

Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor?¹⁾

Joachim Funke

1. Einleitung

Im Kontext von Untersuchungen zum Verhalten von Menschen in hochkomplexen Problemsituationen taucht die sicher naheliegende Frage nach dem Zusammenhang derartigen Verhaltens mit dem Konstrukt „Intelligenz“ auf. So spricht etwa Kreuzig (1979, S. 207) davon, daß bei hochkomplexen intellektuellen Anforderungen eine geradezu exemplarische Beziehung zu Maßen der Intelligenz erwartet werden muß. Jäger (1982) betont gleichfalls, daß sich multivariate Intelligenzstrukturforschung und experimentelle Denkprozeßforschung auf den gleichen Gegenstand beziehen. In den vorgelegten Arbeiten einer deutschen Arbeitsgruppe (z.B. Dörner 1979a; Kreuzig 1979, 1981; Putz-Osterloh 1981; Putz-Osterloh & Lürer 1981; Dörner, Kreuzig, Reither & Stäudel 1983) stellte sich jedoch – grob gesagt – wiederholt heraus, daß „Testintelligenz“ kein Prädiktor für gutes oder schlechtes Abschneiden in den untersuchten computersimulierten Systemen ist, sondern vielmehr andere Persönlichkeitsvariablen wie etwa „Selbstsicherheit“ wesentliche Varianzanteile aufklären. Allerdings gibt es auch im Rahmen hochkomplexer Systeme gegenteilige Befunde wie etwa die von Berkowitz & Szabo (1978), die für das Paradigma „riddle of the death of the woolly mammoth“ deutliche Vorteile hochintelligenter Vpn aufdecken konnten.

Einige der Resultate werden in dieser Arbeit kritisch untersucht mit der Absicht, Unklarheiten und Probleme bei dieser Art von Problemlöseforschung aufzuzeigen. Dies wird am Beispiel des Kleinsystems „Schneiderwerkstatt“ (SWS; Dörner 1979b) erfolgen. Um es gleich vorweg zu betonen: auch wenn einige – meiner Ansicht nach bedenkliche – Schwachpunkte an den berichteten Befunden aufgezeigt werden, steht der heuristische Wert der bisherigen Studien außer Zweifel.

Bei meiner Kritik geht es um folgende Punkte, die zunächst nur angedeutet werden sollen:

- (1) Die Operationalisierung der Lösungsgüte ist unzureichend. Einmal abgesehen von der Unmöglichkeit zu zeigen, wie eine optimale Eingriffsstrategie aussieht, muß bestritten werden, daß sich durch die bisherige Bestimmung der Lösungsgüte ein valider Indikator für komplexes Problemlösen ergibt.
- (2) Bereits auf der Stufe zuvor, der Reliabilität der abhängigen Variablen „Lösungsgüte“, tauchen Zweifel auf: Die interne Konsistenz des gewählten Summenscores überzeugt nicht und seine Test-Retest-Stabilität ist – wie gezeigt werden kann – schlecht.

¹⁾ Überarbeitetes Manuskript zum Vortrag, gehalten auf der 25. Tagung experimentell arbeitender Psychologen in Hamburg 1983 im Arbeitskreis „Computer-Simulation“ (Uecker/Möbus). Für die kritischen Anregungen und Anmerkungen zu Manuskript und Vortrag danke ich Prof. Dr. J. Bredenkamp, Prof. Dr. A.O. Jäger, Prof. Dr. R. Klüwe und Prof. Dr. G. Lürer.

2.1 Kurzbeschreibung des Systems „Schneiderwerkstatt“

In der vorliegenden Literatur zur SWS (Putz-Osterloh 1981; Putz-Osterloh & Lüer 1981) finden sich eher spärliche und weitgehend textgleiche Beschreibungen. Neben der Darstellung der Variablenvernetzung (vgl. Abb. 1) durch Angabe von groben Wirkungscharakteristiken (Maßnahme wirkt positiv oder negativ auf andere Variable) wird gezeigt, welche Größen vom Probanden beeinflussbar sind und welche nicht.

