung als auch Hochschulabsolventen.

In dieser Stichprobe wurde eine verkürzte AGRIMAN-Variante ohne zweite Diagnosephase eingesetzt.

**Stichprobe II** besteht aus N=74 Personen (33 männlich, 41 weiblich), die an sechs zweitägigen Development Centern eines Großunternehmens teilnahmen. Die Stichprobe umfasste sowohl Teilnehmer mit Ausbildung als auch Hochschulabsolventen.

Im letzten Verfahren mit N=10 Teilnehmern wurde eine verkürzte AGRIMAN-Variante ohne Handlungswissenstest und ohne zweite Diagnosephase eingesetzt.

*Datenausfälle:* Im ersten Verfahren konnten N=5 Teilnehmer aus Zeitgründen den Wissenstest nicht beenden. Im vierten Verfahren fielen aufgrund von Motivationsproblemen N=2 Teilnehmer beim Wissenstest und bei der zweiten Diagnosephase aus.

---

**Stichprobe III** besteht aus N=58 (30 männlich, 28 weiblich) studentischen Teilnehmern an drei zweitägigen Assessment Center Trainings. Die Stichprobe umfasste ausschließlich Hochschulabsolventen und Doktoranden.

Im dritten Verfahren mit N=12 Teilnehmern wurde eine verkürzte AGRIMAN-Variante ohne zweite Diagnosephase eingesetzt.

Bei allen drei Verfahrenstypen wurde strikt auf die Einhaltung methodischer Standards geachtet. Die Beobachter durchliefen vor dem Verfahren ein eintägiges Beobachtertraining. Jede Übung wurde von mindestens zwei Beobachtern beobachtet. Die Kriterien waren den Teilnehmern vor den Übungen bekannt, jedes Kriterium war durch konkrete Verhaltensbeispiele verankert. Beobachtung und Bewertung waren getrennt. Die Bewertung fand auf einer vierstufigen Skala statt und wurde bei Abweichungen zwischen den Beobachtern in einer Beobachterkonferenz diskutiert.

## 9.2 Das Szenario AGRIMAN

Die komplexe Problemlöseleistung wird über das Problemlöseszenario AGRIMAN operationalisiert. AGRIMAN wurde vom Autor für den Einsatz in der Eignungsdiagnostik entwickelt und programmiert. Dabei wurden einander widersprechende Ziele verfolgt. Zum einen sollten die in Abschnitt 5.3 genannten Punkte erfüllt werden, um die Reliabilität und damit die Voraussetzung für Validität zu optimieren. Zum anderen sollte das Szenario eine reichhaltige semantische Einkleidung haben, um die hohe Akzeptanz komplexer Szenarien nicht in Gefahr zu bringen.

Kröner (2001) entwickelte als eine Art Gegenentwurf das Szenario MULTIFLUX, das zur Optimierung der Reliabilität auf eine semantische Einkleidung und voneinander abhängige Zeittakte verzichtet. Was die Messgenauigkeit und den Zusammenhang mit Intelligenz angeht ist dieses Szenario ein Fortschritt. Es ist einem Intelligenztest allerdings auch schon so ähnlich, dass die Akzeptanz ähnlich gering sein dürfte.



---

## 9.2.1 Inhaltliche Einkleidung und Oberfläche

AGRIMAN simuliert einen kleinen landwirtschaftlichen Betrieb. Aufgabe des Probanden ist es, auf verschiedenen Feldern Pflanzen in optimaler Saatstärke anzubauen, mit Hilfe von Düngern die Nährstoffe auf einem für die angebauten Pflanzen optimalen Niveau zu halten sowie mit Hilfe von Pestiziden die Schädlinge zu bekämpfen, die den angebauten Pflanzen schaden und die sich je nach angebauter Pflanze unterschiedlich vermehren. Diese Einkleidung wurde gewählt, um Vorwissenseffekte zu vermeiden. Die Zahl der Felder sowie die Zahl der Pflanzen, Dünger, Pestizide, Nährstoffe und Schädlinge und deren Bezeichnung kann in Parameterdateien frei gewählt werden. Die Zahl der Variablen ist also selbst variabel. Auch die Anzahl der Zeittakte und die Zeitbegrenzung ist frei einstellbar. Es können beliebig viele Phasen mit unterschiedlichen Parametereinstellungen hintereinander gespielt werden. Abbildung 9.1 zeigt die generelle Verknüpfung der Variablen.

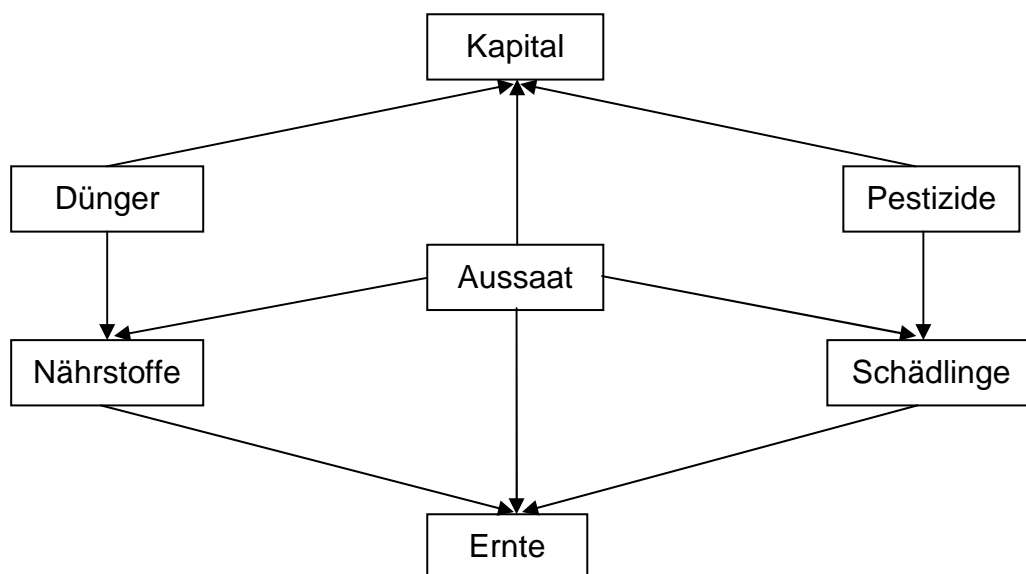


Abbildung 9.1. Verknüpfungsmodell der AGRIMAN-Variablen

Bei Betrachtung des Verknüpfungsmodells ist es wichtig zu berücksichtigen, dass es mehrere Pflanzen, Dünger, Nährstoffe, Pestizide und Schädlinge gibt.

---

Beim Design der Oberfläche und insbesondere der Kommandoübermittlung mussten gegenläufige Ziele berücksichtigt werden. Zum einen sollte sie gemäß Abschnitt 5.3 so einfach wie möglich zu bedienen sein. Zum anderen sollte gezielte Informationssuche Teil der Anforderung der Simulation sein, und das Programm sollte die Strategie des Probanden so umfassend wie möglich protokollieren. Dies ist bei der Darstellung auf einem Bildschirm nicht möglich: Das Programm kann nicht wissen, welche der dargestellten Informationen der Benutzer überhaupt registriert. Die gewählte Oberfläche ist ein Kompromiss: Zwar bleibt der Bildschirm bei AGRIMAN zum Großteil der gleiche, Systeminformationen werden aber nicht automatisch angezeigt. Vielmehr müssen Informationen zum Ertrag des Feldes, zum Nährstoffgehalt im Boden und zur Anzahl der Schädlinge getrennt aufgerufen werden, und zwar für jedes Feld. Dieser Informationsaufruf ist aber sehr einfach und intuitiv über zwei Mausklicks möglich, so dass er der Steuerung nicht im Wege steht. Abbildung 9.2 zeigt einen Screenshot der Oberfläche.

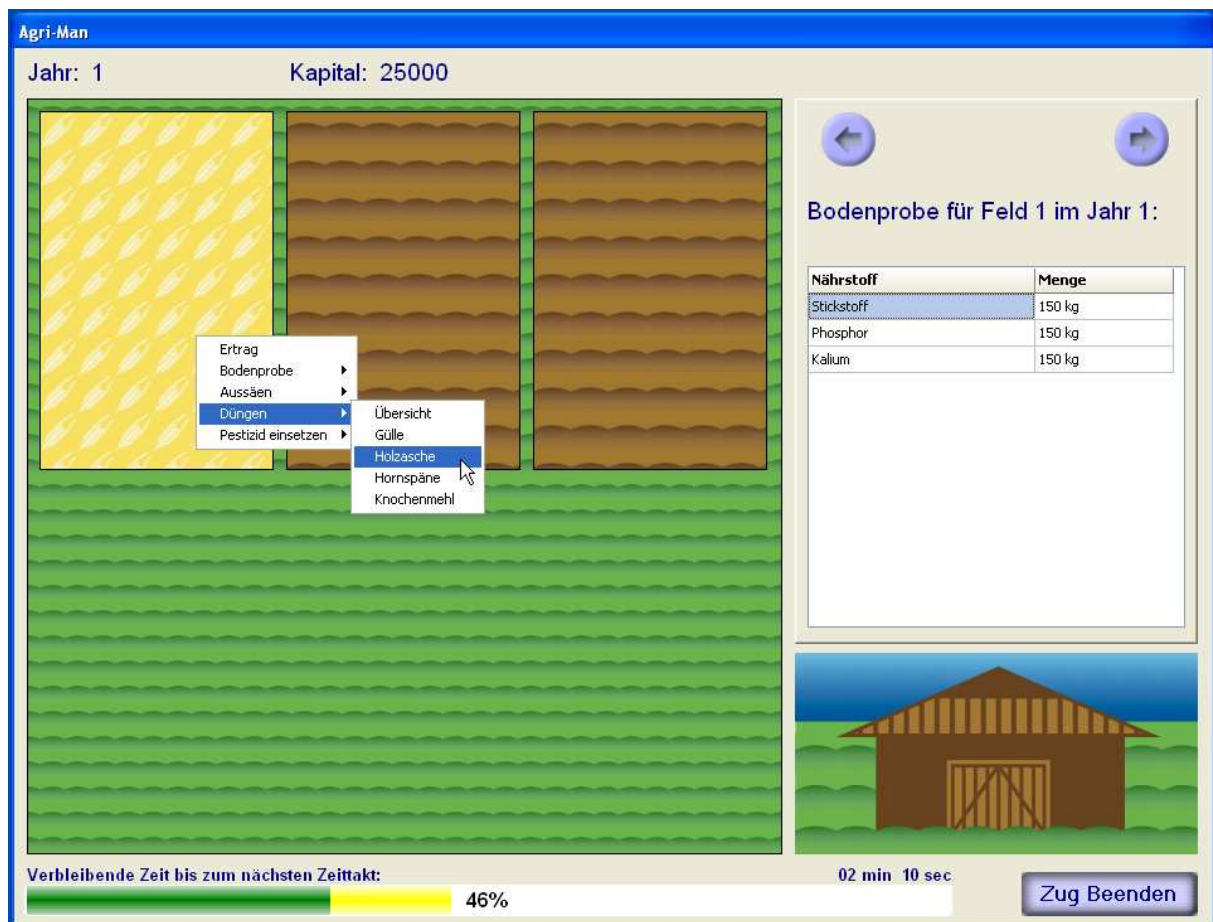


Abbildung 9.2. Die Oberfläche von AGRIMAN

Auf der linken Seite sind grafisch die Felder dargestellt. In diesem Fall sind es drei. Über die Parameterfiles kann eine beliebige Menge festgelegt werden. Rechts unten ist grafisch die Scheune dargestellt. Und rechts oben ist das Aktionsfeld zu sehen. Während des Spiels bleibt diese Oberfläche durchgehend bestehen. Es muss nicht auf andere Bildschirme umgeschaltet werden und es öffnen sich auch keine zusätzlichen Fenster. Sämtliche Aktionen lassen sich über Kontextmenüs steuern, die sich beim Klick auf eines der Felder oder auf die Scheune öffnen. Je nach im Menü ausgewählter Aktion ändert sich dann das Aktionsfeld, so dass dort die Aktion abgeschlossen werden kann. Um ein Feld zu düngen, klickt der Anwender beispielsweise einfach auf das entsprechende Feld. Es öffnet sich ein Kontextmenü. In diesem fährt er über den Menüpunkt „Düngen“ und wählt den gewünschten Dünger aus. Im Aktionsfeld wird er nun gefragt, mit wie viel Kilogramm des Düngers er das Feld düngen möchte. Alles was er noch tun muss, ist, die gewünschte Menge in das Eingabefeld einzugeben und

---

mit einem Klick auf „OK“ zu bestätigen. Unterhalb des Eingabefeldes erscheint dann eine entsprechende Bestätigung.

## **9.2.2 Wissensdiagnostik**

Auf die große Bedeutung des Erwerbs von Systemwissen wurde in Kapitel 6 eingegangen. Dieser sollte durch einen fest im Programm verankerten Wissenstest Rechnung getragen werden. Für den Fall, dass impliziter Wissenserwerb eine Rolle spielt, sollte auch dieser erfasst werden.

AGRIMAN bietet deshalb zwei unterschiedliche Wissenstests an. Der eine erfasst deklarierbares Wissen, während der andere näher an der Simulation bleibt und so Handlungswissen erfassen soll.

### **Der deklarative Wissenstest**

Der deklarative Wissenstest versucht, das Systemwissen der Probanden über Ratingskalen zu erfragen. Von Interesse sind dabei die Zusammenhänge der Variablen, z.B. wie viel jedes Nährstoffs ein bestimmter Dünger enthält oder wie effektiv Pestizide gegen unterschiedliche Schädlinge wirken. Die Items sind so formuliert, dass der Nutzer aufgefordert ist, sich eine konkrete Situation vorzustellen, z.B. dass alle Felder auf ihrem Ausgangszustand sind und nichts angebaut wird. Dann wird er gebeten, die Folgen einer konkreten Handlung einzuschätzen, z.B. dass 1000 Kilogramm Gülle auf die Felder ausgebracht wird. Er soll auf einer Skala von 0 bis 10 einschätzen, wie stark sich die Nährstoffkonzentration verändert. In Abbildung 9.3 ist ein Screenshot dieses Wissenstests zu sehen.

**Agri-Man Wissenstest**

Item 2/6: Stellen sie sich vor, dass auf ihrem Feld alle Werte auf ihrem Ausgangszustand sind.

Würden sie nun die optimale Menge GETREIDE aussäen, welche Auswirkungen hätte das auf die Nährstoffe im Boden? \*

Stickstoff	<input type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input checked="" type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Phosphor	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Kalium	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

Würden sie nun die optimale Menge LEGUMINOSEN aussäen, welche Auswirkungen hätte das auf die Nährstoffe im Boden? \*

Stickstoff	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Phosphor	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Kalium	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

Würden sie nun die optimale Menge RAPS aussäen, welche Auswirkungen hätte das auf die Nährstoffe im Boden? \*

Stickstoff	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Phosphor	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Kalium	<input checked="" type="radio"/> Weiß nicht.	<input type="radio"/> -5	<input type="radio"/> -4	<input type="radio"/> -3	<input type="radio"/> -2	<input type="radio"/> -1	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

\*) -5 bedeutet, dass die Pflanze dem Boden sehr viel des Nährstoffs entzieht. 0 bedeutet, der Nährstoff verändert sich nicht. +5 bedeutet, dass sich durch Anbau der Pflanze der Nährstoffgehalt sehr stark erhöht.

Zurück Verbleibende Zeit: 09 min 32 sec 5% Weiter

Abbildung 9.3. Screenshot des AGRIMAN Wissenstests

Um den Wissenstest auszuwerten wird vom Programm berechnet, welche tatsächlichen Auswirkungen die beschriebenen Handlungen hätten. Danach wird innerhalb eines Bereichs (z.B. Auswirkung von Düngern auf Nährstoffe) jede Skala mit jeder anderen verglichen. Stimmt die Rangordnung der tatsächlichen Auswirkung mit der vom Probanden eingeschätzten Auswirkung überein, gibt es einen Punkt. Die erzielbaren Punkte entsprechen der Fakultät der Anzahl der Ratingskalen. Die tatsächliche Punktzahl wird für jeden Bereich an der möglichen Gesamtpunktzahl relativiert und danach der Mittelwert der Bereiche berechnet. Es ergibt sich ein Wert zwischen 0 und 1.

### Der Handlungswissenstest

Im Handlungswissenstest werden keine Ratingskalen verwendet. Stattdessen werden spezifische Situationen fest vorgegeben, auf die der Nutzer in optimaler Weise reagieren soll. Diese

---

Situationen können in Parameterfiles frei definiert werden. So ist es beispielsweise möglich, ein Feld vorzugeben, auf dem schon eine bestimmte Pflanze ausgesät wurde, auf dem aber bestimmte Nährstoffe fehlen und Schädlinge Überhand genommen haben. Der Proband soll darauf dann mit dem richtigen Dünger und dem richtigen Pestizid reagieren. Die Auswertung kann auf zwei Arten erfolgen. Zum einen können Punkte für die richtige Wahl gegeben werden. Zum Anderen kann natürlich auch das erzielte Kapital ausgewertet werden. Ersteres erfasst vor allem das Wissen über Variablenrelationen. Zweiteres erfasst auch die Stärke der Relationen, ist aber auch fast identisch mit der Simulation selbst.

### **9.2.3 Einordnung in die TAKS**

Dietrich Wagener (2001) entwickelte die „Taxonomie komplexer Szenarios“ (TAKS) zur möglichst detaillierten und intersubjektiv nachvollziehbaren Beschreibung komplexer Szenarios. Sie soll helfen, Studien leichter in einen größeren Zusammenhang einzuordnen. Da AGRIMAN ein neu entwickeltes Szenario ist, soll es hier anhand dieser Taxonomie beschrieben werden.

Die Eigenschaften des zu beschreibenden Szenarios werden dazu in zwei Ebenen dargestellt. Die erste, hierarchisch höhere Ebene, unterscheidet sechs Eigenschaftsbereiche, so genannte Merkmalsknoten:

1. Inhaltliche Einkleidung
2. Bedienungsinterface
3. Formale Systemstruktur
4. Funktionale Systemstruktur
5. Informationsangebot
6. Auswertung

Jeder Merkmalsknoten umfasst in zweiter Ebene eine Reihe von Einzelmerkmalen. Anhand dieser Merkmale soll das komplexe Szenario „AGRIMAN“ beschrieben werden.