Die Autoren beschreiben die SWS wie folgt:

„Das System besteht aus 24 netzartig miteinander verknüpften Variablen, von denen 11 durch Maßnahmen der Pbn direkt beeinflussbar sind. Das Kapital stellt die ‚Kernvariable‘ des Systems dar, da es mit den meisten, nämlich 15 Variablen, in direkter Verbindung steht. – Einige Variablen unterliegen in ihrem Zustand Zufallsschwankungen und werden somit nicht allein durch den Zustand der Beeinflussungsvariablen bestimmt. Der Effekt von Veränderungsmaßnahmen wirkt sich nicht immer sofort, sondern z.T. erst mit einer Zeitverzögerung auf den Zustand direkt benachbarter, d.h. verknüpfter Variablen aus. – Das System ist so programmiert, daß die Werkstatt ohne Eingriffe in kurzer Zeit Konkurs anmelden müßte. Denn Rohmaterial muß eingekauft werden, Reparaturkosten sind z.B. zu erhöhen usw., wenn der Betrieb bei bestehender Produktionskapazität Gewinn erwirtschaften soll. – Aufgabe des Problemlösers ist es, die Schneiderwerkstatt über einen vorgegebenen Zeitraum von simulierten Monaten hinweg mit Hilfe vorgegebener Maßnahmen so zu leiten, daß der Betrieb maximalen Gewinn auf die Dauer erwirtschaftet.“ (Putz-Osterloh & Lüer 1981, S. 313–315).

Tabelle 1:
Liste der Variablen im Programm TAILOR-Shop

Kürzel	Bezeichnung	Anfangswert
ZA	aktuelle Zufriedenheit	0.981
PA	aktuelle Produktion	403.932
PM	mögliche Produktion	403.932
KL	Hemden im Lager	80.716
RL	Rohmaterial im Lager	16.068
NA	Nachfrage aktuell	766.636
VH	verkaufte Hemden	407.216
PH	Preis pro Hemd	52.000
RS	Reparatur- und Servicekosten	1200.000
MA	aktuelle Maschinenkapazität	47.040
WE	Werbe-Kosten	2800.000
LO	Lohnkosten pro Arbeiter	1080.000
SM	Sozialausgaben pro Arbeiter	50.000
N1	Anzahl Arbeiter an 50er	8
N2	Anzahl Arbeiter an 100er	0
A1	Anzahl 50er-Maschinen	10
A2	Anzahl 100er-Maschinen	0
LW	Anzahl Lieferwagen	1
GL	Geschäftslage (0 = Vorort, 0.5 = Cityrand, 1 = City)	0.5
KA	Kapital	15774.659
RP	Rohmaterial-Preis	3.994
ND	Anzahl registrierter Monate	0

In Tabelle 1 findet sich eine Liste der in der SWS beteiligten Variablen sowie ihrer Kürzel, in Tabelle 2 die Gleichungen, aus denen die Vernetzung der Variablen hervorgeht.

Aus Tabelle 2 geht hervor, daß in den Gleichungen (2), (6b) und (8) die Werte für die mögliche Produktion (PM), Nachfrage für den nächsten Monat (NA₂) sowie Preis des Rohmaterials (RP) innerhalb gewisser Grenzen „zufällig“ schwanken können. Nur eine davon – Preis des Rohmaterials – ist dem Probanden bekannt, er kann sich entsprechend darauf einstellen. Auf die beiden restlichen dagegen kann er nicht reagieren, da sie „verdeckt“ sind. Die Zufallssequenzen bleiben für alle Vpn identisch.

Bevor die Ergebnisse einer eigenen Studie präsentiert werden, soll eine für deren Verständnis zentrale Frage, nämlich die nach der Bestimmbarkeit optimaler Systemeingriffe, kurz diskutiert werden.