---

### **Inhaltliche Einkleidung**

Auswahl des Inhaltsbereichs	Landwirtschaftlicher Betrieb: Pflanzen säen, Boden düngen um Nährstoffe optimal zu halten, Schädlinge mit Pestiziden bekämpfen.
Akzeptanz von Inhaltsbereich und Aufgabe	Wird hoch vermutet. Dünger und Pestizide können als biologisch beschrieben werden.
Breite und Tiefe des nötigen Vorwissens	Gering. Sogar die Kenntnis einfachster Beziehungen wie die, dass Dünger die Nährstoffe im Boden verändern, wird im Programm erklärt. Die spezifische Wirkung einzelner Pflanzen, Dünger, Nährstoffe, Pestizide und Schädlinge soll dagegen gar nicht bekannt sein sondern während der Steuerung herausgefunden werden.
Impact des Vorwissens	Kann ganz ausgeschaltet werden, indem nur Phantasienamen für Pflanzen, Dünger, Nährstoffe, Pestizide und Schädlinge verwendet werden. Werden realistische Beziehungen verwendet um die Akzeptanz nicht zu gefährden, kann Vorwissen einen Einfluss haben. Ist beispielsweise bekannt, wie viel Prozent eines Nährstoffs ein Dünger enthält, ist dies ein Vorteil. Die Verbreitung solchen Vorwissens in der Zielgruppe dürfte jedoch gegen Null gehen.

### **Bedienungsinterface**

Kommandoübermittlung	Direkt ohne Vermittlung über einen Versuchsleiter.
Informationsbeschaffung	Direkt ohne Vermittlung über einen Versuchsleiter.
Organisation der Interaktion	Informationen zum Ertrag, zu den Nährstoffen im Boden, zu den Schädlingen müssen getrennt voneinander und für jedes Feld einzeln abgefragt werden, um eine umfassende Protokollierung des Verhaltens zu ermöglichen. Dies geht allerdings sehr einfach und intuitiv.
Eingriffsmodus	Eingabe von Anweisungen, welche Pflanze in welcher Stärke angebaut werden soll, welche Dünger in welcher Menge ausgebracht werden sollen und welche Pestizide in welcher Menge eingesetzt werden sollen.
Ergonomie der Benutzeroberfläche	Hoch. Alle Aktionen werden in der gleichen Art und Weise mit der Maus eingeleitet und abgeschlossen.
Motivationswirkung der Darbietung	Hoch. Die Benutzeroberfläche wurde funktional und dennoch grafisch ansprechend gestaltet.
Notizen	Den Versuchspersonen wird erlaubt und empfohlen, Notizen zu machen. Wird dies einmal vergessen, können jederzeit alle Informationen aus früheren Zeittakten abgerufen werden.

Instruktionsmethode	Empfohlen wird eine mündliche Einführung, um die Akzeptanz zu erhöhen. Ein standardisierter Leitfaden dafür liegt vor. Ein vollautomatisches Lernprogramm steht ebenfalls zur Verfügung.
Bearbeitung durch Gruppen	AGRIMAN wird von Einzelpersonen gesteuert.
<b>Formale Systemstruktur</b>	
Anzahl der Variablen	Je nach Szenario variabel. Für jedes Feld werden Nährstoffe und Schädlinge gespeichert, außerdem welche Pflanze der Nutzer in welcher Stärke aussät und welche Dünger und Pestizide er in welchem Ausmaß verwendet.
Anzahl der Verknüpfungen	Je nach Szenario variabel. Die grundsätzlichen Verknüpfungen zeigt Abbildung 9.1
Verknüpfungsarten	Gewichtete Linearfunktionen, Weibullfunktionen, Logistische Funktionen
Manipulierbarkeit	Hoch. Nährstoffe und Schädlinge lassen sich über Dünger und Pestizide direkt beeinflussen, und diese haben einen direkten Einfluss auf die Ernte.
Taktzahl und Bearbeitungszeit	Je nach Szenario variabel. Möglich ist eine globale Zeitbeschränkung, innerhalb der beliebig viele simulierte Jahre gespielt werden können. Außerdem kann eine feste Zahl von simulierten Jahren, die jeweils individuell zeitbegrenzt werden können, festgelegt werden. Keine Echtzeit.
<b>Funktionale Systemstruktur</b>	
Vorwissenskompatibilität	Hoch, wenn entsprechend eingestellt. Wo in dieser Studie reale Bedingungen simuliert werden, basiert diese Simulation auf agrarwissenschaftlichen Forschungsergebnissen. Es werden verschiedene Dinge nicht simuliert, wie z.B. das Wetter oder unterschiedliche Erntemaschinen. Darauf wird in der Einführung ausdrücklich hingewiesen.
Umfang und Reduzierbarkeit der Regeln	Die Regelmenge ergibt sich aus der Anzahl der Pflanzen, Dünger, Nährstoffe, Pestizide und Schädlinge. Eine Reduktion der Regeln ist durch Konzentration auf eine einzelne Pflanze möglich, allerdings wird es dadurch schwieriger, Bestleistungen zu erreichen. „Faustregeln“ reichen zur Steuerung nicht aus.
Eigendynamik	Es kann frei eingestellt werden, wie schnell Nährstoffe sich wieder ihrem Normalwert annähern. Dies entspricht einer negativen Eigendynamik, die jedoch schwach ist. Die Schädlinge wachsen logistisch.
Zeitverzögerung	Kommt nicht vor.



Nebenwirkungen	Keine direkten. Die Reduktion der Nährstoffe im Boden durch den Anbau von Pflanzen könnte vielleicht dazugehören.
Nötiger Planungshorizont	Gering. Die richtige Düngermenge ist jederzeit für den nächsten Zug berechenbar. Bei der Pestizidmenge ist zu beachten, die Schädlinge möglichst auf einem niedrigen Niveau zu halten, bei dem sie sich noch nicht schnell vermehren. Es ist möglich, das eigene Ergebnis durch die Nutzung von Fruchtfolgen zu steigern.
Reversibilität	Hoch. Durch gezielten Einsatz von Düngern und Pestiziden lässt sich das System innerhalb eines Zuges wieder zu Optimalbedingungen steuern. Einzige Ausnahme liegt vor, wenn die Nährstoffe sich durch entsprechende Pflanzen extrem erhöht haben. Das ist relativ unwahrscheinlich, die Umkehr dauert dann aber zwei bis drei Zeittakte.
Übermächtige Entscheidungen	Es gibt keine Entscheidungen, die allein den Erfolg massiv beeinflussen.
Zufallseinflüsse	Keine. Dies wird den Probanden auch mitgeteilt.
Antagonisten, Fallen, Kompensationen	Keine.
Schwierigkeitsverlauf	Es ist möglich, unterschiedlich schwierige Phasen hintereinander zu hängen. Dies wird normalerweise aber nicht getan.
Strukturelle Redundanz	Die Felder sind bei AGRIMAN strukturell redundante Teilsysteme, die nach den gleichen Regeln funktionieren. Der relative Anteil dieser Redundanz ist hoch, außer den Feldervariablen gibt es nur noch das Gesamtkapital.
<b>Informationsangebot</b>	
Anteil trivialer Regeln	Gering. Saatgut, Dünger und Pestizide kosten Geld, Ernte bringt Geld, Dünger wirken auf Nährstoffe, Pestizide wirken auf Schädlinge.
Information über Regeln	Je nach Szenario variabel. Standard ist, dass nur die trivialen Regeln erklärt werden bzw. im Szenario abrufbar sind und alles andere während der Steuerung entdeckt werden muss.
Information über Zustände	Sämtliche Variablen sind transparent und werden numerisch exakt rückgemeldet. Lediglich die Rückmeldung über zu hohe oder zu niedrige Nährstoffkonzentration erfolgt verbal und lässt nicht auf das Ausmaß der Abweichung von Optimalzustand zurückschließen.

---

Information über Trends	Zu jeder Variable bis auf das Kapital lassen sich durch wenige Mausklicks die Werte aus vergangenen Zeiten ablesen. Grafische Trendinformationen sind nicht verfügbar.
Information über Erfolgskriterien	Den Probanden wird mitgeteilt, dass ihr einziges Ziel die Maximierung des Kapitals ist.
Verfälschung von Zustandsinformationen	Keine.
Assistentenfunktion	Keine.
<b>Auswertung</b>	
Ebene	Aktuell wird die Problemlösefähigkeit auf der allgemeinsten Ebene als Erreichung des Steuerungsziels gemessen.
Methode	Automatisch.

---

### 9.2.4 Parameter in dieser Untersuchung

AGRIMAN kann über Parameter fast beliebig verändert werden. In diesem Abschnitt sollen die konkreten Parameter dieser Untersuchung vorgestellt werden. Alle drei Stichproben bearbeiteten die gleiche Simulation.

AGRIMAN lief in dieser Untersuchung in mehreren Phasen ab.

1. *Lernphase*: Die Einführung in die Programmsteuerung erfolgte anhand eines standardisierten Leitfadens durch den Organisator des Verfahrens und dauerte 30 Minuten.
2. *Experimentalphase*: Nach der Lernphase hatten die Teilnehmer 45 Minuten Zeit, um Wissen über das System zu sammeln. Das Kapital war in dieser Phase unwichtig, es gab sechs Felder und es durften so viele Zeittakte gespielt werden wie gewünscht. In dieser Phase hatten die Teilnehmer die Möglichkeit, ein Feld jederzeit wieder auf seine Ausgangsdaten zurückzusetzen, um kontrollierte Experimente durchzuführen.
3. *Diagnosephase*: Anschließend waren acht Zeittakte zu absolvieren, alle mit einer Zeitbegrenzung von vier Minuten. Drei Felder sollten bestellt werden, die natürlich nicht mehr auf ihre Ursprungswerte zurückgesetzt werden konnten. Das Startkapital betrug 25.000.

- 
4. *Handlungswissenstest*: Auf die Diagnosephase folgten acht Mini-Szenarios, in denen jeweils ein Feld mit spezifischen Ausgangsbedingungen vorgegeben wurde, auf die in optimaler Weise zu reagieren war. Für jede Situation waren anderthalb Minuten Zeit.
  5. *Wissenstest*: Schließlich wurde der deklarative Wissenstest vorgegeben, für den die Probanden zehn Minuten Zeit hatten.
  6. *Diagnosephase II*: In den Stichproben II und III folgte auf den Wissenstest noch eine zweite Diagnosephase, die mit der ersten identisch war.

Die verwendeten Variablen waren in allen Phasen und Stichproben die gleichen.

- Es gab drei Nutzpflanzen: Getreide, Leguminosen und Raps.
- Es gab drei Nährstoffe: Stickstoff, Phosphor und Kalium.
- Es gab vier Dünger: Gülle, Holzasche, Hornspäne, und Knochenmehl.
- Es gab zwei Schädlinge: Admiralsspinne und Grinlaus.
- Es gab drei Pestizide: Xotal, Chlomitran und Arcenid.

Die Wirkung der Nutzpflanzen, Nährstoffe und Dünger entsprach grob den Bedingungen in der Realität. Die Schädlinge und Pestizide sind Phantasieprodukte, eine realistische Abbildung wäre hier zu komplex gewesen. Das Simulationsmodell wurde auf Basis einer Voruntersuchung mit N=27 Studenten auf eine mittlere Schwierigkeit optimiert.

### **9.3 Weitere Instrumente**

Im Rahmen der Assessment Center wurde neben dem Szenario AGRIMAN eine ganze Reihe von Übungen und Fragebogenverfahren eingesetzt, die im Folgenden beschrieben werden sollen.

---

### 9.3.1 Planspielbasierte Gruppendiskussion

Die zentrale Übung zur Messung beruflicher Problemlösekompetenz stellt das komplexe Unternehmensplanspiel COSMETICA dar. Das Computerprogramm simuliert umfassend und realitätsnah ein mittelständisches Kosmetikunternehmen, das in die Bereiche Entwicklung, Marketing und Produktion untergliedert ist. Es ist die Aufgabe der Spieler, ein Produkt zu entwickeln, Arbeiter einzustellen und Maschinen zu kaufen, durch geschickte Lohnpolitik die Motivation aufrechtzuerhalten, ggf. die Mitarbeiter fortzubilden, auf unterschiedlichen Kanälen Werbung für unterschiedliche Zielgruppen zu machen und durch richtige Festlegung des Produktpreises möglichst gewinnbringend zu verkaufen. Es können Lagerhäuser gebaut, Kredite aufgenommen und Unternehmensberater engagiert werden. All das wird mit über 3000 Variablen simuliert, wodurch die Simulation zu den hochkomplexen Szenarios gerechnet werden kann.

COSMETICA wird in Gruppen von sechs Personen gesteuert. Die Unternehmensdaten können von allen auf einem schwenkbaren Flachbildschirm eingesehen werden. Zusätzlich werden zu Beginn jedes Zeittakts automatisch die aktuellen Unternehmensdaten ausgedruckt. Gesteuert wird die Simulation von einem der Gruppenmitglieder auf einem Laptop. Aufgabe der Gruppe ist es, gemeinsam eine Strategie zu entwickeln und das Unternehmen möglichst gewinnbringend zu steuern. Abstimmung, Kooperation und Koordination sind hierbei entscheidend. COSMETICA wird je nach Verfahren in mehreren getrennt voneinander bewerteten Phasen bearbeitet. Zwischen den Phasen werden die Management-Teams neu zusammengesetzt, außerdem wechseln teilweise die Beobachter.

Während der Simulation wurden die Teilnehmer von mehreren Beobachtern beobachtet und auf den drei Verhaltensdimensionen „Strategisches Denken und Handeln“, „Führungskompetenz“ und „Teamorientierung“ bewertet.

COSMETICA wurde in Stichprobe I in drei Phasen bearbeitet, in Stichprobe III in zwei Phasen. Die ersten beiden Phasen waren dabei identisch.

---

### **9.3.2 Fallstudienbasierte Gruppendiskussion**

Stichprobe II bearbeitete statt des Unternehmensplanspiels eine komplexe Fallstudie. Vor der Diskussion erhielt jeder Teilnehmer knapp 50 Seiten mit Informationen, die zunächst in Einzelarbeit durchgearbeitet wurden. Darauf folgte eine Gruppendiskussion mit sechs Teilnehmern, in der auf Basis der Informationen Entscheidungen getroffen werden mussten. Diese Diskussion lief in zwei Phasen ab, wobei zwischen den Phasen die Teams neu zusammengesetzt wurden und die Beobachter wechselten.

Während der Diskussion wurden die Teilnehmer von mehreren Beobachtern beobachtet und auf den Verhaltensdimensionen „Strategisches Denken und Handeln“, „Ergebnisorientierung“, „Engagement“ und „Kommunikation“ bewertet.

### **9.3.3 Interview**

In den Stichproben I und II wurden neben Assessment Center Übungen auch teilstrukturierte Interviews durchgeführt. Die Fragen wurden dabei durch einen Interviewleitfaden vorgegeben, es mussten aber nicht unbedingt alle Fragen in der vorgegebenen Reihenfolge gestellt werden. Von Interesse für diese Untersuchung ist vor allem Stichprobe II, da in diesem Interview das Kriterium „Strategisches Denken und Handeln“ erfragt wurde.

### **9.3.4 Rollenspiele**

In allen drei Stichproben wurden Rollenspiele durchgeführt, die hier nicht genauer beschrieben werden. Sie dienen zur diskriminanten Validierung des Problemlöseszenarios.

### **9.3.5 Intelligenztests**

In Stichprobe III wurden zwei Intelligenztests zur Erfassung der Verarbeitungskapazität durchgeführt. Dabei fielen drei Teilnehmer aus.

---

Der erste eingesetzte Test ist der Berliner Intelligenzstrukturtests, Form 4 (Jäger et al, 1997). Auf das Berliner Intelligenzstrukturmodell wurde in Abschnitt 6.1.1 schon eingegangen. Von Interesse war für diese Arbeit nur die Intelligenzfacette Verarbeitungskapazität, daher wurden nur die Items dieser Facette vorgegeben. Außerdem wurde die Kurzform des Tests eingesetzt, die die Aufgaben mit der höchsten Trennschärfe verwendet.

Als zweites Maß diente der IQ-Test Online (IQO), der vom Autor als ökonomisches Maß der Verarbeitungskapazität entwickelt wurde. Die Verarbeitungskapazität wird über sechs Itemtypen (Matrizen, Figurenalogien, Zahlenreihen, Schätzen, Wortanalogien, Tatsache/Meinung) sehr ähnlich wie im BIS-4 erfasst. Die interne Konsistenz dieses Instruments beträgt  $\alpha=.9$ . Die Korrelation mit der Facette Verarbeitungskapazität aus dem BIS-4 betrug in einer Validierungsstudie  $r=.78$ .

### **9.3.6 Persönlichkeitsfragebogen**

In den Stichproben II und III kam der Persönlichkeitsfragebogen im Internet (PFI) zum Einsatz. Dieser erfasst auf bis zu 49 Skalen für den Berufserfolg als relevant erachtete Persönlichkeitsmerkmale, Motive und Interessen. In Abhängigkeit von den Anforderungen, die ein bestimmter Beruf stellt, kann daraus ein beliebiges Subset ausgewählt werden.

Der PFI wurde nicht faktorenanalytisch konstruiert, die Skalen stehen daher auch nicht orthogonal zueinander. Moderate Interkorrelationen wurden bei der Konstruktion und Itemselektion bewusst in Kauf genommen, um eine breite Anwendbarkeit und hohe Auflösung zu gewährleisten. Die interne Konsistenz der Skalen liegt durchgängig über  $\alpha =.80$ . Die Skalen korrelieren in erwarteter Weise mit bis zu  $r=.39$  mit Assessment Center Leistungen.

## **9.4 Statistische Hypothesen**

Die in Kapitel 7 vorgestellten inhaltlichen Hypothesen sollen nun in statistische Hypothesen zu konkreten Operationalisierungen übersetzt werden.

- 
- Die Verarbeitungskapazität bzw. Arbeitsgedächtniskapazität wird über die entsprechende Facette im Intelligenztest BIS-4 sowie den Intelligenztest IQO operationalisiert. Für ein allgemeines Maß der Verarbeitungskapazität werden beide Testwerte z-standardisiert und gemittelt.
  - Das erworbene systemspezifische Wissen wird über den deklarativen Wissenstest des Szenarios AGRIMAN operationalisiert.
  - Die komplexe Problemlöseleistung wird über das Endkapital im Szenario AGRIMAN operationalisiert.
  - Die Berufsleistung wird über die durchschnittliche Bewertung aller durchgeführten Übungen in den Assessment Centern bzw. Development Centern operationalisiert.

Für eine Validierungsstudie ist weniger die reine Signifikanz eines Zusammenhanges interessant, als vielmehr die absolute Höhe. Deshalb sollen hier vor allem Mutmaßungen über die Höhe der zu erwartenden Zusammenhänge aufgestellt werden.

- H1: Ausgehend von den in den Abschnitten 4.4 und 4.6 geschilderten Studienergebnissen ist ein Zusammenhang über  $r=.50$  zwischen der gemessenen Verarbeitungskapazität und dem Ergebnis im deklarativen Wissenstest zu erwarten. Jeder geringere Zusammenhang würde darauf hindeuten, dass neben der Verarbeitungskapazität weitere zentrale Konstrukte beteiligt sind.
- H2: Wiederum ausgehend von den in den Abschnitten 4.4 und 4.6 geschilderten Ergebnissen ist auch hier ein Zusammenhang über  $r=.50$  zu erwarten. Wiederum würde ein geringerer Zusammenhang darauf hindeuten, dass Systemwissen allein nicht reicht, um ein System erfolgreich zu steuern.
- H3: In einer multiplen Regression sollte die Verarbeitungskapazität keine systematische Varianz in der Steuerungsleistung über das Systemwissen hinaus aufklären.
- H4: Ein Zusammenhang über  $r=.30$  ist zu erwarten, wenn AGRIMAN als valides Instrument der Personaldiagnostik eingesetzt werden soll.

Neben den vier Haupthypothesen sollen hier noch Überlegungen zu drei weiteren Zusammenhängen angestellt werden. Diese zeigt Tabelle 9.1.

---

Tabelle 9.1. Nebenhypothesen zur vorliegenden Untersuchung

Problemlöseleistung und Verarbeitungskapazität	Ausgehend von den Ergebnissen aus Abschnitt 6.1.4 kann man mit einem Zusammenhang von Problemlöseleistung und Verarbeitungskapazität um $r=.37$ rechnen.
Problemlöseleistung und Persönlichkeit	Bisherige Studien zum Zusammenhang von Problemlöseleistung und Persönlichkeit haben nur schwache und nicht replizierbare Zusammenhänge zu Tage gefördert (Kersting, 1999). Persönlichkeitsmaße dienen in dieser Studie daher eher der diskriminanten Validierung in dem Sinne, dass das kognitive Problemlösemaß nicht durch Persönlichkeitseigenschaften konfundiert sein sollte. Es werden keine substanziellen Zusammenhänge erwartet.
Problemlöseleistung und Interesse / Motivation	Da Motivation und Interesse eher die typische Leistung als die maximale Leistung beeinflussen ist nicht zu erwarten, dass sie über die Verarbeitungskapazität hinaus Problemlöseleistung aufklären.

---



---

## 10. AUSWERTUNG

Die Auswertung der Daten gliedert sich in vier Teile. In Abschnitt 10.1 werden die Skalen der Simulation AGRIMAN deskriptiv ausgewertet. In Abschnitt 10.2 wird auf den Zusammenhang der Problemlöseleistung mit erworbenem Systemwissen und Intelligenz eingegangen. In Abschnitt 10.3 steht der Zusammenhang mit Persönlichkeitsskalen und Motivation im Vordergrund. Schließlich wird in Abschnitt 10.4 auf den Zusammenhang von AGRIMAN mit beruflicher Problemlöseleistung eingegangen.

### 10.1 Deskriptive Auswertung

Die wichtigen durch AGRIMAN erfassten Daten sind das Endkapital in den Diagnosephasen I und II sowie die Auswertung des deklarativen Wissenstests und des Handlungswissenstests.

#### 10.1.1 Diagnosephase I

Die Problemlöseleistung wird bei AGRIMAN primär über das in der Diagnosephase I erwirtschaftete Kapital operationalisiert. Diesem Indikator kommt daher zentrale Bedeutung zu.

Das Startkapital betrug in dieser Untersuchung 25.000. Tabelle 10.1 zeigt die nach Stichprobe gruppierte Auswertung.

Tabelle 10.1. Deskriptive Auswertung des Kapitalendwerts Diagnosephase I

Stichprobe	N	Mittelwert	Median	Standardabweichung	Minimum	Maximum
1	53	21665,79	22337,11	6535,07	-598,65	35887,92
2	74	22342,80	22783,93	6403,90	-2100,81	36830,20
3	58	26161,20	25926,64	5888,76	11368,86	37382,35
Insgesamt	185	23345,96	23853,16	6542,04	-2100,81	37382,35

Der Mittelwert in der Gesamtstichprobe lag bei 23.345,96, der Median bei 23.853,16. Insgesamt waren N=71 (38%) der Teilnehmer in der Lage, ihr Startkapital zu steigern. Die Stan-

---

dardabweichung von 6.542,04 zeigt, dass ausreichend Varianz vorhanden ist, was sich auch in der breiten Spannweite der Endkapitalwerte von -2.100,81 bis 37.382,35 zeigt.

An den Daten ist darüberhinaus klar ersichtlich, dass Mittelwert und Median in der ausschließlich aus Hochschulabsolventen und Doktoranden bestehenden Stichprobe 3 deutlich höher ausfällt als in den anderen beiden Stichproben, die auch Auszubildende beinhalten.

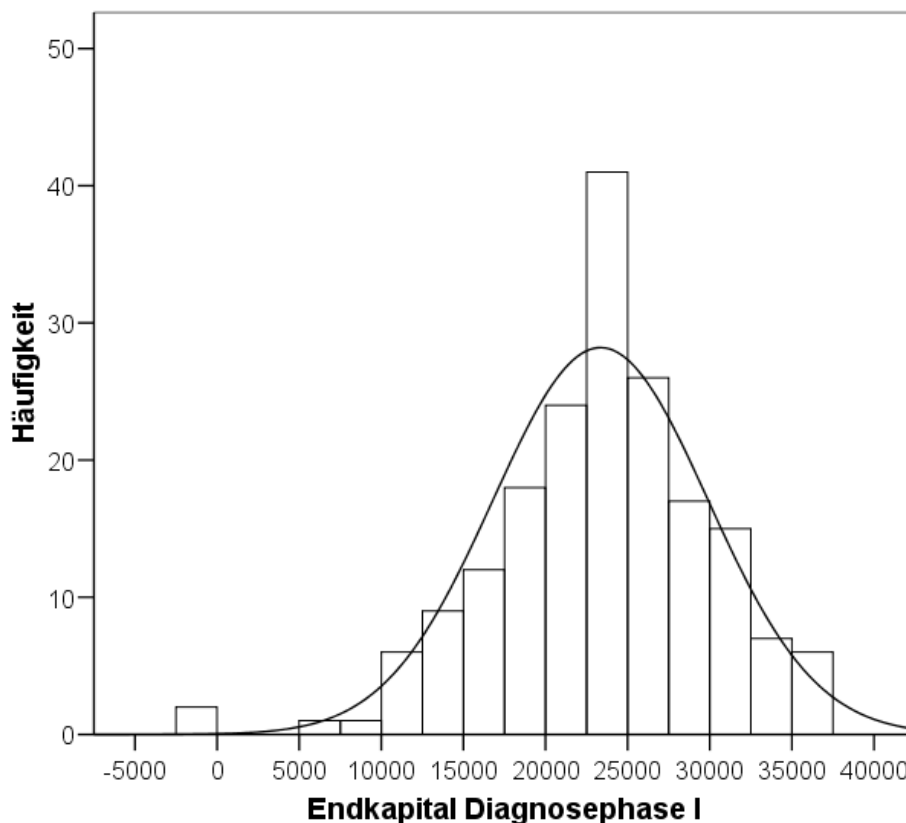


Abbildung 10.1. Histogramm des Kapitalendwerts Diagnosephase I

Das Histogramm zeigt eine annähernde Normalverteilung der Kapitalendwerte. Der Kolmogorov-Smirnov-Test auf Normalverteilung wird dementsprechend nicht signifikant (Asymptotische zweiseitige Signifikanz 0,494).

---

## 10.1.2 Diagnosephase II

In den Stichproben 2 und 3 wurde nach den Wissenstest eine zweite Diagnosephase geschaltet. Tabelle 10.2 zeigt die deskriptiven Kennwerte des Kapitalendwerts dieser Phase.

Tabelle 10.2. Deskriptive Auswertung Endkapital Diagnosephase II

Stichprobe	N	Mittelwert	Median	Standardabweichung	Minimum	Maximum
2	62	24402,94	24265,84	6386,14	2641,78	37556,59
3	46	27052,98	26814,41	5695,90	15670,38	37487,19
Insgesamt	108	25531,66	25539,26	6215,11	2641,78	37556,59

Im Vergleich zur Diagnosephase I zeigt sich ein höherer Mittelwert und Median, offensichtlich lernen die Teilnehmer noch in der Diagnosephase I dazu.

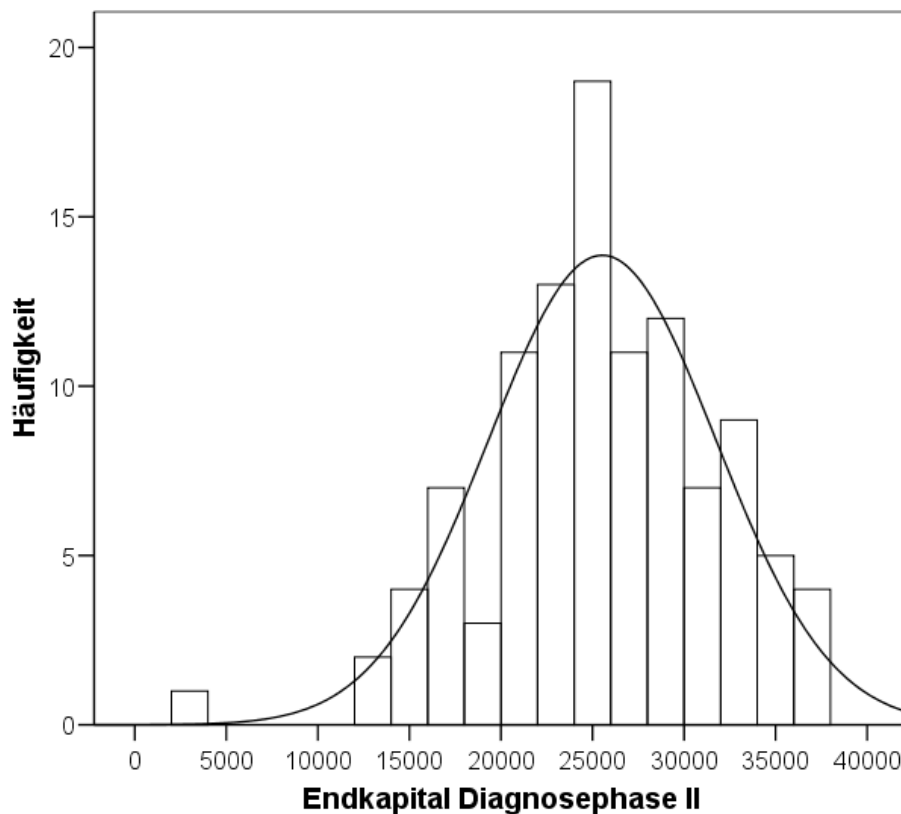


Abbildung 10.2. Histogramm des Kapitalendwerts Diagnosephase II

Auch in der Diagnosephase II zeigt das Histogramm eine annähernde Normalverteilung des Kapitalendwerts. Der Kolmogorov-Smirnov-Test auf Normalverteilung wird ebenfalls nicht signifikant (Asymptotische zweiseitige Signifikanz 0,913).

### 10.1.3 Handlungswissenstest

Im Handlungswissenstest wurden acht Mini-Szenarios vorgegeben, in denen jeweils ein Feld mit spezifischen Ausgangsbedingungen präsentiert wurde, auf die in optimaler Weise zu reagieren war. Für die richtige Wahl gibt es bis zu zwei Punkte, so dass der maximal erreichbare Punktwert 16 ist. Zusätzlich kann auch der Kapitalendwert jeder Situation ausgewertet werden.

Tabelle 10.3. Deskriptive Auswertung des Handlungswissenstests

Stichprobe	N		Mittelwert	Median	Standardabweichung	Minimum	Maximum
1	53	Punkte	6,47	6	3,37	0	16
		Kapital	42139,75	42359,06	1281,12	38935,50	45097,23
2	64	Punkte	6,94	7	3,55	1	16
		Kapital	42583,31	42559,26	1412,90	39262,07	45492,01
3	58	Punkte	8,40	8,5	3,73	2	16
		Kapital	43144,29	43037,22	1065,39	40889,79	45739,77
Insgesamt	175	Punkte	7,28	7	3,63	0	16
		Kapital	42634,90	42620,36	1321,84	38935,50	45739,77

Der Mittelwert von 7,28 zeigt eine annähernd mittlere Schwierigkeit an. Wie schon bei den Kapitalendwerten ist sowohl der Punktwert als auch der summierte Kapitalwert im Handlungswissenstest in Stichprobe 3 höher ausgeprägt als in den anderen beiden Stichproben.

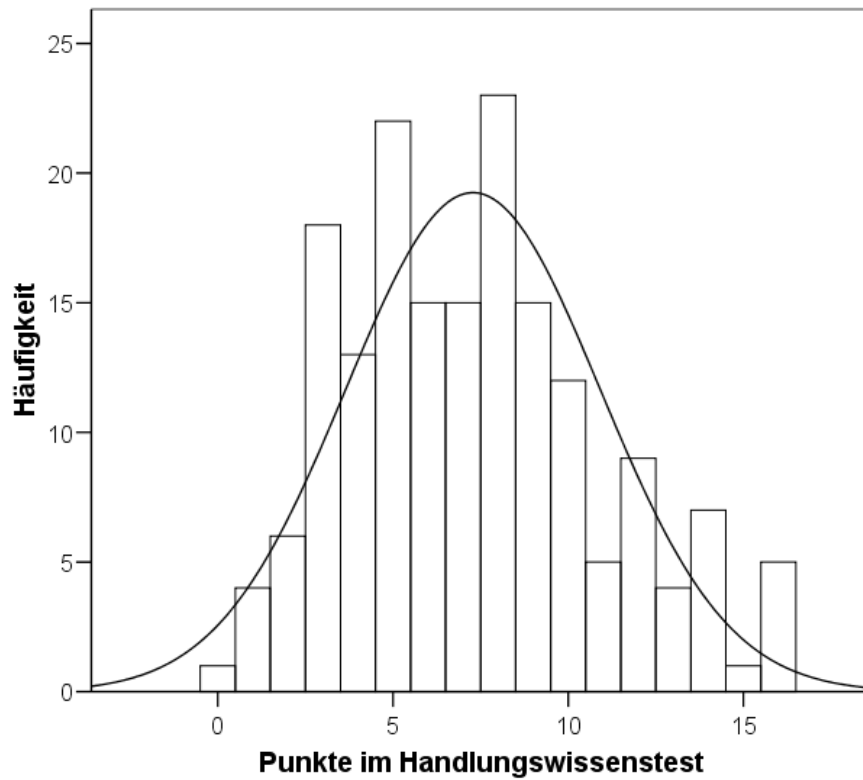


Abbildung 10.3. Histogramm des Punktwerts im Handlungswissenstest

Das Histogramm zeigt eine größere Abweichung der Werte von der Normalverteilung als bei den Kapitalendwerten. Der Kolmogorov-Smirnov-Anpassungstest wird trotzdem knapp nicht signifikant (Asymptotische zweiseitige Signifikanz 0,057).

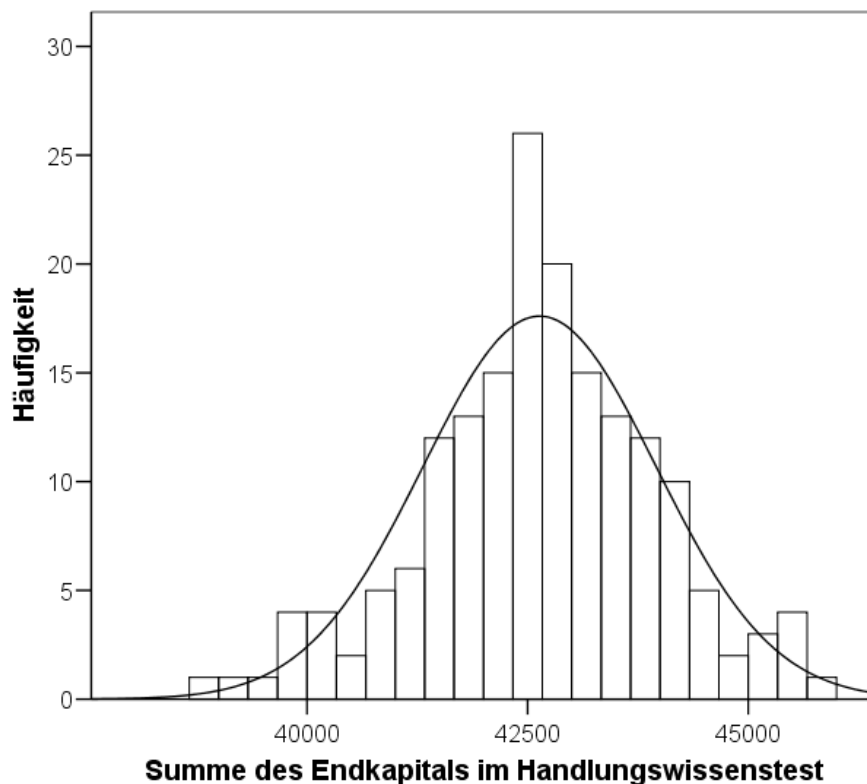


Abbildung 10.4. Histogramm des Punktwerts im Handlungswissenstest

Das summierte Endkapital ist wiederum annähernd normalverteilt, der Kolmogorov-Smirnov-Test wird nicht signifikant (Asymptotische zweiseitige Signifikanz 0,623).

### 10.1.4 Deklarativer Wissenstest

Der Wissenstest wird ausgewertet, indem die erreichten Punkte für jedes Item an der Maximalzahl erreichbarer Punkte relativiert wird. Danach wird der Mittelwert aller Items berechnet. Das theoretische Minimum liegt damit bei 0, das theoretische Maximum bei 1. Tabelle 10.4 zeigt die deskriptive Auswertung der Werte im Wissenstest.