Tabelle 2:
Vernetzung der Variablen in der Schneiderwerkstatt*

Nr. der Gleichung	Variable	Beziehung
(1)	ZA	$ZA = \text{Min}(ZM, (0.5 + ((LO - 850)/550) + SM/800))$
(2)	PM	$PM = (\text{Min}(N1, A1) \cdot (MA + \xi \cdot 4 - 2) + \text{Min}(N2, A2) \cdot (MA \cdot 2 + \xi \cdot 6 - 3)) \cdot \sqrt{ZA}$
(3)	PA	$PA = \text{Min}(PM, RL)$
(4)	HL	$HL = HL + PA - VH$
(5)	RL	$RL = RL - PA$
(6a)	NA ₁	$NA_1 = NA_2/2 + 280 \cdot 1.25 \cdot e^{-(PH^2/4250)}$
(6b)	NA ₂	$NA_2 = \text{Min}(WE/5, NM) + LW \cdot 100$ $NA_2 = NA_2 + NA_2 \cdot g + \xi \cdot 100 - 50$ wobei $g = 0$ für $GL = \text{Vorort}$ $= 0.1$ für $GL = \text{Cityrand}$ $= 0.2$ für $GL = \text{City}$
(7)	VH	$VH = \text{Min}(HL, NA_1)$
(8)	RP	$RP = 2 + \xi \cdot 6.5$
(9)	MA	$MA = \text{Min}((MA - 0.1 \cdot MA + (RS/(A1+A2)) \cdot 0.017), MM)$
(10a)	KA	$KA = KA - SM \cdot (N1+N2) - PA \cdot 1 - RL \cdot 5$ $- HL \cdot 1 + VH \cdot PH - WE - LW \cdot 500$ $- GL \cdot 2000 - RS - (N1+N2) \cdot LQ$
(10b)	KA	$KA = KA + KA \cdot z$ wobei $z = 0.0025$, wenn $KA \geq 0$ $= 0.0066$, wenn $KA < 0$

* Das Zeichen ξ steht für eine Zufallszahl im Bereich zwischen Null und Eins, erzeugt von einem rechnerinternen Pseudozufallszahlengenerator.

2.2 Bestimmung der Lösungsgüte: das zentrale Kriterium

Von zentraler Bedeutung ist die Bestimmung eines Kriteriums der Lösungsgüte, anhand dessen Probanden als „gute“ oder „schlechte“ Problemlöser zu identifizieren sind. In allen vorliegenden Studien findet dieses *post-experimentelle* Klassifikationsprinzip Anwendung, bei dem nicht nur die *Möglichkeit*, sondern auch die *Richtigkeit* einer Extremgruppenbildung oder Stichprobenhalbierung zur Beschreibung von Befunden aus Verhaltensdaten wie aus Daten der vielfach erstellten Denkprotokolle unterstellt wird. Deswegen sei diese Vorgehensweise gründlicher untersucht.

Mit der Feststellung der Lösungsgüte in L o h h a u s e n hat sich K r e u z i g (1979) beschäftigt. Dort kamen in der Mehrzahl subjektive Kriterien zum Einsatz, die recht hoch untereinander korrelieren und somit auch als reliable Meßwerte anzusehen sind. Im einzelnen erfolgte nach der Auswahl sogenannter „Kernvariablen“ (insgesamt 17) ein unabhängiges Experten-Rating durch 5 Versuchsleiter sowie durch „naive“ Probanden. Innerhalb und zwischen den Gruppen fanden sich substantielle Korrelationen der Beurteilungen, die von den Autoren als Hinweis auf ein „von möglichen Interaktionseffekten zwischen Vpn und VI unbeeinflusstes Gütekriterium“ (D ö r n e r et al. 1983, S. 162) interpretiert wurden. Nebenbei bemerkt halten wir eine andere Sichtweise dieses Sachverhalts für möglich: daß nämlich die Experten keine so profunden Expertisen ausstellten, sondern wohl eher naiven Kriterien wie dem Auf und Ab der Kurven oder der Bedeutung der gezeigten Variablen etc. erlegen sind. Im Normalfall würde man ja gerade eine Diskrepanz zwischen naivem und Expertenrating erwarten.

In der SWS finden sich vergleichsweise spärliche Angaben zum zentralen Gütekriterium. Als „Erfolgsmaß“ wird hier die Anzahl der Monate ausgezählt, in denen das (Flüssig-)Kapital KA einen Aufwärtstrend zeigt. Bei P u t z - O s t e r l o h (1981) schätzen analog zur Prozedur bei L o h h a u s e n sieben Beurteiler anhand von Grafiken die Aufwärtsentwicklung von „12 wichtige Systemvariablen“ (welcher?) auf einer fünfstufigen Ratingskala ein. Diese Urteile korrelieren mit dem Kriterium zu 0.83 (N = 83), der Übereinstimmungskoeffizient der sieben Beurteiler lag bei 0.94. Damit hält die Autorin die „Gültigkeit des Kriteriums zur Beschreibung des Problemlöseerfolgs . . . für hinreichend gesichert.“ (S. 88). In der Replikationsstudie von P u t z - O s t e r l o h & L ü e r (1981) werden dazu keine Angaben mehr gemacht, wie überhaupt diese Problematik dort nicht weiter thematisiert wird.