Tabelle 10.4. Deskriptive Auswertung des deklarativen Wissenstests

Stichprobe	N	Mittelwert	Median	Standardabweichung	Minimum	Maximum
1	53	0,2793	0,2503	0,17482	0,02	0,85
2	67	0,3608	0,3499	0,17303	0,06	0,84
3	58	0,4294	0,3935	0,18159	0,14	0,84
Insgesamt	178	0,3589	0,3350	0,18516	0,02	0,85

Das Maximum von 0,85 zeigt, dass es durchaus möglich ist, beinahe alle Variablenrelationen korrekt zu entdecken. Es zeigt sich außerdem wieder, dass Stichprobe 3 im Mittel besser abschneidet als die anderen beiden Stichproben.

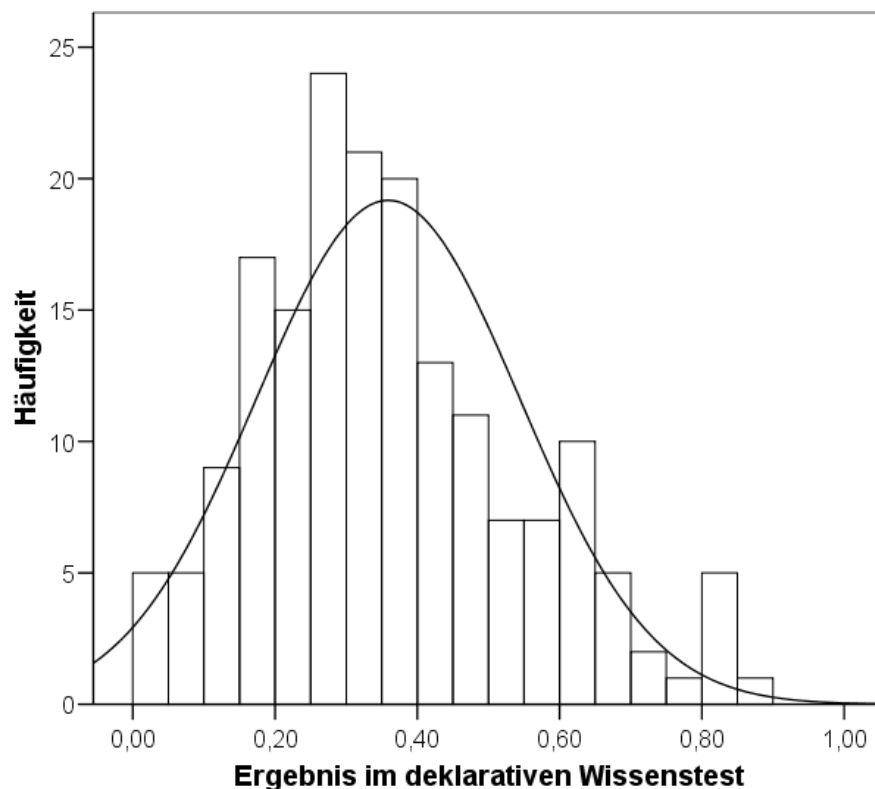


Abbildung 10.5. Histogramm der Ergebniswerte im deklarativen Wissenstest

Hier ist die Abweichung von der Normalverteilung wieder stärker ausgeprägt als bei den Kapitalendwerten, der Kolmogorov-Smirnov-Test wird aber auch hier knapp nicht signifikant (Asymptotische zweiseitige Signifikanz 0,067).

## 10.2 Zusammenhang mit Wissen und Intelligenz

Der Zusammenhang der Problemlöseleistung mit erworbenem Systemwissen und Intelligenz ist vor allem unter Forschungsgesichtspunkten interessant.

### 10.2.1 Zusammenhang mit Wissen

Tabelle 10.5 zeigt den Zusammenhang der mit AGRIMAN erfassten Ergebnisse untereinander.

Tabelle 10.5. Interkorrelation der erfassten AGRIMAN-Ergebnisse

	Endkapital (Diagnose I)	Endkapital (Diagnose II)	Handlungs- wissen (Punkte)	Handlungs- wissen (Endkapital)	Deklaratives Wissen
Endkapital (Diagnosephase I)		,887(**) n=108	,429(**) n=175	,717(**) n=175	,581(**) n=178
Endkapital (Diagnosephase II)	,887(**) n=108		,483(**) n=108	,670(**) n=108	,623(**) n=103
Handlungswissen (Punkte)	,429(**) n=175	,483(**) n=108		,675(**) n=175	,602(**) n=168
Handlungswissen (Endkapital)	,717(**) n=175	,670(**) n=108	,675(**) n=175		,628(**) N=168
Deklaratives Wissen	,581(**) n=178	,623(**) n=103	,602(**) n=168	,628(**) n=168	

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Wie bei einem reliablen Messinstrument zu erwarten ergibt sich ein mir  $r=.89$  sehr hoher Zusammenhang zwischen beiden Diagnosephasen. Ebenfalls sehr hoch fällt mir  $r=.72$  und  $r=.67$  der Zusammenhang der beiden Diagnosephasen mit dem summierten Endkapital des Handlungswissenstests aus. Das ergibt Sinn, da die Anforderungen der Miniszenerarien denen der Simulation sehr ähnlich sind.

Interessant sind vor allem die hohen Zusammenhänge der Problemlöseleistung mit dem erworbenen Systemwissen. Der Zusammenhang zwischen der Problemlöseleistung in Diagnosephase I und dem deklarativem Systemwissen beträgt  $r=.58$ . Die Problemlöseleistung in der



---

nach dem Wissenstest stattfindenden Diagnosephase II lässt sich mit dem deklarativen Systemwissen sehr gut vorhersagen, der Zusammenhang beträgt  $r=.62$ .

Geringer, aber immer noch hoch fallen die Zusammenhänge zwischen dem Punktwert im Handlungswissenstest und der Problemlöseleistung aus. Auch hier ist der Zusammenhang mit der zweiten Diagnosephase ( $r=.48$ ) höher als der Zusammenhang mit der ersten Diagnosephase ( $r=.43$ ).

Die in den Studien von Süß, Kerting und Oberauer (1991) sowie Wittmann, Süß und Oberauer (1996) gefundenen hohen Zusammenhänge zwischen Problemlöseleistung und systemspezifischem Wissen erfahren so eine Bestätigung. Die Annahme von Broadbent, dass vor allem implizites Wissen erworben wird, muss zurückgewiesen werden.

Die **Hypothese H2** kann als bestätigt gelten. Der Zusammenhang zwischen erworbenem Systemwissen und Problemlöseleistung beträgt  $r=.58$  (Diagnosephase I) bzw.  $r=.62$  (Diagnosephase II).

### 10.2.2 Zusammenhang mit Intelligenz

Tabelle 10.6 zeigt die Zusammenhänge der AGRIMAN-Ergebnisse mit der Intelligenzfacette Verarbeitungskapazität. Wie schon in der Validierungsstudie fiel auch in dieser Untersuchung die Korrelation beider Intelligenztests mit  $r=.73$  hoch aus. Deshalb wurde entschieden, beide Instrumente zu aggregieren, um so durch Testverlängerung ein reliableres Maß der Verarbeitungskapazität zu erhalten. Dazu wurden die Ergebniswerte z-standardisiert und dann der Mittelwert berechnet.

Mit der gleichen Methode wurden auch aggregierte AGRIMAN-Skalen berechnet. Die erste (*AGRIMAN I*) ist ein Aggregat aus Diagnosephase I, deklarativem Wissen und dem Punktwert im Handlungswissenstest. In die zweite aggregierte Skala (*AGRIMAN II*) wurde außerdem die zweite Diagnosephase eingerechnet. Tabelle 10.6 zeigt die Zusammenhänge.

Tabelle 10.6. Interkorrelation der erfassten AGRIMAN-Ergebnisse mit der Verarbeitungskapazität

	Verarbeitungskapazität (BIS-K und IQO)
Endkapital (Diagnose I)	,363(**) n=54
Endkapital (Diagnose II)	,454(**) n=44
Handlungswissen (Punkte)	,394(**) n=54
Deklaratives Wissen	,552(**) n=54
AGRIMAN I	,522(**) n=54
AGRIMAN II	,539(**) n=44

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

In Abschnitt 6.1.4 wurde ein über alle gefundenen Studien gemittelter Zusammenhang zwischen komplexer Problemlöseleistung und Intelligenz von  $r=.37$  ermittelt. Dies entspricht ziemlich genau dem in dieser Studie gefundenen Zusammenhang zwischen der Problemlöseleistung in Diagnosephase I und der Verarbeitungskapazität von  $r=.36$ .

Höher fällt der Zusammenhang mit der Problemlöseleistung in Diagnosephase II mit  $r=.45$  aus, und insbesondere die aggregierten Skalen zeigen einen deutlich höheren Zusammenhang ( $r=.52$  bzw.  $r=.54$ ).

Der höchste Zusammenhang findet sich zwischen der Verarbeitungskapazität und dem erworbenen deklarierbaren Systemwissen ( $r=.55$ ).

Die *Hypothese H1* kann als bestätigt gelten. Der Zusammenhang zwischen Verarbeitungskapazität und erworbenem Systemwissen beträgt  $r=.55$ .

### 10.2.3 Zusammenhang mit Wissen und Intelligenz

Um zu prüfen, inwiefern sich die Problemlöseleistung durch Intelligenz und Wissen vorher-sagen lässt, wurden jeweils multiple Regressionen mit der Problemlöseleistung in einer der

beiden Diagnosephasen als Kriterium berechnet. Die Ergebnisse zeigen die Tabellen 10.7 bis 10.12.

Tabelle 10.7. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=54), Verarbeitungskapazität vor deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,363(a)	,132	,115	5320,40	,132	,007
2	,495(b)	,245	,216	5009,49	,113	,008

a Einflußvariablen : (Konstante),Verarbeitungskapazität

b Einflußvariablen : (Konstante),Verarbeitungskapazität, deklaratives Wissen

Tabelle 10.8. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=54), Verarbeitungskapazität nach deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,481(a)	,232	,217	5005,94	,232	,000
2	,495(b)	,245	,216	5009,49	,014	,340

a Einflußvariablen : (Konstante), deklaratives Wissen

b Einflußvariablen : (Konstante), deklaratives Wissen, Verarbeitungskapazität

Tabelle 10.9. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=54): Koeffizienten

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	,339	,112		3,029	,004
	Deklaratives Wissen	,422	,107	,481	3,958	,000
2	(Konstante)	,363	,115		3,163	,003
	Deklaratives Wissen	,354	,128	,404	2,767	,008
	Verarbeitungskapazität	,203	,211	,140	,963	,340

Tabelle 10.10. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44), Verarbeitungskapazität vor deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,454(a)	,206	,187	5046,78	,206	,002
2	,591(b)	,349	,317	4625,95	,143	,005

a Einflußvariablen : (Konstante),Verarbeitungskapazität

b Einflußvariablen : (Konstante),Verarbeitungskapazität, deklaratives Wissen

Tabelle 10.11. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44), Verarbeitungskapazität nach deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,555(a)	,308	,291	4712,63	,308	,000
2	,591(b)	,349	,317	4625,95	,041	,115

a Einflußvariablen : (Konstante),deklaratives Wissen

b Einflußvariablen : (Konstante), deklaratives Wissen, Verarbeitungskapazität

Tabelle 10.12. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44): Koeffizienten

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	,158	,118		1,332	,190
	Deklaratives Wissen	,498	,115	,555	4,323	,000
2	(Konstante)	,194	,118		1,641	,108
	Deklaratives Wissen	,392	,131	,437	2,998	,005
	Verarbeitungskapazität	,376	,233	,235	1,609	,115

Zunächst fällt auf, dass der Zusammenhang zwischen deklarativem Systemwissen und Problemlöseleistung in Stichprobe 3 geringer ausfällt als in der Gesamtstichprobe. Trotzdem ergibt sich, egal welche der beiden Diagnosephasen als Kriterium genommen wird, das gleiche Bild: Das deklarative Systemwissen erklärt signifikant zusätzliche Varianz über die Verarbeitungskapazität hinaus. Die Verarbeitungskapazität hingegen erklärt nicht signifikant zusätzliche Varianz über das deklarative Wissen hinaus auf.

Es kann deshalb angenommen werden, dass die Verarbeitungskapazität vor allem über das erworbene Systemwissen vermittelt auf die Problemlöseleistung wirkt. Dass das erworbene Systemwissen über die Verarbeitungskapazität hinaus zusätzliche Varianz aufklärt zeigt jedoch, dass noch andere Faktoren beteiligt sind.

Die **Hypothese H3** kann damit als bestätigt gelten. Verarbeitungskapazität erklärt nicht signifikant zusätzliche Varianz der Steuerungsleistung über das Systemwissen hinaus.

## 10.3 Zusammenhang mit Motivation und Interesse

Vergangene Befunde zum Zusammenhang von Persönlichkeitsmerkmalen und Problemlöseleistung ergaben zumeist nur geringe und darüberhinaus inkonsistente Ergebnisse (Kersting, 1999). Der Schwerpunkt dieser Untersuchung lag deshalb nicht in diesem Bereich. Trotz allem sollen die Ergebnisse hier kurz berichtet werden. Tabelle 10.13 zeigt die Zusammenhänge mit den erhobenen Persönlichkeitsdimensionen.

Tabelle 10.13. Zusammenhang der AGRIMAN-Ergebnisse mit Persönlichkeitsskalen

	Endkapital (Diagnose I)	Endkapital (Diagnos. II)	Handlungs- wissen	Deklaratives Wissen	AGRIMAN I	AGRIMAN II
Altruismus	-,050 n=56	-,038 n=56	-,133 n=56	-,014 n=56	-,080 n=56	-,071 n=56
Ausgeglichenheit	,020 n=56	,026 n=56	,129 n=56	,140 n=56	,120 n=56	,097 n=56
Empathie	-,097 n=56	-,072 n=56	-,229 n=56	-,037 n=56	-,147 n=56	-,131 n=56
Flexibilität	,180 n=64	,234 n=62	-,113 n=64	-,094 n=57	,011 n=57	,090 n=57
Kontaktfreude	-,055 n=108	-,096 n=106	-,145 n=108	,052 n=101	-,044 n=101	-,053 n=101
Optimismus	-,034 n=64	-,033 n=62	-,072 n=64	-,025 n=57	-,063 n=57	-,048 n=57
Organisationstalent	,067 n=108	,050 n=106	-,077 n=108	,149 n=101	,064 n=101	,070 n=101
Selbstwirksamkeit	,015 n=108	,056 n=106	,143 n=108	,154 n=101	,162 n=101	,156 n=101
Teamführung	,144 n=108	,150 n=106	,023 n=108	,139 n=101	,129 n=101	,147 n=101
Verträglichkeit	,045 n=108	-,017 n=106	,047 n=108	,120 n=101	,068 n=101	,040 n=101

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Wie erwartet ergeben sich keine signifikanten Zusammenhänge. Von größerem Interesse ist der Zusammenhang der AGRIMAN-Ergebnisse mit Maßen der Motivation und Interessen. Es könnte schließlich die Vermutung aufgestellt werden, dass diese die nicht durch Intelligenz aufgeklärte Varianz erklären. Tabelle 10.14 zeigt die Ergebnisse.

Tabelle 10.14. Zusammenhang der AGRIMAN-Ergebnisse mit Motiven und Interessen

	Endkapital (Diagnose I)	Endkapital (Diagnos. II)	Handlungs- wissen (Punkte)	Deklaratives Wissen	AGRIMAN I	AGRIMAN II
Karriereorientierung	,369(**) n=108	,335(**) n=106	,151 n=108	,320(**) n=101	,328(**) n=101	,345(**) n=101
Leistungsmotivation	-,069 n=108	-,071 n=106	-,111 n=108	-,029 n=101	-,087 n=101	-,085 n=101
Lernmotivation	-,018 n=64	,006 n=62	,037 n=64	,089 n=57	,098 n=57	,094 n=57
Materielle Orientierung	,102 n=64	,157 n=62	,019 n=64	,172 n=57	,150 n=57	,179 n=57
Mathematisch- technisches Interesse	,264(*) n=64	,240 n=62	,311(*) n=64	,198 n=57	,351(**) n=57	,347(**) n=57
Wirtschaftliches Interesse	,340(**) n=108	,318(**) n=106	,304(**) n=108	,390(**) n=101	,404(**) n=101	,398(**) n=101

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Wie der Tabelle zu entnehmen ist zeigen sich signifikante Zusammenhänge mit Motiven und Interessen. Insbesondere die Motivation, beruflich voranzukommen, sowie das Interesse an Wirtschaft korreliert mittelhoch mit den AGRIMAN-Skalen. Daneben scheint auch das Interesse an Mathematik eine Rolle zu spielen.

Keine signifikanten Zusammenhänge ergeben sich mit der Lernmotivation und der Leistungsmotivation. Auch das Motiv, ein hohes Einkommen zu erzielen, scheint von geringerer Bedeutung zu sein.

Um zu prüfen, ob diese Faktoren zusätzliche Varianz über Intelligenz hinaus aufklären, wurden sie in eine multiple Regression aufgenommen. Dabei konnten nur die Karriereorientierung und das wirtschaftliche Interesse berücksichtigt werden, das mathematisch-technische Interesse und die materielle Orientierung wurden nur in Stichprobe 2 erfasst. Die Tabellen 10.15 bis 10.20 zeigen die Ergebnisse der Regressionen.

Tabelle 10.15. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=44), deklaratives Wissen vor Motivation und Interesse eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,407(a)	,166	,145	5328,77	,166	,007
2	,516(b)	,266	,229	5061,63	,100	,026
3	,566(c)	,321	,247	5000,11	,054	,240

a Einflußvariablen : (Konstante), Verarbeitungskapazität

b Einflußvariablen : (Konstante), Verarbeitungskapazität, deklaratives Wissen

c Einflußvariablen : (Konstante), Verarbeitungskapazität, deklaratives Wissen, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen

Tabelle 10.16. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=44), Motivation und Interesse vor deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,407(a)	,166	,145	5328,77	,166	,007
2	,475(b)	,225	,164	5269,05	,059	,246
3	,566(c)	,321	,247	5000,11	,095	,028

a Einflußvariablen : (Konstante), Verarbeitungskapazität

c Einflußvariablen : (Konstante), Verarbeitungskapazität, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen

c Einflußvariablen : (Konstante), Verarbeitungskapazität, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen, deklaratives Wissen,

Tabelle 10.17. Multiple Regression mit der Problemlöseleistung (Diagnosephase I) als Kriterium (n=44): Koeffizienten

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	,394	,126		3,136	,003
	Verarbeitungskapazität	,652	,231	,407	2,820	,007
2	(Konstante)	,301	,126		2,387	,022
	Verarbeitungskapazität	,357	,254	,223	1,403	,169
	Deklaratives Wissen	,316	,137	,367	2,309	,026
3	(Konstante)	-,924	,731		-1,263	,214
	Verarbeitungskapazität	,290	,257	,181	1,129	,266
	Deklaratives Wissen	,311	,136	,361	2,280	,028
	Karriereorientierung	,389	,368	,284	1,058	,297
	Interesse an Wirtschaft	-,048	,235	-,056	-,204	,840

Tabelle 10.18. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44), deklaratives Wissen vor Motivation und Interesse eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,462(a)	,214	,194	5103,11770	,214	,002
2	,592(b)	,351	,318	4695,85462	,137	,007
3	,610(c)	,372	,304	4742,79674	,021	,546

a Einflußvariablen : (Konstante), Verarbeitungskapazität

b Einflußvariablen : (Konstante), Verarbeitungskapazität, deklaratives Wissen

c Einflußvariablen : (Konstante), Verarbeitungskapazität, deklaratives Wissen, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen

Tabelle 10.19. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44), Motivation und Interesse vor deklarativem Wissen eingeführt

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,462(a)	,214	,194	5103,11770	,214	,002
2	,488(b)	,239	,178	5152,88472	,025	,546
3	,610(c)	,372	,304	4742,79674	,133	,008

a Einflußvariablen : (Konstante), Verarbeitungskapazität

c Einflußvariablen : (Konstante), Verarbeitungskapazität, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen

c Einflußvariablen : (Konstante), Verarbeitungskapazität, Karriereorientierung, Interesse an wirtschaftlichen Zusammenhängen, deklaratives Wissen,

Tabelle 10.20. Multiple Regression mit der Problemlöseleistung (Diagnosephase II) als Kriterium (n=44): Koeffizienten und Signifikanz

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	,331	,127		2,607	,013
	Verarbeitungskapazität	,769	,233	,462	3,299	,002
2	(Konstante)	,217	,123		1,763	,086
	Verarbeitungskapazität	,410	,248	,247	1,653	,106
	Deklaratives Wissen	,383	,133	,429	2,870	,007
3	(Konstante)	-,578	,730		-,792	,434
	Verarbeitungskapazität	,370	,256	,223	1,445	,157
	Deklaratives Wissen	,381	,136	,427	2,803	,008
	Karriereorientierung	,272	,367	,192	,742	,463
	Interesse an Wirtschaft	-,049	,235	-,055	-,209	,836

Die Regression mit der zweiten Diagnosephase als Prädiktor spricht eine sehr deutliche Sprache. Motivation und Interesse klären kaum zusätzliche Varianz gegenüber der Verarbeitungskapazität auf, das erworbene deklarierbare Systemwissen dagegen klärt signifikant Varianz



---

auch über Verarbeitungskapazität, Motivation und Interesse hinaus auf. Wird die erste Diagnosephase als Kriterium verwendet, ist dieser Befund nicht ganz so eindeutig, aber immer noch vorhanden. In beiden Regressionen wird der Einfluss von Motivation und Interesse nicht signifikant.

## **10.4 Zusammenhang mit Berufsleistung**

Die Untersuchung des Zusammenhangs zwischen der Leistung in komplexen Problemlöse-szenarien und Berufsleistung sowie beruflicher Problemlöseleistung steht im Vordergrund dieser Untersuchung. Berufsleistung wird dabei über die Gesamtleistung in Assessment Centern operationalisiert, berufliche Problemlöseleistung über die Leistung in komplexen, intransparenten, dynamischen Assessment Center Übungen. In den folgenden Abschnitten wird zunächst auf den Zusammenhang mit der Gesamtleistung eingegangen, um dann einzelne Übungen zu untersuchen.

### **10.4.1 Gesamtergebnis**

Die Gesamtleistung in einem anforderungsbezogen konstruierten Assessment Center wird als inhaltsvalides Sample von Berufsleistung betrachtet. Das Ergebnis errechnet sich als Mittelwert aller beobachteten Übungen, das AGRIMAN-Ergebnis und sonstige kognitive Leistungstests wurden dabei nicht eingerechnet. Tabelle 10.21 zeigt die Zusammenhänge.

Tabelle 10.21. Zusammenhang der AGRIMAN-Ergebnisse der Assessment Center Leistung

	Endkapital (Diagnose I)	Endkapital (Diagnose II)	Handlungs- wissen (Punkte)	Deklaratives Wissen	AGRIMAN I	AGRIMAN II
Stichprobe I	,323(*) n=53		,191 n=53	,300(*) n=53	,345(*) n=53	
Stichprobe II	,280(*) n=74	,277(*) n=62	,142 n=64	,135 n=67	,226 n=57	,251 n=57
Stichprobe III	,313(*) n=58	,355(*) n=46	,325(*) n=58	,428(**) n=58	,432(**) n=58	,402(**) n=46
Alle Stichproben	,346(**) n=185	,335(**) n=108	,261(**) n=175	,322(**) n=178	,378(**) n=168	,341(**) n=103

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Interessant ist zunächst einmal der Zusammenhang der über das Endkapital operationalisierten komplexen Problemlöseleistung mit der Gesamtleistung im Assessment Center. Werden alle Stichproben in die Auswertung einbezogen fällt dieser Zusammenhang mit  $r=.35$  mittel-hoch aus. Angesichts der Stichprobengröße und der Tatsache, dass in Assessment Centern kommunikative und soziale Kompetenzen betont werden, ist das ein beachtlicher Wert. Werden die Stichproben einzeln ausgewertet, ergibt sich der geringste Zusammenhang von  $r=.28$  in Stichprobe II.

Ein etwas höherer Zusammenhang von  $r=.38$  ergibt sich in der Gesamtstichprobe, wenn Endkapital, Handlungswissen und deklaratives Wissen aggregiert werden. Hier zeigen sich jedoch gegenläufige Trends in unterschiedlichen Stichproben. Während in Stichprobe III der Zusammenhang der beiden Maße des Systemwissens mit dem Gesamtergebnis höher ausfällt als der Zusammenhang des Endkapitals mit dem Gesamtergebnis, ist dies in den Stichproben I und II genau umgekehrt.

Die **Hypothese H4** kann damit als bestätigt gelten. Der Zusammenhang der Problemlöseleistung mit der über die Gesamtleistung im Assessment Center bzw. Development Center operationalisierten Berufsleistung beträgt  $r=.35$ .

## 10.4.2 Planspielbasierte Gruppendiskussion

Als primäres Maß beruflicher Problemlöseleistung dient in dieser Untersuchung das Unternehmensplanspiel COSMETICA. Wie in Abschnitt 9.3.1 beschrieben lief das Planspiel in zwei (Stichprobe III) bzw. drei (Stichprobe I) Runden ab und es wurden drei Kriterien beobachtet: Strategisches Denken und Handeln, Führungskompetenz und Entscheidungsfähigkeit sowie Teamorientierung. In der Auswertung werden zunächst nur die ersten beiden COSMETICA-Runden berücksichtigt und auf eine Auswertung der zweiten AGRIMAN-Diagnosephase verzichtet, um für alle Zusammenhänge auf den vollen Stichprobenumfang von n=111 Teilnehmern zurückgreifen zu können. Tabelle 10.22 zeigt die Ergebnisse.

Tabelle 10.22. Zusammenhang der AGRIMAN-Ergebnisse mit der COSMETICA-Leistung

	Endkapital (Diagnose I)	Handlungs- wissen (Punkte)	Deklaratives Wissen	AGRIMAN I
Cosmetica (Runde 1+2)	,335(**) n=111	,232(*) n=111	,412(**) n=111	,394(**) n=111
Cosmetica (Runde 1)	,279(**) n=111	,168 n=111	,340(**) n=111	,317(**) n=111
Cosmetica (Runde 2)	,321(**) n=111	,246(**) n=111	,398(**) n=111	,388(**) n=111
Strategisches Denken & Handeln (Runde 1+2)	,398(**) n=111	,285(**) n=111	,404(**) n=111	,437(**) n=111
Führungskompetenz & Entscheidungsfähigkeit (Runde 1+2)	,254(**) n=111	,149 n=111	,352(**) n=111	,304(**) n=111
Teamorientierung (Runde 1+2)	,208(*) n=111	,169 n=111	,307(**) n=111	,275(**) n=111
Strategisches Denken & Handeln (Runde 1)	,373(**) n=111	,240(*) n=111	,351(**) n=111	,388(**) n=111
Führungskompetenz & Entscheidungsfähigkeit (Runde 1)	,256(**) n=111	,091 n=111	,282(**) n=111	,253(**) n=111
Teamorientierung (Runde 1)	,035 n=111	,085 n=111	,195(*) n=111	,128 n=111
Strategisches Denken & Handeln (Runde 2)	,331(**) n=111	,262(**) n=111	,362(**) n=111	,384(**) n=111
Führungskompetenz & Entscheidungsfähigkeit (Runde 2)	,195(*) n=111	,176 n=111	,345(**) n=111	,289(**) n=111
Teamorientierung (Runde 2)	,293(**) n=111	,186 n=111	,300(**) n=111	,313(**) n=111

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Wie beim Gesamtergebnis ergeben sich auch mit COSMETICA mittelhohe Zusammenhänge. Die Gesamtleistung hängt zu  $r=.34$  mit dem Endkapital zusammen. Der Zusammenhang mit dem aggregierten Maß AGRIMAN I fällt mit  $r=.39$  höher aus.

Werden beide Runden getrennt ausgewertet, fällt der Zusammenhang der zweiten Runde mit der AGRIMAN-Leistung höher aus. Das ergibt Sinn, da die Leistung in der ersten Runde stärker von sozialen Aspekten des Gruppenbildungsprozesses überstrahlt wird und die kognitiven Anforderungen der Simulation in der zweiten Runde höher sind.

Bei einer getrennten Auswertung nach Kriterien fällt erwartungsgemäß der Zusammenhang mit dem Kriterium Strategisches Denken und Handeln am höchsten aus ( $r=.40$  mit dem Endkapital und  $r=.44$  mit dem aggregierten Maß). Vor dem Hintergrund dass hier eine Teamsituation mit starkem sozialen Anteil beobachtet wird ist das ein beachtlicher Wert.

Um auch die dritte COSMETICA-Runde auswerten zu können wurde Stichprobe I gesondert betrachtet. Die Ergebnisse sind Tabelle 10.23 zu entnehmen.

Tabelle 10.23. Zusammenhang der AGRIMAN-Ergebnisse mit der COSMETICA-Leistung, Stichprobe

	Endkapital (Diagnose I)	Handlungs- wissen (Punkte)	Deklaratives Wissen	AGRIMAN I
Cosmetica (Runde 1+2+3)	,329(*) n=53	,390(**) n=53	,231 n=53	,401(**) n=53
Cosmetica (Runde 1+2)	,329(*) n=53	,399(**) n=53	,173 n=53	,382(**) n=53
Cosmetica (Runde 1)	,261 n=53	,316(*) n=53	,133 n=53	,301(*) n=53
Cosmetica (Runde 2)	,332(*) n=53	,404(**) n=53	,179 n=53	,387(**) n=53
Cosmetica (Runde 3)	,254 n=53	,286(*) n=53	,291(*) n=53	,350(*) n=53
Strategisches Denken & Handeln (Runde 1+2+3)	,358(**) n=53	,387(**) n=53	,262 n=53	,426(**) n=53
Führungskompetenz & Entscheidungsfähigkeit (Runde 1+2+3)	,249 n=53	,345(*) n=53	,174 n=53	,325(*) n=53
Teamorientierung (Runde 1+2+3)	,281(*) n=53	,315(*) n=53	,182 n=53	,329(*) n=53

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Es zeigt sich, dass die dritte COSMETICA-Runde den Zusammenhang mit dem AGRIMAN-Ergebnis nicht mehr deutlich erhöht. Werden die Runden getrennt voneinander ausgewertet, zeigt sich, dass die dritte Runde weniger mit AGRIMAN zusammenhängt als die zweite. Auch das ergibt Sinn, da es in der dritten Runde viel stärker um Stress- und Krisenmanagement sowie Arbeit unter Zeitdruck geht als um strategisches Denken.

## 10.4.2 Fallstudienbasierte Gruppendiskussion

In Stichprobe II wurde statt einer planspielbasierten Gruppendiskussion eine fallstudienbasierte Gruppendiskussion durchgeführt. Sie wird in Abschnitt 9.3.2 beschrieben. Tabelle 10.24 zeigt die Zusammenhänge mit AGRIMAN-Ergebnissen.

Tabelle 10.24. Zusammenhang der AGRIMAN-Ergebnisse mit der Gruppendiskussion

	Endkapital (Diagn. I)	Endkapital (Diagn. II)	Handlungs- wissen (Punkte)	Deklarati- ves Wissen	AGRIMAN I	AGRIMAN II
Gruppendiskussion (Runde 1+2)	,192 n=74	,149 n=67	,113 n=64	,199 n=62	,225 n=57	,231 n=57
Gruppendiskussion (Runde 1)	,117 n=74	,022 n=67	-,092 n=64	,131 n=62	,045 n=57	,072 n=57
Gruppendiskussion (Runde 2)	,229(*) n=74	,239 n=67	,267(*) n=64	,228 n=62	,342(**) n=57	,330(*) n=57
Strategisches Denken & Handeln (Runde 1+2)	,227 n=74	,145 n=67	,128 n=64	,244 n=62	,260 n=57	,276(*) n=57
Ergebnisorientierung (Runde 1+2)	,275(*) n=74	,167 n=67	,215 n=64	,292(*) n=62	,301(*) n=57	,312(*) n=57
Engagement (Runde 1+2)	,099 n=74	,123 n=67	-,004 n=64	,114 n=62	,133 n=57	,142 n=57
Kommunikation (Runde 1+2)	,106 n=74	,125 n=67	,095 n=64	,100 n=62	,157 n=57	,145 n=57
Strategisches Denken & Handeln (Runde 1)	,157 n=74	,017 n=67	-,069 n=64	,157 n=62	,078 n=57	,103 n=57
Ergebnisorientierung (Runde 1)	,194 n=74	,051 n=67	,011 n=64	,247 n=62	,112 n=57	,149 n=57
Engagement (Runde 1)	,023 n=74	-,016 n=67	-,212 n=64	,022 n=62	-,053 n=57	-,025 n=57
Kommunikation (Runde 1)	,040 n=74	,028 n=67	-,061 n=64	,030 n=62	,028 n=57	,033 n=57

Strategisches Denken & Handeln (Runde 2)	,249(*) n=74	,236 n=67	,273(*) n=64	,273(*) n=62	,371(**) n=57	,376(**) n=57
Ergebnisorientierung (Runde 2)	,286(*) n=74	,231 n=67	,339(**) n=64	,260(*) n=62	,391(**) n=57	,379(**) n=57
Engagement (Runde 2)	,148 n=74	,227 n=67	,172 n=64	,176 n=62	,272(*) n=57	,264(*) n=57
Kommunikation (Runde 2)	,144 n=74	,186 n=67	,207 n=64	,139 n=62	,234 n=57	,210 n=57

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Der Zusammenhang des AGRIMAN-Endkapitals mit dem Gesamtergebnis der Gruppendiskussion fällt mit  $r=.19$  deutlich geringer aus als der Zusammenhang mit dem COSMETICA-Ergebnis. Etwas höher ist der Zusammenhang mit den aggregierten Skalen:  $r=.23$  mit AGRIMAN I und  $r=.23$  mit AGRIMAN II.

Werden beide Runden getrennt ausgewertet zeigt sich, dass auch hier die zweite Runde stärker mit AGRIMAN zusammenhängt als die erste. Dieser Effekt zeigt sich besonders bei den aggregierten Skalen: Der Zusammenhang mit AGRIMAN I beträgt  $r=.05$  in der ersten und  $r=.34$  in der zweiten Runde. Offenbar überstrahlen die sozialen Aspekte der Gruppenbildung in der ersten Runde die strategischen Elemente stärker als bei COSMETICA.

Eine getrennte Auswertung nach Kriterien zeigt erwartungsgemäß nur geringe Zusammenhänge mit Engagement und Kooperation. Überraschenderweise korreliert das Kriterium Ergebnisorientierung höher mit den AGRIMAN-Ergebnissen als das Kriterium Strategisches Denken und Handeln.

Die höchsten Zusammenhänge mit den aggregierten Skalen ergeben sich, wenn man die Kriterien der zweiten Runde gesondert betrachtet. AGRIMAN I korreliert zu  $r=.37$  mit dem Strategischen Denken und Handeln in Runde 2 und sogar zu  $r=.39$  mit der Ergebnisorientierung in Runde 2.

Nicht unerwähnt bleiben soll ein Suppressor-Effekt, der sich in den Daten zeigt. In einer multiplen Regression mit AGRIMAN I als Kriterium wurden nacheinander zuerst das Strategische Denken und Handeln in Runde 2 und danach das Engagement in Runde 1 als Prädiktoren eingeführt. Der Zusammenhang steigt dadurch von  $r=.37$  für Strategisches Denken und Handeln allein auf ein multiples  $R=.49$ . Wie Tabelle 10.21 zeigt erhält das Engagement, das selber kaum mit dem Kriterium korreliert, dabei ein negatives Beta-Gewicht.

Sollte sich dieser Suppressor-Effekt als stabil erweisen könnte er so interpretiert werden, dass das Urteil der Beobachter über die strategischen Fähigkeiten des Teilnehmers durch dessen anfängliches Engagement überstrahlt wird. Allerdings erweisen sich Suppressor-Effekte meist nicht als stabil.

Tabelle 10.25. Beta-Gewichte der multiplen Regression mit AGRIMAN I als Prädiktor (n=57)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-,957	,321		-2,981	,004
	Strategisches Denken und Handeln (Runde 2)	,382	,129	,371	2,965	,004
2	(Konstante)	-,458	,357		-1,281	,206
	Strategisches Denken und Handeln (Runde 2)	,605	,148	,589	4,092	,000
	Engagement (Runde 1)	-,411	,153	-,385	-2,678	,010

### 10.4.3 Interview

Wie in Abschnitt 9.3.3 beschrieben wurde in Stichprobe 2 ein Interview durchgeführt, in dem das Kriterium Strategisches Denken und Handeln erfasst werden sollte. Tabelle 10.26 zeigt die Zusammenhänge mit diesem Kriterium und zur diskriminanten Validierung die Zusammenhänge mit der ebenfalls erfassten sozialen Kompetenz des Teilnehmers.