Die verschiedenen Bemühungen zur Definition von Systemgütekriterien für L o h h a u s e n und SWS reflektieren die Einsicht, mit der Maßfrage an einem zentralen Punkt angelangt zu sein. Was bei diesen Anstrengungen jedoch am meisten überrascht, ist der weitgehende Verzicht auf *objektive* Gütemaßstäbe (vgl. dagegen F u n k e 1981). Selbst die Auswahl einiger als wesentlich eingeschätzter Variablen, die den Ehrentitel „Kern“-Variable tragen dürfen, ist nicht mathematisch bzw. systemtheoretisch fundiert (es sei denn, man akzeptiert die Auswahl der Variable mit den meisten Verzweigungen als eine systemtheoretische), obwohl sich an einigen Stellen Ansätze dazu finden lassen (etwa bei D ö r n e r et al. 1983, S. 44f.). Dabei fragt man sich, warum Untersucher, die ihre Probanden an computersimulierten Systemen herumwerkeln lassen, nicht gelegentlich ihren eigenen Rechner befragen, wie er das Problem

lösen würde. Dies ist als Hinweis auf die so sehr gewünschte Realitätsnähe zu verstehen: welcher Manager, Minister oder Wissenschaftler ließe nicht Rechner einsetzen zur Bearbeitung von Fragestellungen, die seine beschränkten menschlichen Kapazitäten überfordern?

Am konkreten Beispiel der SWS erscheint es naheliegend, sich zunächst einmal zu fragen, wie solche Betriebe optimal geführt werden können (vgl. P e l z m a n n 1982). Seit den 40er Jahren existiert mit den Methoden der linearen und nicht-linearen Optimierungsverfahren eine Möglichkeit, komplexe Zusammenhänge optimal zu lösen bzw. festzustellen, ob überhaupt ein Optimum existiert. „Die ersten großen und bedeutenden Anwendungen auf wirtschaftliche Probleme erfolgten bei der Planung von Ölraffinerien. Mittlerweile gibt es Anwendungen auf fast allen Gebieten der Wirtschaft. Der Grund hierfür liegt in der zunehmenden Kompliziertheit der ökonomischen und technischen Prozesse, die häufig nur noch mit Hilfe mathematischer Methoden bewältigt werden können.“ (S c h i c k 1981, S. V).

Ein Versuch zur Optimierung des Systems für den ersten Monat wurde von mir unter Mithilfe von Mathematikern³⁾ der Universität Trier durchgeführt, allerdings nur für die Linearfunktionen des Systems (als bekannt wurden dabei angenommen: alle „Zufallsgrößen“ sowie PH, N1, N2, A1, A2, GL). Als Zielfunktion diente das Flüssigkapital (KA). Mittels des Simplexalgorithmus (BASIC-Programm ‚LINEAR-OPTI‘ von P o o l e & B o r c h e r s 1980) wurde dieses restringierte Problem eingegeben mit dem Resultat, daß eine *Optimierung des linearen Gleichungssystems nicht möglich war!*

Die Tatsache, daß selbst ein verkürztes System von einem speziellen Algorithmus zur Optimierung linearer Gleichungen als nicht lösbar bewertet wird, macht es schwierig, nun Vpn-Verhalten als „gut“ bzw. „schlecht“ im Hinblick auf die zu optimierende Zielfunktion zu klassifizieren. Offensichtlich beschreiben die Funktionen des Systems SWS keinen im mathematischen Sinn beschränkten Lösungsraum (d.h. der Lösungsraum läßt sich nicht als Polyeder kennzeichnen).

Versucht man dennoch, Qualitätsunterschiede zu definieren, können sich diese nicht mehr auf die Handhabung des ganzen Systems beziehen, sondern nur noch auf Teile davon. Das hat aber zur Folge, daß eine Definition guter oder schlechter Problemlöser auf der Basis von Teilerfolgen nur eine unvollständige und möglicherweise falsche Beurteilung impliziert, somit also die Validität der Einteilung in bezug auf die Handhabung des Gesamtsystems fraglich ist.

3. Schilderung einer „Replikations“-studie

Nachfolgend wird über ein Experiment zur SWS berichtet, das einige zentrale Annahmen und Ergebnisse bisheriger Untersuchungen überprüft. Die implementierte Fassung des SWS weicht dabei in einigen Punkten von der bekannten Standard-Prozedur ab. Deswegen ist es vielleicht nicht ganz richtig, von einer Replikation zu sprechen; dennoch besteht eine so hohe Ähnlichkeit der Untersuchungen, daß dieser Begriff zumindest nahegelegt wird.

³⁾ Ich danke Herrn Dr. W r o b e l sowie Herrn Dr. Z e n c k e für ihre freundliche Hilfe.

Zunächst sollen die Untersuchungsziele kurz dargelegt werden, danach folgen eine Beschreibung von Versuchsplanung und Durchführung, Erläuterungen zur Wahl der AV sowie schließlich die Resultate.

3.1 Hypothesen

Von zentraler Bedeutung ist die Frage nach der Vorhersagbarkeit der Problemlösungsgüte durch Testintelligenz sowie die mögliche Modifikation dieser Beziehung durch den Faktor „Transparenz der Problemstellung“. Aus der Diskussion der genannten Arbeiten lassen sich folgende Prognosen ableiten:

- H1:* Es besteht ein Haupteffekt „Testintelligenz“ in bezug auf die AV Lösungsgüte. Diese Annahme steht in Widerspruch zu den berichteten Befunden von D ö r n e r et al. (1983) sowie P u t z - O s t e r l o h (1981) und P u t z - O s t e r l o h & L ü e r (1981), berücksichtigt aber die von uns geäußerte Kritik an den genannten Untersuchungen. In der Terminologie von C o h e n (1977) wird ein „large effect“ (aufgeklärte Varianz von mindestens 14%) erwartet.
- H2:* Es besteht ein Haupteffekt „Transparenz“ derart, daß unter transparenten Bedingungen Probanden besser als unter intransparenten mit dem System umgehen können. Diese Prognose basiert auf der Annahme, daß mehr Informationen über ein System dessen „Regierbarkeit“ erhöhen. Angesichts der von P u t z - O s t e r l o h (1981) sowie P u t z - O s t e r l o h & L ü e r (1981) berichteten Befunde zur Wirkung des Faktors Transparenz wird ebenfalls ein „large effect“ erwartet.
- H3:* Es besteht eine Interaktion „Testintelligenz x Transparenz“ derart, daß unter transparenten Bedingungen Hochintelligente einen deutlicheren Vorteil gegenüber Niedrigintelligenten besitzen als unter Intransparenz. Diese Hypothese folgt der Argumentation, daß testintelligente Vpn durch das zusätzliche Informationsangebot, das die Problemsituation stärker an die Intelligenztestsituation annähert, ihre Fähigkeiten effektiver einsetzen können. Hierfür wird allerdings ein schwächerer Effekt („medium“; Varianzaufklärung von 6%) erwartet, da die Haupteffekte als im wesentlichen unabhängige Größen angesehen werden.

Die apriori Berechnung des Stichprobenumfangs (siehe C o h e n 1977) stellt nach B r e d e n : a m p (1980) einen wesentlichen Bestandteil nicht nur von Replikationsstudien dar. Vier Parameter sind zur Bestimmung von N bzw. n' ($\hat{=}$ Zellbesetzung) wichtig: ein Maß für die in der Population vermutete Effektstärke (hier mit f abgekürzt), die Zahl der Freiheitsgrade für die Hypothese (u) sowie die akzeptierten Werte für Alpha- und Beta-Fehler. Für H1, die im Zentrum unseres Interesses steht, ergibt sich nach den genannten Angaben über die vermutete Effektstärke ($f = .40, u = 1$), einem akzeptierten Alpha-Risiko von 0.10 und einem Beta-Risiko von 0.20 ein n' von 10.5, also ein Gesamt-N von 44. Diese Angabe gilt gleichermaßen für H2. Für H3 ($f = .25, u = 1$) errechnet sich n' mit 25.5, d.h. einem Gesamt-N von 104. Da H1 von stärkstem Interesse ist, soll sich daran die Stichprobengröße orientieren. Die Teststärke für die vermutete Interaktion wird somit gering sein.