Tabelle 10.26. Zusammenhang der AGRIMAN-Ergebnisse mit Interviewkriterien

	Endkapital (Diagn. I)	Endkapital (Diagn. II)	Handlungswissen (Punkte)	Deklaratives Wissen	AGRIMAN I	AGRIMAN II
Strategisches Denken und Handeln (Interview)	,241(*) n=74	,313(*) n=62	,080 n=64	,200 n=67	,182 n=57	,225 n=57
Soziale Kompetenz (Interview)	,125 n=74	,052 n=62	,022 n=64	-,019 n=67	,013 n=57	,010 n=57

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Wie erwartet hängt die erfragte Soziale Kompetenz nicht mit AGRIMAN zusammen. Auch beim erfragten Strategischen Denken und Handeln ergeben sich allerdings nur für das Endkapital in beiden Diagnosephasen signifikante Zusammenhänge.

#### 10.4.4 Rollenspiele

Rollenspiele sind klassische Assessment Center Übungen. Da sie primär soziale Kompetenz und kommunikative Fähigkeiten messen dienen sie an dieser Stelle der diskriminanten Validierung. Tabelle 10.27 zeigt die Zusammenhänge.

Tabelle 10.27. Zusammenhang der AGRIMAN-Ergebnisse mit Rollenspiel-Ergebnissen

	Endkapital (Diagn. I)	Endkapital (Diagn. II)	Handlungs- wissen (Punkte)	Deklarati- ves Wissen	AGRIMAN I	AGRIMAN II
Mitarbeitergespräch (Stichprobe 1)	,147 n=53		,088 n=53	,017 n=53	,108 n=53	
Verhandlungsübung (Stichprobe 1)	,100 n=53		,183 n=53	,005 n=53	,121 n=53	
Mitarbeitergespräch (Stichprobe 2)	,135 n=64	,050 n=62	-,048 n=64	-,002 n=57	,023 n=57	,034 n=57
Konfliktgespräch (Stichprobe 2)	,042 n=47	,060 n=45	,065 n=47	-,147 n=45	-,027 n=45	-,001 n=45
Konfliktgespräch (Stichprobe 3)	-,010 n=58	,121 n=46	,191 n=58	,275(*) n=58	,191 n=58	,191 n=46

Wie erwartet zeigen sich praktisch keine Zusammenhänge mit den Rollenspielergebnissen, lediglich das Konfliktgespräch in Stichprobe 3 zeigt einen signifikanten Zusammenhang mit dem deklarativen Systemwissen, was bei 26 berechneten Korrelationen aber nicht verwunderlich ist. Auch eine Auswertung der erfassten Einzelkriterien zeigt keine signifikanten Zusammenhänge.



---

## 10.4.5 Im Vergleich mit dem Intelligenztest

AGRIMAN wurde mit dem Ziel entwickelt, eine sozial valide und damit in der angewandten Eignungsdiagnostik einsetzbare Alternative zu Intelligenztests zu schaffen. In diesem Abschnitt soll die Validität deshalb der des eingesetzten Intelligenzmaßes gegenübergestellt werden. Dabei wird auf n=54 Personen der Stichprobe 3 zurückgegriffen, die sowohl AGRIMAN als auch den Intelligenztest bearbeitet haben.

Tabelle 10.28. Zusammenhang von AC-Maßen mit AGRIMAN und Intelligenz

	AGRIMAN I	Intelligenz
AC-Gesamtergebnis	,410(**) n=54	,528(**) n=54
Cosmetica (Runde 1+2)	,320(*) n=54	,470(**) n=54
Strategisches Denken & Handeln (Runde 1+2)	,411(**) n=54	,487(**) n=54
Führungskompetenz & Entscheidungsfähigkeit (Runde 1+2)	,246 n=54	,450(**) n=54
Teamorientierung (Runde 1+2)	,135 n=54	,212 n=54

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Im Vergleich übertrifft die Validität des Intelligenztests die in dieser Stichprobe ohnehin schon hohe Validität von AGRIMAN noch einmal deutlich. Die in der Vergangenheit immer wieder bestätigte hohe Validität von Intelligenztests findet damit in den vorliegenden Daten einmal mehr eine eindrucksvolle Bestätigung. Im Vergleich fällt vor allem auf, dass der Intelligenztest nicht nur das Kriterium Strategisches Denken und Handeln sehr gut vorhersagt, sondern auch das Kriterium Führungskompetenz und Entscheidungsfähigkeit.

Aufgrund der vergleichsweise geringen Stichprobengröße müssen diese Ergebnisse allerdings als vorläufig betrachtet werden. Ein Blick auf die Scatterplots soll veranschaulichen, warum.

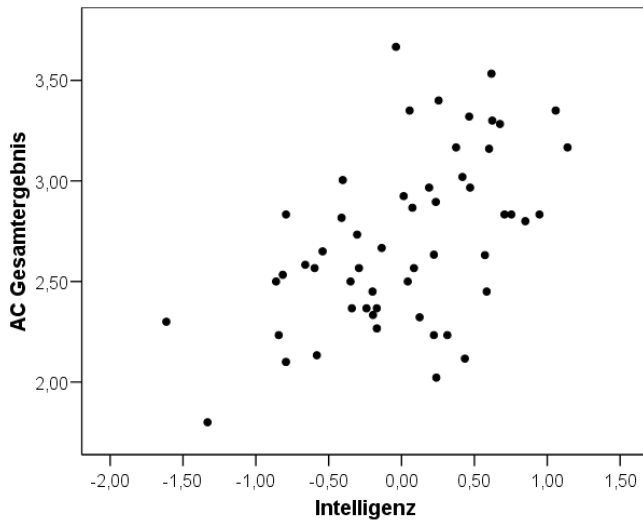


Abbildung 10.6. Streudiagramm: Zusammenhang von Intelligenz und AC Gesamtergebnis

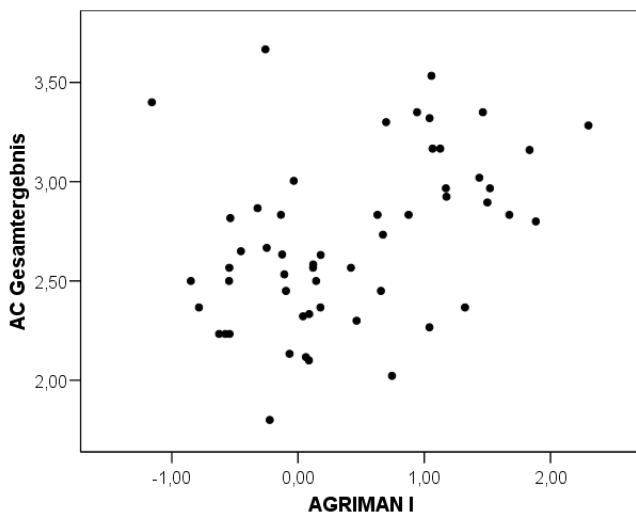


Abbildung 10.7. Streudiagramm: Zusammenhang von AGRIMAN I und AC Gesamtergebnis

Die Inspektion von Abbildung 10.7 zeigt im linken oberen Bereich des Diagramms zwei Datensätze, die dem Zusammenhang komplett widersprechen. Diese Personen schnitten im Assessment Center hervorragend ab, erreichten aber nur unterdurchschnittliche Werte in AGRIMAN. Woran das liegt kann hier nicht beantwortet werden. Der Zusammenhang mit dem AC Gesamtergebnis würde ohne diese beiden Fälle  $r=.58$  betragen statt  $r=.42$ . Dies zeigt den großen Einfluss, den Einzelfälle bei dieser Stichprobengröße noch haben.

Trotzdem soll auch die inkrementelle Validität beider Instrumente übereinander untersucht werden. Die Tabellen 10.29 und 10.30 zeigen das Ergebnis multipler Regressionen mit dem AC Gesamtergebnis als Kriterium.

Tabelle 10.29. Multiple Regression mit dem AC Gesamtergebnis als Kriterium (n=54)

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,528(a)	,278	,265	,36305	,278	,000
2	,551(b)	,303	,276	,36021	,025	,183

a Einflußvariablen : (Konstante), Verarbeitungskapazität

b Einflußvariablen : (Konstante), Verarbeitungskapazität, AGRIMAN I

Tabelle 10.30. Multiple Regression mit dem AC Gesamtergebnis als Kriterium (n=54)

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderung in R-Quadrat	Änderung in Signifikanz von F
1	,410(a)	,168	,152	,38985	,168	,002
2	,551(b)	,303	,276	,36021	,135	,003

a Einflußvariablen : (Konstante), AGRIMAN I

b Einflußvariablen : (Konstante), AGRIMAN I, Verarbeitungskapazität

Tabelle 10.31. Multiple Regression mit dem AC Gesamtergebnis als Kriterium (n=54): Koeffizienten und Signifikanz

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	2,699	,049		54,627	,000
	Verarbeitungskapazität	,374	,084	,528	4,480	,000
2	(Konstante)	2,662	,056		47,238	,000
	Verarbeitungskapazität	,306	,097	,431	3,148	,003
	AGRIMAN I	,096	,071	,185	1,350	,183

Es zeigt sich, dass AGRIMAN keine signifikante inkrementelle Validität über die Verarbeitungskapazität hinaus aufweist, die Verarbeitungskapazität umgekehrt aber signifikant inkrementelle Validität über AGRIMAN hinaus aufweist.



---

# 11. DISKUSSION

Im Rahmen dieser Arbeit wurde ein einfaches Arbeitsmodell postuliert, aus dem zentrale Hypothesen abgeleitet wurden. Dieses Modell geht davon aus, dass Probanden beim komplexen Problemlösen ihre Verarbeitungskapazität einsetzen, um deklaratives Systemwissen zu erwerben und dieses Wissen dann verwenden, um das System optimal zu steuern. Demnach sollte Verarbeitungskapazität hoch mit dem erworbenen deklarativen Systemwissen zusammenhängen (Hypothese 1) und das erworbene Systemwissen wiederum die Problemlöseleistung vorhersagen (Hypothese 2). Weiterhin wurde davon ausgegangen, dass die Verarbeitungskapazität über das deklarative Systemwissen hinaus keinen Einfluss auf die Problemlöseleistung hat (Hypothese 3).

Die vierte und letzte Hypothese ist der Kern der Validierung des komplexen Problemlöse Szenarios. Sie geht davon aus, dass die Problemlöseleistung mit der über kontextvaliden Assessment Center und Development Center operationalisierten Berufsleistung zusammenhängt. Entlang dieses Arbeitsmodells sollen zunächst die Ergebnisse gesichtet werden, um dann deren Relevanz für die Forschung zum komplexen Problemlösen und die angewandte Eignungsdiagnostik zu diskutieren.

## 11.1 Sichtung der Ergebnisse

Die deskriptive Auswertung der Daten zeigt zum einen, dass alle Ergebniswerte von AGRIMAN normalverteilt sind, Kolmogorov-Smirnov-Anpassungstests werden trotz der Stichprobengröße nicht signifikant. Zum anderen zeigt sich eine angemessene Schwierigkeit der Simulation. 38% der Teilnehmer waren in der Lage, das Startkapital zu steigern. Das ist sehr wichtig: Ist die Simulation zu schwer eingestellt, frustriert sie, ist ein Kapitalanstieg zu leicht zu erreichen, motiviert sie nicht zu einer konstanten Optimierung der Steuerung. Durch die angemessene Schwierigkeit ist eine wichtige Voraussetzung für Zusammenhänge mit Außenkriterien wie Verarbeitungskapazität und Berufsleistung gegeben.

---

## Hypothese 1

Im Rahmen des Arbeitsmodells wird ein hoher Zusammenhang zwischen Verarbeitungskapazität und deklarativem Systemwissen erwartet (Hypothese 1). Diese Erwartung findet in dem gefundenen Zusammenhang von  $r=.55$  eine Bestätigung. Die Höhe dieses Zusammenhanges ist umso bemerkenswerter, da es sich um einen nicht minderungskorrigierten Zusammenhang handelt. Eine doppelte Minderungskorrektur mit einer angenommenen Reliabilität des Verarbeitungskapazitätsmaßes von  $r_{tt}=.9$  und des Wissensmaßes von  $r_{tt}=.8$  ergäbe einen Zusammenhang von  $r=.65$ , was einer gemeinsamen Varianz von 42,4% entspricht. Eine solche Aufwertung ist durchaus gerechtfertigt, da die Ratingskalen des Wissenstests das deklarative Systemwissen keineswegs in seiner Gänze erfassen: Es werden nicht die angegebenen Verhältnisse verschiedener Antworten ausgewertet, sondern lediglich die Rangfolge.

An dieser Stelle muss angemerkt werden, dass der besondere Aufbau von AGRIMAN den Aufbau von Systemwissen begünstigt und somit auch den hohen Zusammenhang von Verarbeitungskapazität und Systemwissen mit bedingen könnte. In der Experimentalphase kann der Proband das System ohne Rücksicht auf Verluste analytisch sezieren. Am erfolgversprechendsten sind experimentelle Eingriffe, bei denen nur eine Variable systematisch variiert wird, um aus den unterschiedlichen Resultaten deren Einfluss auf andere Variablen zu extrahieren. Unterstützt wird eine solche Herangehensweise außerdem dadurch, dass der Proband durch das Zurücksetzen der Felder stets gleiche Ausgangsbedingungen für seine Experimente vorfindet. Dieses analytische Vorgehen erfordert natürlich in besonderem Maße induktive und deduktive Denkprozesse, die der Verarbeitungskapazität zugeordnet sind. AGRIMAN ist Intelligenztests damit sicher näher als hochkomplexe Simulationen ohne Experimentierphase, in denen Systemexploration und Systemsteuerung von Anfang an gegeneinander abzuwägen sind. Es könnte argumentiert werden, dass nur in einem solchen Fall andere Konstrukte wie die „operative Intelligenz“ wichtiger werden als die Verarbeitungskapazität. Doch das ist unwahrscheinlich. Auch wenn die systematische Variation von Variablen fehlt, wird aus Eingriffen und daraus resultierenden Systemzuständen Wissen extrahiert. Der Unterschied ist lediglich, dass dies unsystematisch geschieht und vermutlich insgesamt weniger Wissen erworben wird.

Der Zusammenhang der Verarbeitungskapazität mit der Problemlöseleistung fällt geringer aus als der Zusammenhang mit dem erworbenen Systemwissen. Der gefundene Zusammenhang

---

von  $r=.36$  mit der ersten Diagnosephase entspricht fast genau dem in Abschnitt 6.1.4 über alle recherchierten Studien gemittelten Zusammenhang von Intelligenz und komplexer Problemlöseleistung von  $r=.37$ .

## **Hypothese 2**

Das Arbeitsmodell nimmt an, dass die Problemlöseleistung entscheidend von dem erworbenen Systemwissen beeinflusst wird (Hypothese 2). Auch diese Annahme kann aus den Daten heraus bestätigt werden: Der Zusammenhang beträgt  $r=.58$  für die erste Diagnosephase und  $r=.62$  für die zweite Diagnosephase. Auch diese Werte sind schon ohne statistische Korrektur sehr hoch. Eine doppelte Minderungskorrektur mit einer angenommenen Reliabilität des Wissensmaßes und des Problemlösemaßes von  $r_{tt}=.8$  ergäbe einen Zusammenhang von  $r=.73$  für die erste Diagnosephase und  $r=.78$  für die zweite Diagnosephase. Das entspricht einer gemeinsamen Varianz von 52,7% bzw. 60,7%. Auch diese Aufwertung ist nicht ungerechtfertigt. Auf die Unvollkommenheit des Wissenstests wurde schon eingegangen. Im Rahmen dieser Hypothese muss noch ein weiterer Faktor einbezogen werden. Nicht wenige Probanden konzentrierten sich im Rahmen der Diagnosephase schnell auf die Optimierung einer einzelnen, erfolgversprechenden Pflanze. Das in dieser Untersuchung eingesetzte Systemmodell ermöglicht es, auch mit Monokulturen ein hohes Endkapital zu erzielen, so dass dieses Vorgehen durchaus intelligent und erfolgversprechend ist. Der Wissenstest reflektiert diese Spezialisierung allerdings nicht – das mangelnde Wissen über die zwei anderen Pflanzen wird mit einem geringeren Score bestraft.

Die Annahme einer nahezu monokausalen Wirkung von erworbenem Systemwissen auf die komplexe Problemlöseleistung erscheint vor dem Hintergrund dieser hohen Zusammenhänge nicht ungerechtfertigt. Wichtig für die Einordnung des Befundes ist, dass es sich bei dem gemessenen Systemwissen definitiv nicht um Vorwissen handelt, sondern um während der Steuerung erworbenes Wissen. Vorwissen im Bereich Landwirtschaft wurde von allen Probanden verneint.

Auch hier muss angemerkt werden, dass dem Aufbau von Wissen bei AGRIMAN durch die Experimentalphase möglicherweise eine größere Bedeutung zukommt als in anderen komplexen Simulationen. Im Fall der zweiten Hypothese kann das entscheidend sein. In einer Simulation ohne Experimentalphase haben die anfänglichen Annahmen eines Probanden ein hohes

---

Gewicht. Sind sie fehlerhaft, wird das System nicht nur in den ersten Zeittakten suboptimal gesteuert, sondern auch für die folgenden Takte in einen unvorteilhaften Ausgangszustand versetzt. Dieser unerwünschte Effekt verstärkt sich, wenn die Reversibilität des Systems gering ist. Sind die anfänglichen Annahmen dagegen korrekt, kann das System ohne große Abweichungen von der ursprünglichen Strategie erfolgreich gesteuert werden. Der Einfluss auf den Zusammenhang von Systemwissen und Problemlöseleistung kann paradox sein: Sind keine oder nur wenig Anpassungen der Steuerungseingriffe nötig, werden auch weniger resultierende Systemzustände beobachtet. Es ist dann kein so umfassender Wissenserwerb möglich wie bei wiederholten Fehlsteuerungen und Anpassungen. Ohne Experimentalphase ist daher im schlimmsten Fall sogar eine negative Korrelation von Systemwissen und Steuerungsleistung denkbar. Daraus darf nicht geschlossen werden, dass das Arbeitsmodell nicht stimmt. Die massive Überbewertung der anfänglichen Annahmen stellt dann vielmehr die eingesetzte Simulation in Frage.

### **Hypothese 3**

Im Arbeitsmodell wird davon ausgegangen, dass Verarbeitungskapazität nur vermittelt über das erworbene Systemwissen auf die Steuerungsleistung wirkt, also keinen unabhängigen Einfluss auf sie hat (Hypothese 3). Diese Annahme wurde durch multiple Regressionen geprüft. Egal, welche der beiden Diagnosephasen dabei als Kriterium verwendet wird, Verarbeitungskapazität klärt keine zusätzliche signifikante Varianz über Wissen hinaus auf. Das bedeutet nicht, dass die Probanden ihre Verarbeitungskapazität in der Diagnosephase nicht benötigen. In diesem Fall müsste die reine Vermittlung von Systemwissen Probanden in die Lage versetzen, das System erfolgreich zu steuern, was jedoch in der Untersuchung von Süß (1996) nicht der Fall ist. Der Befund besagt aber, dass für die Anwendung von Wissen offenbar keine andere Facette der Verarbeitungskapazität und auch keine andere Fähigkeit benötigt wird als für den Erwerb.

### **Bewährung des Arbeitsmodells**

Die ersten drei Hypothesen konnten bestätigt werden. Auf der Basis dieser Ergebnisse kann davon ausgegangen werden, dass die Probanden die Simulation in der Experimentalphase systematisch explorieren, um Wissen zu sammeln, und dass dies auch gelingt. In der Diagno-



---

sephase wird das erworbene Wissen dann gezielt eingesetzt, um die Simulation optimal zu steuern. Für den Erfolg bei der Steuerung ist der Umfang des Wissenserwerbs entscheidend. Die Bestätigung der Hypothesen zeigt darüber hinaus, dass bei der Systemsteuerung sehr wohl explizites Wissen erworben wird. Die These von Broadbent, dass Probanden vor allem implizites, nicht verbalisierbares Wissen erwerben, konnte damit erneut widerlegt werden. Aus diesem Grund ist das Ergebnis des in dieser Untersuchung eingesetzten Handlungswissenstests weniger interessant, da er primär für den Fall entwickelt wurde, dass nicht deklarierbares Systemwissen erfasst werden muss. Die Konzeption von Handlungswissenstests über die Vorgabe von Situationen, die so auch in der Simulation vorkommen, ist ohnehin problematisch. Solche Tests kopieren letztlich 1:1 die Anforderungen der Simulation. Insbesondere wenn der Planungshorizont wie bei AGRIMAN kurz ist, also vor allem der aktuelle Zeittakt optimal gesteuert werden muss, ohne weit in die Zukunft zu denken, sind die erfassten Kompetenzen kaum zu unterscheiden. Es muss dann eher von einem Retest gesprochen werden, als von der Erfassung impliziten Handlungswissens. Das zeigt sich auch in den Daten. Werden die Situationen anhand des Endkapitals ausgewertet, korrelieren Sie zu .717 mit der ersten Diagnosephase und .670 mit der zweiten.

Was kann aus der Höhe der Zusammenhänge geschlossen werden? Der Zusammenhang zwischen deklarativem Wissen und Steuerungsleistung ist so hoch, dass von einer monokausalen Beziehung ausgegangen werden kann. Das bedeutet, der Steuerungserfolg hängt einzig vom erworbenen Systemwissen ab, und wird damit auch indirekt nur von der Fähigkeit zum Wissenserwerb beeinflusst.

Doch was ist diese Fähigkeit zum Wissenserwerb? In dieser Arbeit wurde davon ausgegangen, dass es sich dabei vor allem um die Verarbeitungskapazität bzw. die dahinterstehende Arbeitsgedächtniskapazität handelt. Verarbeitungskapazität sagt das erworbene Systemwissen jedoch nicht perfekt vorher, auch nach doppelter Minderungskorrektur überlappen beide Konstrukte nur zu 42,4%. Dazu kommt, dass das Systemwissen über die Verarbeitungskapazität hinaus Varianz der komplexen Problemlöseleistung aufklärt. 11,3% der Leistung in Diagnosephase 1 und sogar 14,3% der Leistung in Diagnosephase 2 werden durch das Systemwissen, nicht aber die Verarbeitungskapazität aufgeklärt.

---

Es kann deshalb nicht ausgeschlossen werden, dass noch andere Faktoren am Wissenserwerb und der komplexen Problemlöseleistung beteiligt sind. In dieser Arbeit wurde untersucht, ob es sich dabei um Persönlichkeitsmerkmale, Motivation oder Interesse handelt, also nichtkognitive Merkmale. Das konnte in Abschnitt 10.3 nicht bestätigt werden. Mit Persönlichkeitsmerkmalen konnte überhaupt kein Zusammenhang gefunden werden. Motivation und Interesse korrelieren mittelhoch mit der Problemlöseleistung. Sie erklären allerdings nicht signifikant zusätzliche Varianz über die Verarbeitungskapazität hinaus. Der mittelhohe Zusammenhang resultiert daher vermutlich indirekt daraus, dass Probanden mit hoher Verarbeitungskapazität auch höhere Motivation aufweisen und mehr Interesse zeigen. Diese Ergebnisse machen durchaus Sinn. Persönlichkeitsmerkmale, Motivation und Interesse modulieren Verhalten eher langfristig. In den zwei bis drei Stunden, in denen AGRIMAN läuft, kann dagegen davon ausgegangen werden, dass die Probanden unabhängig von Persönlichkeit, Motivation und Interessen ihre maximale Performance bringen.

Ist also doch ein weiteres kognitives Konstrukt im Spiel, gar die „operative Intelligenz“? Das muss bezweifelt werden. Wahrscheinlicher ist, dass es sich gar nicht um ein generelles Konstrukt handelt, sondern um szenariospezifische Faktoren. Kersting (1999) zitiert Studien, in denen verschiedene Simulationen gemeinsam eingesetzt wurden. Der höchste sich ergebende Zusammenhang zweier Simulationen beträgt .51. Er schließt daraus, dass Problemlöseszenarien keine transsituationalen Anforderungen an eine Problemlösefähigkeit stellen, sondern neben Intelligenz und Wissen nur hochgradig situations- und aufgabenspezifische Fähigkeiten erfordern, die keine diagnostische Relevanz haben. Auch die Befunde von Süß (1996), nach denen sich die gesamte gemeinsame Varianz verschiedener Szenarien durch Intelligenz und Wissen aufklären lässt, passt zu dieser Annahme.

#### **Hypothese 4 - Die Validierung**

Das Hauptziel dieser Untersuchung war nicht die Untersuchung von Determinanten der komplexen Problemlöseleistung, sondern die saubere Validierung des Diagnoseinstrumentes AGRIMAN in einer großen Stichprobe. Dabei wird ein Zusammenhang zwischen Problemlöseleistung und Berufsleistung erwartet (Hypothese 4).

Berufsleistung wird in dieser Untersuchung über das Gesamtergebnis in kontextvaliden Assessment Centern und Development Centern operationalisiert. Daneben kann ein genauerer

---

Blick auf berufliche Problemlöseleistung geworfen werden, indem gezielt der Zusammenhang mit der Leistung in planspielbasierten und fallstudienbasierten Gruppendiskussionen untersucht wird. Zur diskriminanten Validierung dient die Leistung in Rollenspielen, die soziale Kompetenzen erfordern und damit von AGRIMAN unabhängig sein sollten.

Es zeigen sich ermutigende Ergebnisse. Der Zusammenhang zwischen dem AGRIMAN-Endkapital und dem Ergebnis im Assessment Center bzw. Development Center beträgt in der Gesamtstichprobe von  $n=185$  Teilnehmern  $r=.35$ . Das aggregierte Maß aus Endkapital, Handlungswissenstest und deklarativem Wissenstest, das für  $n=168$  Teilnehmer berechnet werden konnte, hängt sogar zu  $r=.38$  mit dem Gesamtergebnis zusammen. Dies ist ein beachtlicher Zusammenhang wenn man bedenkt, dass Assessment Center trotz der in dieser Untersuchung verwendeten planspiel- und fallstudienbasierten Gruppendiskussionen noch immer stark soziale Kompetenz und kommunikative Fähigkeiten betonen. Die vierte Hypothese ist damit ebenfalls bestätigt.

Die Untersuchung der stärker kognitiv geprägten Einzelübungen, die berufliche Problemlöseleistung operationalisieren, zeigt ebenfalls ermutigende Zusammenhänge. Der Zusammenhang der Gesamtleistung in der planspielbasierten Gruppendiskussion mit dem AGRIMAN-Endkapital beträgt in der großen Substichprobe von  $n=111$  Teilnehmern  $r=.34$ . Der Zusammenhang mit der aggregierten Skala AGRIMAN I beträgt  $r=.39$ . Betrachtet man nur das Kriterium Strategisches Denken und Handeln, steigt der Zusammenhang sogar auf  $r=.40$  für das Endkapital und  $r=.44$  für die aggregierte Skala.

Insgesamt geringer fallen die Zusammenhänge mit der Leistung in der fallstudienbasierten Gruppendiskussion aus. Die Korrelation mit der Gesamtleistung beträgt nur  $r=.19$  für das Endkapital und  $r=.23$  für die aggregierte Skala. Der geringe Zusammenhang ist höchstwahrscheinlich darauf zurückzuführen, dass die soziale und kommunikative Kompetenz das strategische Denken und Handeln stärker überstrahlen als das in der planspielbasierten Gruppendiskussion der Fall ist. Das ergibt durchaus Sinn. In der planspielbasierten Gruppendiskussion haben die Gruppe und auch die Beobachter mit der Simulation ein klares Kriterium zur nachträglichen Bewertung von Diskussionsbeiträgen und daraus folgenden Aktionen. Gruppenmitglieder, die wiederholt schlechte Vorschläge machen, treten dadurch fast automatisch in den Hintergrund, auch wenn sie ihre Vorschläge noch so eloquent vorgetragen haben. Eine solche objektive Bewertung geschieht in der Gruppendiskussion nicht: Hier können inhaltli-

---

che Schwächen von Diskussionsbeiträgen beliebig durch Rhetorik überdeckt werden, ohne den Praxistest befürchten zu müssen.

Mit der zweiten Runde der Diskussion ergeben sich immerhin substantziellere Zusammenhänge von  $r=.23$  für das Endkapital und  $r=.34$  für die aggregierte Skala. Auch das ist nicht überraschend, da es im Rahmen der Gruppenfindungsprozesse der ersten Runde sehr stark auf soziale Kompetenzen ankommt, wohingegen die zweite Runde stärker von strategischen Anforderungen geprägt ist. Mit dem Kriterium Strategisches Denken und Handeln in der zweiten Runde liegt der Zusammenhang dann auch noch etwas höher:  $r=.25$  für das Endkapital und  $r=.37$  für das aggregierte Maß.

Dass reine soziale und kommunikative Kompetenz durch AGRIMAN nicht gemessen wird, zeigt der nicht vorhandene Zusammenhang der AGRIMAN-Ergebnisse mit der Leistung in den unterschiedlichen eingesetzten Rollenspielen. Die durchschnittliche Korrelation des Endkapitals in der ersten Diagnosephase mit den fünf verschiedenen Rollenspielen beträgt  $r=.08$ , keine der fünf Korrelationen wird signifikant. Auch mit der im Interview gemessenen sozialen Kompetenz ergeben sich keine signifikanten Zusammenhänge, die Korrelation beträgt  $r=.13$ . Das liegt nicht an der Erhebungsform. Im gleichen Interview wurde auch das Kriterium Strategisches Denken und Handeln erfasst. Mit diesem Kriterium korreliert das Ergebnis der ersten Diagnosephase signifikant zu  $r=.24$ . Dass AGRIMAN nicht mit sozialer Kompetenz zusammenhängt ist im Sinne einer diskriminanten Validierung positiv zu werten. Es bedeutet, dass keine sozialen Faktoren die Messung der Problemlöseleistung in komplexen Situationen konfundieren.

Da es keine Zusammenhänge der Problemlöseleistung mit sozial geprägten Übungen gibt erscheint es zunächst schwer erklärbar, warum die stark kognitiv geprägten planspielbasierten bzw. fallstudienbasierten Gruppendiskussionen kaum stärker mit der Problemlöseleistung zusammenhängen als das Gesamtergebnis des Verfahrens. Die plausibelste Erklärung ist vermutlich messtheoretischer Natur: Das Gesamtergebnis eines Assessment Centers oder Development Centers setzt sich aus mehreren weitgehend unabhängigen Übungen zusammen und dürfte dadurch deutlich reliabler sein als das Ergebnis einer einzelnen Übung. Die geringere Reliabilität der einzelnen Übung begrenzt an dieser Stelle möglicherweise den Zusammen-

---

hang mit AGRIMAN. Es wäre interessant zu untersuchen wie hoch der Zusammenhang ausfällt, wenn eine Reihe von kognitiv geprägten Übungen summiert und mit der komplexen Problemlöseleistung korreliert wird.

## 11.2 Bewertung der Ergebnisse

Die Sichtung der Ergebnisse bestätigt alle vier Hypothesen dieser Untersuchung und damit das postulierte Arbeitsmodell. Die komplexe Problemlöseleistung ist abhängig vom Umfang des erworbenen Systemwissens, das wiederum direkt von der Verarbeitungskapazität abhängig ist. Die komplexe Problemlöseleistung korreliert mit der über inhaltsvalide Assessment Center und Development Center operationalisierten Berufsleistung. In diesem Abschnitt soll die Bedeutung dieser Befunde zum einen für die Forschung zum komplexen Problemlösen und zum anderen für die angewandte Eignungsdiagnostik diskutiert werden.

### Komplexes Problemlösen

Mit der Forschung zu komplexen Problemlöseszenarien verbanden sich ursprünglich große Hoffnungen und Ansprüche. Es wurde postuliert, dass die Simulationen als Modelle der Realität eine hohe ökologische Validität aufweisen. Komplexe Problemlösesimulationen wurden so in die Nähe inhaltsvalider Arbeitsproben gerückt. Daraus folgend wurde eine gegenüber anderen Instrumenten überlegene Validität behauptet, insbesondere der klassischen Intelligenzdiagnostik wurde regelrecht der Kampf angesagt. Dem über Intelligenztests gemessenen Intelligenzkonstrukt wurde die nur über komplexe Problemlöseszenarien messbare „Operative Intelligenz“ gegenübergestellt.

Diese hohen Ansprüche müssen mit dieser Arbeit einmal mehr als unbegründet abgewiesen werden.

In Abschnitt 4.1.2 wurde auf die vermeintliche ökologische Validität existierender Szenarien am Beispiel des TAILORSHOP eingegangen. Zahlreiche unrealistische Systemeigenschaften zeigen schnell, dass hier mitnichten Realität simuliert wird. Nach Kersting (1999) sind die für die Nutzung des Simulationsbegriffs nötigen Voraussetzungen wie Modellanalysen, Spezifikationen von Abbildungsvorschriften und Modellvalidierungen für die meisten Szenarien

---

nicht einmal ansatzweise erfüllt. Komplexe Problemlöseszenarien stellen keine inhaltsvaliden Arbeitsproben dar und müssen die postulierte überlegene Validität wie jedes andere Messinstrument in empirischen Studien beweisen.

Der Anspruch, ein von Intelligenz unabhängiges, neues Konstrukt zu messen, wurde meist mit einer Gegenüberstellung der sehr unterschiedlichen Oberflächenanforderungen von Intelligenztests und komplexen Problemlöseszenarien begründet. Empirisch stützte er sich auf die in frühen Studien gefundenen Nullkorrelationen zwischen komplexer Problemlöseleistung und Intelligenz. Spätere Studien führten diese Befunde auf die mangelnde Reliabilität der eingesetzten Simulationen zurück und fanden vermehrt substanzielle Zusammenhänge insbesondere mit der Intelligenzfacette Verarbeitungskapazität. Im Rahmen dieser Arbeit wurden Studien zum Zusammenhang von komplexer Problemlöseleistung mit Intelligenz gesammelt. Es ergab sich ein über alle Studien gemittelter nach Stichprobengröße gewichteter Zusammenhang von  $r=.32$ . Werden unreliable Simulationen ausgeschlossen, beträgt der Zusammenhang sogar  $r=.37$ . In dieses Bild passen auch die Ergebnisse dieser Untersuchung, die einen über das erworbene Systemwissen vermittelten Zusammenhang der Problemlöseleistung mit Verarbeitungskapazität von  $r=.36$  findet.

Das nährt Zweifel an der Existenz eines die Intelligenzdiagnostik erweiternden, valideren neuen Konstrukts „Operative Intelligenz“. Der in dieser Untersuchung in Stichprobe 3 durchgeführte Vergleich von AGRIMAN mit den eingesetzten Maßen für Verarbeitungskapazität bestätigt diese Zweifel. AGRIMAN korreliert zu  $r=.41$  mit dem Gesamtergebnis des Assessment Centers, die Verarbeitungskapazität hingegen zu  $r=.53$ , was die postulierte Überlegenheit der Simulation eindrucksvoll widerlegt. Die Validität des Problemlöseszenarios resultiert dabei komplett aus der gemeinsamen Varianz mit der Verarbeitungskapazität, es besitzt keine inkrementelle Validität über sie hinaus. Das passt zu der schon zitierten Annahme von Kersting (1999), dass Problemlöseszenarien neben Intelligenz und Wissen nur hochgradig situations- und aufgabenspezifische Fähigkeiten erfordern, die keine diagnostische Relevanz haben. Diese Ergebnisse müssen aufgrund der geringen Stichprobengröße von  $n=54$  sicher mit Vorsicht betrachtet werden, zeigen aber eine klare Tendenz. Für die Problemlöseforschung bedeuten sie, dass mit Simulationen weder ein neues Konstrukt noch ein überlegenes Messinstrument entdeckt wurde. Die Schlussfolgerung von Süß (1996), dass keinerlei empirische Hinweise ein neues Konstrukt rechtfertigen, ist damit einmal mehr bestätigt. Das Forschungsfeld

---

hat sich in dieser Hinsicht als Sackgasse erwiesen, was sicher erklärt, dass die Forschung zum komplexen Problemlösen nahezu zum Erliegen gekommen ist.

Die Ergebnisse sind trotzdem interessant. Immerhin bestätigen Sie das gängige Modell, nach dem Intelligenz über erworbenes Wissen vermittelt auf Leistung wirkt, mit einer neuen Methodik – der Computersimulation.

### **Angewandte Eignungsdiagnostik**

Dass sich die Erwartungen an komplexe Problemlöseszenarien nicht erfüllt haben, spricht zunächst nicht gegen ihren Einsatz in der angewandten Eignungsdiagnostik. Hier kommt es vor allem auf drei Dinge an: Ökonomie, Akzeptanz, und vor allem Validität.