3.2 Versuchsplan

Zur Überprüfung von H1, H2 und H3 wurde ein Versuch zur SWS mit folgenden orthogonalen Faktoren realisiert: (1) Faktor A „Testintelligenz“ mit den Stufen niedrig/hoch, (2) Faktor B „Transparenz des Systems“ mit den Stufen niedrig/hoch. Beim ersten Faktor handelt es sich nicht um eine „echte“ experimentelle Variable, da sie nicht vom VI manipuliert werden kann.

Die Auswahl von Vpn gemäß den Stufen von Faktor A erfolgte nach einer Voruntersuchung: insgesamt 105 Psychologiestudenten des ersten Semesters bearbeiteten Set I und II der „Advanced Progressive Matrices“ von Raven (1965). Abbildung 2 zeigt die Verteilung des APM-Rohwertes in der Stichprobe.

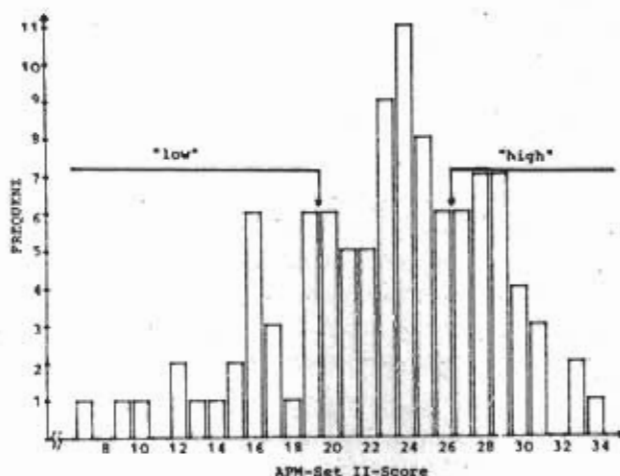


Abb. 2: Verteilung der 105 Vpn auf die Rohwerte im APM-Set II mit Angabe der cut-off-Werte.

Anhand dieser Verteilung wurde unter Berücksichtigung des geplanten Stichprobenumfangs das obere und untere Quartil der Stichprobe ermittelt (siehe Grenzwerte in Abbildung 2) und die jeweiligen Vpn zur Untersuchung gebeten.

Faktor B wurde analog zu Putz-Osterloh (1981) operationalisiert: Der einzige Unterschied der Transparenz- gegenüber der intransparenten Bedingung besteht in der Verfügbarkeit des Vernetzungsschemas (vgl. Abbildung 1). Allerdings war das Niveau der Transparenz in unserer Studie insgesamt höher. Alle Informationen standen auf dem Bildschirm des Kleincomputers zur Verfügung. Eingriffswünsche wurden jeweils vom Rechner abgefragt (siehe unten) und nicht – wie in der ursprünglichen Fassung – aktiv von der Vp formuliert und dann über einen VI vermittelt. Zusätzliche und in bezug auf die Systemtransparenz wichtige Veränderungen sind: (1) die Anzeige der in der originalen SWS verdeckten Variablen „Nachfrage“ und „Zufriedenheit“ sowie (2) die Transformation von Parametern in prozentuale Angaben (aktueller Wert

bezogen auf Maximum) bei Maschinenkapazität, Produktionsquote und Zufriedenheit. Mit den zuletzt genannten Umwandlungen erspart man der Vp überflüssige Berechnungen. Daß sich bei einer Vereinfachung der Diagnose bestimmter Bereiche auch eine Veränderung der Transparenz ergibt, ist nicht anzunehmen: der Kenntnisstand über das System wächst im Unterschied zur erstgenannten Änderung damit nicht.

Bei der Zuweisung der Vpn zu den beiden Transparenzbedingungen wurde darauf geachtet, zunächst die Zellen mit intransparenter Bedingung zu füllen, um den kaum kontrollierbaren Informationsfluß zwischen den Vpn gering zu halten.