Komplexe Problemlöseszenarien eignen sich vor allem, eine vorhandene Lücke in Assessment Centern und Development Centern zu füllen. Diese Verfahren fokussieren sehr stark die kommunikative und soziale Kompetenz eines Teilnehmers, kognitive Faktoren bleiben tendenziell unterbelichtet. Intelligenztests sind wenig geeignet, diesen Mangel zu beseitigen, da sie von den Teilnehmern mehrheitlich abgelehnt, ja gefürchtet werden. Vor allem in Development Centern, die ja explizit zur Personalentwicklung durchgeführt werden, ist ihr Einsatz schwer zu rechtfertigen. Intelligenz gilt als angeboren, kaum trainierbar und über die Lebensspanne stabil – auf dieser Basis Entwicklungsempfehlungen zu geben ist schwer. Die Etikettierung eines internen Mitarbeiters als „dumm“ kann darüber hinaus einen negativen Einfluss auf dessen zukünftige Motivation und Leistung haben.

Komplexe Problemlöseszenarien haben dieses Problem nicht. Sie sind augenscheinlicher als Intelligenztests, ihre Relevanz für den Berufsalltag wird deshalb von den Teilnehmern eher akzeptiert. Darüber hinaus wird Problemlösefähigkeit als veränderbarer angesehen als Intelligenz, so dass auch einem Einsatz zur Personalentwicklung nichts entgegensteht.

Solcherart von überzogenen Anforderungen befreit können die Ergebnisse angemessen gewürdigt werden. Erstmals wird in dieser Untersuchung der Zusammenhang zwischen Leistungen im Assessment Center und einem komplexen Problemlöseszenario in einer wirklich großen Stichprobe von  $n=185$  Teilnehmern untersucht. Der sich in dieser Stichprobe ergebende Zusammenhang mit dem Gesamtergebnis der Assessment Center oder Development Center von  $r=.35$  für das Endkapital bzw.  $r=.38$  für die aggregierte Skala kann damit als relativ

---

stabil angesehen werden. AGRIMAN steht im Vergleich gut da: In Kapitel 3 wurden Intelligenztests und Assessment Center als valide Auswahlinstrumente vorgestellt, die eine durchschnittliche unkorrigierte Validität um .30 erreichen. Es muss allerdings beachtet werden, dass in den zugrundeliegenden Studien vor allem Vorgesetztenbeurteilungen als Kriterium eingesetzt wurden. Diese weisen eine geringere Reliabilität als Assessment Center auf. Geht man wie Hunter und Hunter (1984) von einer Reliabilität von  $r_{tt}=.60$  für Vorgesetztenbeurteilungen aus und postuliert eine Reliabilität von  $r_{tt}=.80$  für die in dieser Untersuchung eingesetzten Verfahren, ergeben sich sowohl für Intelligenztests und Assessment Center als auch für AGRIMAN eine einfach minderungskorrigierte Validität von  $r=.39$ .

AGRIMAN kann damit guten Gewissens als Baustein zur simulationsbasierten Messung kognitiver Fähigkeiten in Assessment Centern und Development Centern eingesetzt werden. Die Validität ist mit dieser Untersuchung belegt.

Süß (1996) kritisiert am Einsatz komplexer Problemlöseszenarien, dass die Diagnose von Intelligenz und Wissen nicht getrennt werden kann – sie würden nur als *Amalgam* diagnostiziert. Die Frage, welche Kompensationsmöglichkeiten von Einzelfähigkeiten bei einer diagnostischen Fragestellung sinnvoll oder erwünscht ist, kann deshalb nicht vom Diagnostiker bestimmt werden, sondern schmilzt in einen Gesamtscore ein. Diese Kritik geht davon aus, dass komplexe Problemlöseszenarien tatsächlich Vorwissen erfassen. Wie in Abschnitt 6.2 beschrieben wurde, ist jedoch eher davon ausgegangen, dass Sie eine Kompetenz zum Erwerb von Wissen messen. Kersting (1999) sieht darin trotz aller Kritik eine Rechtfertigung des Einsatzes komplexer Problemlöseszenarien zur Provokation und Messung voraussetzungsfreier Wissenserwerbsprozesse.

Doch es bleiben schwere Kritikpunkte. AGRIMAN wurde mit dem klaren Ziel entwickelt, ein messgenaues Diagnoseinstrument zu schaffen. Die in Kapitel 5.3 genannten Punkte zur Optimierung der Reliabilität wurden alle beachtet, und auch die Trennung in Systemexploration und Systemsteuerung sollte zu einer höheren Reliabilität führen. Trotz allem lässt sich ein großer Teil der Varianz in der Problemlöseleistung nicht auf Verarbeitungskapazität zurückführen. Und wenn neben Verarbeitungskapazität ein weiteres Konstrukt beteiligt ist, ist dessen alleinige Varianz nicht valide. Die Simulation weist in dieser Untersuchung keine inkrementelle Validität über Verarbeitungskapazität hinaus auf. Trifft zu, dass über Verarbeitungs-



---

kapazität hinaus nur szenariospezifische Fähigkeiten ohne diagnostischen Wert erfasst werden, ist dies sehr negativ zu werten. Je nach eingesetzter Simulation könnten Ergebnisse der Personalauswahl sehr unterschiedlich ausfallen.

Ein weiterer wichtiger Kritikpunkt ist der im Vergleich zu Intelligenztests höhere Aufwand. AGRIMAN läuft mit Instruktion auch in der kürzesten Version knapp zwei Stunden und erfordert eine Person, die die Programmsteuerung erklärt. Verglichen mit dem ebenfalls in dieser Untersuchung eingesetzten IQO, der weniger als 45 Minuten läuft, ist das ein hoher Aufwand. Leider schlägt sich dieser nicht in einer erhöhten Validität nieder.

Die gefundene beachtliche Kriterienvalidität des Instruments hat also einen bitteren Beigeschmack: Mit der ungeklärten Konstruktvalidität und dem höheren administrativen Aufwand zahlt man einen hohen Preis für die höhere soziale Validität des Instruments. Es sollte deshalb weiter nach sozial validen Maßen kognitiver Fähigkeiten gesucht werden. Inhaltsvalide Wissenstests und Situational Judgment Tests sind ein Ansatz, auch der Versuch, Intelligenztests sozial valider zu gestalten, indem sie anforderungsbezogen eingekleidet werden. Solange solche Alternativen ihre Überlegenheit allerdings nicht bewiesen haben, bleiben komplexe Problemlöseszenarios brauchbare Bausteine in der Personalauswahl und Personalentwicklung.

---

---

## 12. LITERATUR

- Arbeitskreis Assessment Center (Hrsg.) (2004). *Standards der Assessment Center Technik*. Hamburg.
- Beckmann, J.F. & Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. In P.A. Frensch & J. Funke (Eds.). *Complex problem solving. The European perspective*
- Bobko, P. & Roth, P.L. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors and job performance. *Personnel Psychology*, 52, 561-589.
- Borman, W.C., White, L.A., Pulakos, E.D., Oppler, S.H. (1991). Models of Supervisory Job Performance Ratings. *Journal of Applied Psychology*, 76, 863-872.
- Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York: Cambridge University Press.
- Dörner, D. (1975). Wie Menschen eine Welt verbessern wollten. *Bild der Wissenschaft*, 12, 2, 48-53.
- Dörner, D. (1979). Programm Taylorshop in der Version für Ti-59 mit Drucker PC-100. Modifizierte und kommentierte Fassung von Norbert Streitz, Institut für Psychologie der TH Aachen (unveröffentlichtes Computerprogramm).
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290-308.
- Dörner, D., Drewes, U., & Reither, F. (1975). Über das Problemlösen in sehr komplexen Realitätsbereichen. In W. H. Tack (Ed.), *Bericht über den 29. Kongreß der DGfPs in Salzburg 1974* (Vol. Band 1, pp. 339-340). Göttingen: Hogrefe.
- 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 Komplexität. Bern: Huber.
- Dye, D.A., Reck, M., & McDaniel, M.A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 14, 283-302.
- Funke, J. (1985). Steuerung dynamischer System durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie*, 193, 443-465.
- Funke, J. (1985). Problemlösen in komplexen computersimulierten Realitätsbereichen. *Sprache & Kognition*, 3, 113-129.
- Funke, J. (1986). Komplexes Problemlösen: Kritische Bestandsaufnahme und weiterführende Perspektiven. Berlin: Springer.

- 
- Funke, J. (1988). Bedingungen und Auswirkungen der Informationssuche und -aufnahme beim Bearbeiten des komplexen Simulationssystems „TAILORSHOP“, *Berichte aus dem Psychologischen Institut der Universität Bonn*, 14, 5.
- Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Berlin: Springer.
- Funke, J. (2005). Komplexes Problemlösen. In J. Funke (Ed.), *Denken und Problemlösen (=Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 8: Denken und Problemlösen)*. Göttingen: Hogrefe.#
- Funke, U. (1992). *Diagnostisches interaktives System zur Komplexitätssimulation „DISKO/c“*. Handbuch. Filderstadt: Care applications Hoffmann KG (Vertrieb).
- Funke, U. (1993). Computergestützte Eignungsdiagnostik mit komplexen dynamischen Szenarios. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 109-118.
- Funke, U. (1995). Szenarios in der Eignungsdiagnostik und im Personaltraining. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimulierte Szenarios in der Personalarbeit (S.145-216)*. Göttingen: Verlag für Angewandte Psychologie.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. & Benson, C. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology*, 72, 493-511.
- Gediga, G., Schöttke, H. & Tücke-Bressler, M. (1984). Problemlösen und Intelligenz. *Psychologische Forschungsberichte FB 8 der Universität Osnabrück*, Nr. 34.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personel selection. *Personnel Psychology*, 26, 461-477.
- Hartung, S. & Schneider, I. (1995). Entwicklung und Anwendung computersimulierter Szenarien. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimulierte Szenarios in der Personalarbeit (S.219-236)*. Göttingen: Verlag für Angewandte Psychologie.
- Hasselmann, D. (1993). Computersimulierte komplexe Problemstellungen in der Management-Diagnostik. Hamburg: Windmühle.
- Hasselmann, D. (1995). Die Konstruktion computersimulierter Szenarios in der Personalarbeit. In B. Strauß & M. Kleinmann (Hrsg.) *Computersimulierte Szenarios in der Personalarbeit (S. 237-259)*. Göttingen: Verlag für Angewandte Psychologie.
- Hesse, F.W. (1982). Effekte des semantischen Kontextes auf die Bearbeitung komplexer Probleme. *Zeitschrift für experimentelle und angewandte Psychologie*, 29, 62-91.
- Hörmann, H.J. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexen Problemlösen. *Sprache & Kognition*, 1, 23-31.
- Horn, J.L. & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253-270.
- Hunter, J.E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization tot he General Aptitude Test Battery (GATB)*. Washington DC: U.S. Department of Labor, Employment Service.

- 
- Hunter, J.E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hussy, W. (1984). Zum Begriff der Problemschwierigkeit beim komplexen Problemlösen. *Trierer Psychologische Berichte*, 11, 4.
- Hussy, W. (1989). Intelligenz und komplexes Problemlösen. *Diagnostica*, 35, 1-16.
- Hussy, W. (1991). Komplexes Problemlösen und Verarbeitungskapazität. *Sprache & Kognition*, 10, 208-220.
- Jäger, A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jäger, A.O., Süß, H.-M. & Beauducel, A. (1997). Berliner Intelligenzstrukturtest (BIS, Form 4). Göttingen: Hogrefe.
- Kersting, M. (1999). Diagnostik und Personalauswahl mit computergestützten Problemlösenszenarios? Zur Kriteriumsvalidität von Problemlösenszenarios und Intelligenztests. Göttingen: Hogrefe.
- Klingner, Y. & Schuler, H. (2004). Improving participants' evaluations while maintaining validity by a work sample-intelligence hybrid. *International Journal of Selection and Assessment*, 12, 120-134.
- Kröner, S. (2001). Intelligenzdiagnostik per Computersimulation. Berlin: Waxmann.
- Kühle, H.J. & Badke, P. (1986). Die Entwicklung von Lösungsvorstellungen in komplexen Problemsituationen und die Gedächtnisstruktur. *Sprache & Kognition*, 5, 95-105.
- Locher, J. (1997). Transfereffekte bei der Bearbeitung computersimulierter Problemlösenszenarien (Unveröffentlichte Dissertation). Paderborn: Universität-Gesamthochschule Paderborn.
- Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, 18, 685-697.
- Misiak, C. & Kluwe, R.H. (1991). Complex System Control. Research Report No.13, Institut für Kognitionsforschung, Universität der Bundeswehr Hamburg
- Misiak, C., Haider, H. & Kluwe, R.H. (1988). Formale Strukturen von Systemen zur Erfassung kognitiver Leistungsmerkmale. In W. Schönplflug (Hrsg.), *Bericht über den 36. Kongress der DGfP* (Bd. 1, S. 311). Göttingen: Hogrefe.
- Müller, H. (1993). Komplexes Problemlösen: Reliabilität und Wissen. Bonn: Holos.
- Neick, S. (2008). Eine prognostische Validierung des polizeilichen Auswahlverfahrens in Mecklenburg-Vorpommern. Frankfurt: Verlag für Polizeiwissenschaft.
-

- 
- Obermann, C. (1991). Airport Problemlösesimulation V 2.2 Handbuch. Göttingen: Hogrefe.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189, 79-100.
- Putz-Osterloh, W. (1985). Selbstreflexion, Testintelligenz und interindividuelle Unterschiede bei der Bewältigung komplexer Probleme. *Sprache & Kognition*, 4, 203-216.
- Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme. *Zeitschrift für Psychologie*, 195, 63-84.
- Putz-Osterloh, W. (1991). Computergestützte Eignungsdiagnostik: Warum Strategien informativer als Leistungen sein können. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S.97-102). Stuttgart: Verlag für Angew. Psychologie.
- Putz-Osterloh, W., Köster, K. (1988). Diagnostik komplexer Denk- und Entscheidungsstrategien bei einem computersimulierten Planspiel. In Bundesminister für Verteidigung (Hrsg.), *Untersuchungen des Psychologischen Dienstes der Bundeswehr*, 223-255.
- 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.
- Putz-Osterloh, W., Haupts, I. (1990) Diagnostik komplexer Organisations- und Entscheidungsstrategien in dynamischen Situationen. *Untersuchungen des psychologischen Dienstes der Bundeswehr*, 24, 5-48.
- Reichert, U. & Dörner, D. (1988). Heuristiken beim Umgang mit einem „einfachen“ dynamischen System. *Sprache & Kognition*, 7, 12-24.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. & Rolland, J.P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, 88, 1068-1081.
- Schmidt, F.L. & Hunter, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F.L., Hunter, J.E. & Outerbridge, A.N. (1986). The impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432-439.
- Schmitt, N., Gooding, R.Z., Noe, R.A. & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schoppek, W. (1991). Spiel und Wirklichkeit – Reliabilität und Validität von Verhaltensmustern in komplexen Simulationen. *Sprache & Kognition*, 10, 15-27.
- Schoppek, W. (1996). Kompetenz, Kontrollmeinung und komplexe Probleme. Zur Vorhersage individueller Unterschiede bei der Systemsteuerung. Bonn: Holos.
- Schuler, H. (2003). (Hrsg.) Lehrbuch Organisationspsychologie. Bern: Huber.
- Schuler, H. (2005). (Hrsg.) Lehrbuch der Personalpsychologie. Göttingen: Hogrefe.

- 
- Schuler, H. & Funke, U. (1995). Diagnose beruflicher Eignung und Leistung. In: Lehrbuch Organisationspsychologie, 2.Aufl., hrsg. V. Heinz Schuler, Bern 1995.
- Schuler, H. & Klingner, Y. (2005). *Arbeitsprobe zur berufsbezogenen Intelligenz. Büro- und kaufmännische Tätigkeiten (AZUBI-BK)*. Göttingen: Hogrefe.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spies, K., & Hesse, F.W. (1987). Problemlösen. In G. Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (S.371-425). Stuttgart: G. Fischer.
- Stäudel, T. (1987). *Problemlösen, Emotionen und Kompetenz*. Regensburg: Roderer.
- Stern, W. (1911). *Intelligenzproblem und Schule*. Leipzig: Teubner.
- Sternberg, R.J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225-307). Cambridge: Cambridge University Press.
- Streifert, S., Pogash, R. & Piasecki, M. (1988). Simulated-based assessment of managerial competence: Reliability and validity. *Personnel psychology*, 41, 537-557.
- Strohschneider, St. (1986). Zur Stabilität und Validität von Handeln in komplexen Realitätsbereichen. *Sprache & Kognition*, 5, 42-48.
- Strohschneider, St. (1991). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. *Diagnostica*, 4, 353-371.
- Süß, H.M. (1996). Intelligenz, Wissen und Problemlösen. Göttingen: Hogrefe.
- Süß, H.M. (2006). Eine Intelligenz - viele Intelligenzen? - neuere Intelligenztheorien im Widerstreit. In: Wagner, H. (Hrsg.) *Intellektuelle Hochbegabung*. Bonn: Bock.
- 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 Steuerungsleistung an computersimulierten Systemen durch Wissen und Intelligenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 189-203.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Vineberg, R. & Taylor, E.N. (1972). *Performance in four army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance*. Alexandria, VA: Human Resources Research Organization (HUMRRO).
- Wagener, D. (1994). Entwicklung des computergestützten Systems FSYS 1.0 zur Diagnose komplexen Problemlösens. Unveröffentlichte Diplomarbeit, Universität Mannheim.
- Wagener, D. (2001). Psychologische Diagnostik mit komplexen Szenarios. Taxonomie, Entwicklung, Evaluation. Lengerich; Berlin; Riga; Rom; Wien; Zagreb: Pabst Science Publishers.
-

- 
- Wittman, W.W., Süß, H.M., Oberauer, K., Schulze, R. & Wilhelm, O. (1995). Der Zusammenhang von Arbeitsgedächtniskapazität und Konstrukten der Intelligenzstrukturforschung (Berichte des Lehrstuhls Psychologie II der Universität Mannheim). Mannheim: Universität Mannheim, Heft 1.
- Wittman, W.W., Süß, H.M., & Oberauer, K. (1996). Determinanten komplexen Problemlösens (Berichte des Lehrstuhls Psychologie II der Universität Mannheim). Mannheim: Universität Mannheim, Heft 9.
- Wittmann, W.W. & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik-symmetry (pp. 77-108). In P.L. Ackermann, P.C. Kyllonen & R.D. Roberts (Eds.), *Learning and individual differences: Process, trait and content*. Washington: American Psychological Association.