3.3 Durchführung der Untersuchung⁴⁾

Insgesamt 53 Probanden bearbeiteten die SWS in rund zweistündigen Sitzungen. Zunächst wurde der „Fragebogen für kognitive Prozeßvariablen“ (FKP) von Kreuzig (1981) vorgelegt, dessen 18 Items die beiden Subskalen „kontrolliertes divergentes Denken“ (KDD; 7 Items) und „Steuerbarkeit der Aktivierung von divergentem und konvergentem Denken“ (STA; 11 Items) konstituieren. Der Summenwert dieser Skalen erwies sich in der Lohhausen-Studie mit einem $r = 0.57$ als optimaler Prädiktor des Erfolgs.

Nach Beantwortung der FKP-Items setzten sich Vp und VI gemeinsam vor den Kleincomputer. Dort wurde zunächst eine allgemeine Instruktion erteilt, bei der die Vp aufgefordert wurde, die SWS „so gut wie möglich“ zu leiten. Zugleich gab es Informationen über wichtige Konstanten (z.B. Anschaffungskosten, Zinsen, Miete etc.) sowie einschränkende Bedingungen. Im Anschluß an diese allgemeinen Erläuterungen bestand dann in einer simulierten zweimonatigen Probezeit Gelegenheit, unter Anleitung des VI die Handhabung des Computers (Eingabe, Korrekturen etc.) sowie den Ablauf des Programms zu erfahren. Im ersten Monat wurde jeder mögliche Eingriff ausprobiert, wobei die Vp dessen Form bestimmte. Bereits im zweiten Probemonat konnte sie gänzlich eigenständig – wenngleich noch immer unter den Augen des VI – entscheiden. Die zwölf Ernstmonate bearbeitete die Vp dann selbständig und ohne Zeitbegrenzung. Als „Honorar“ wurden $1\frac{1}{2}$ Vpn-Stunden bescheinigt.

Jeder Monat begann mit der Präsentation des aktuellen Standes der Parameter (siehe Abbildung 3).

Nach Inspektion dieser Daten durch die Vp fragte das Programm dann für jeden Eingriffsbereich ab, ob eine Veränderung vorgenommen werden solle oder nicht. Durch einfachen Tastendruck (J oder N) gab die Vp ihren Wunsch bekannt; bei positivem Votum wurde der alte Stand noch einmal präsentiert und der neue erfragt. Die Vp teilte daraufhin ihre Entscheidung via Tastatur dem Rechner mit, der zugleich die Eingabe auf Zulässigkeit prüfte. So konnte man etwa – gemäß den Spielregeln der Originalform – den Lohn nicht unterhalb von 850 festsetzen oder keine 100-Hemden-Maschine kaufen, wenn der Anlagenzustand mehr als 30% Schäden aufwies. Auch Plausibilitätskontrollen waren bei jedem Eingriff automatisch vorgesehen, etwa wenn die Vp mehr Maschinen verkaufen wollte als sie besaß. Abbildung 4 zeigt einen Ausschnitt aus dem „Dialog“ mit dem SWS-Programm.

⁴⁾ Für die gelungene Durchführung der Untersuchung danke ich ganz herzlich den Studenten A. Buchner, I. Kirchner, W. Kolb und B. Müller. Ihrer Sorgfalt und ihrem Engagement ist u.a. die geringe Ausfallquote zu verdanken.

----- DER ZUSTAND IHRES LADENS IM MONAT 0: -----	
KAPITAL	: 15774.66
VERKAUFTE HEMDEN	: 407
NACHFRAGE (AKTUELL)	: 767
ROHMATERIAL: PREIS	: 3.99
ROHMATERIAL: IM LAGER	: 16
FERTIGE HEMDEN IM LAGER	: 81
50-HEMDEN-MASCHINEN	: 10
ARBEITER FÜR 50ER	: 8
100-HEMDEN-MASCHINEN	: 0
ARBEITER FÜR 100ER	: 0
REPARATUR & SERVICE	: 1200
LOHN PRO ARBEITER	: 1080
SOZIALKOSTEN PRO ARBEITER	: 50
PREIS PRO HEMD	: 52
AUSGABEN FÜR WERRUNG	: 2800
ANZAHL LIEFERWAGEN	: 1
GESCHÄFTSLAGE	: CITYRAND
ARBEITZUFRIEDENHEIT	IN#: 57.69
MASCHINENSCHADEN	IN#: 5.92
PRODUKTIONSFAHLL	IN#: 0
(WEITER MIT 'RETURN'-TASTE)	

Abb. 3: Präsentation des monatlichen Zustands der Schneiderwerkstatt auf dem Monitor (Beispiel).

----- EINGABE VON MASSNAMEN FÜR MONAT 1: -----	
ROHMATERIAL BESORGEN	(J/N)_J
IM LAGER BEFINDET SICH ROHMATERIAL FÜR 16 HEMDEN. ROHMATERIAL FÜR EIN HEMD KOSTET MOMENTAN 3.99 DM	
FÜR WIEVIEL HEMDEN WÜCHTEN SIE ROHMATERIAL BESCHAFFEN: ?_100	
PREIS PRO HEMD ÄNDERN	(J/N)_N
KOSTEN FÜR WERRUNG ÄNDERN	(J/N)_N
ARBEITER ANSTELLEN/ENTLASSEN	(J/N)_N
MASCHINEN: KAUF-VERKAUF-TAUSCH	(J/N)_N
REPARATUR/SERVICEKOSTEN ÄNDERN	(J/N)_N
(usw.)	

Abb. 4: Ausschnitte aus dem Mensch-Maschinen-Dialog über gewünschte Veränderungen und Eingriffe in das System (Probendateneingaben sind mit „-“ angezeigt).

Durch diese Präsentationsform konnte eine Problembearbeitung erzielt werden, bei der der Filter „Versuchsleiter“ ausgeschaltet blieb. Dies scheint mir von großer Bedeutung zu sein, wenn man etwa zum Vergleich das Vorgehen bei L o h h a u s e n betrachtet. Dort mußte etwa der VI entscheiden, auf welchem Niveau sich die Frage einer Vp bewegte, um eine adäquate Antwort zu geben: fragte z.B. jemand nach dem Zustand der Maschinen, war die Antwort „mittelmäßig“; fragte er nach der Zahl voll funktionsfähiger Maschinen, bekam er die exakten Angaben (nach D ö r n e r et al. 1983, S. 130). Inwiefern hier VI-Effekte von Bedeutung sein können, möge der Leser selbst entscheiden.

Während der Bearbeitung durfte sich die Vp Aufzeichnungen machen. Die Systemdaten der 12 Ernstmonate sowie intern gemessene Entscheidungszeiten wurden nach Abschluß der Bearbeitung auf Datenträger aufgezeichnet, eine Registrierung der Aktionen während der Probezeit erfolgte nicht.

Nach zeitlicher unbegrenzter Bearbeitung der 12 „Ernst“monate beantwortete die Vp einige Standardfragen über die empfundene Wichtigkeit von Systemvariablen, über die verfolgten Ziele sowie – in der intransparenten Gruppe – über die vermutete Vernetzung der Variablen. Außerdem erfolgte eine nochmalige Vorlage des FKP.

Insgesamt erstreckte sich die Haupt-Untersuchung auf einen Zeitraum von ca. 3 Wochen im Januar 1983. Fünf Vpn fielen aus technischen Gründen (defektes Aufzeichnungsgerät) aus, die definitive Zahl verwertbarer Vpn lag bei $N = 12$ pro Zelle. Angaben zu Stichprobenmerkmalen findet man in Tabelle 3.

Tabelle 3:
Deskriptive Merkmale der vier Untersuchungsgruppen
(I = Intelligenz, T = Transparenz, „-“ = niedrig, „+“ = hoch)

Gruppe	Geschlecht		APM-Set II		KDD*		STA*	
	w	m	\bar{x}	s	\bar{x}_1	\bar{x}_2	\bar{x}_1	\bar{x}_2
I- T-	7	5	14.7	3.92	35.9	32.7	40.3	41.2
I- T+	8	4	16.5	3.06	34.0	32.3	36.7	37.8
I+ T-	7	5	29.4	1.83	42.0	41.8	45.7	48.1
I+ T+	5	7	29.2	2.08	39.1	39.2	40.0	40.2

* Aus dem „Fragebogen zur Erfassung kognitiver Prozeßvariablen“ von K r e u z i g (1981) abgeleitete Skalenwerte; KDD = kontrolliertes divergentes Denken, STA = Steuerbarkeit der Aktivierung von divergentem und konvergentem Denken.

Die Aufteilung der Geschlechter auf die Versuchsbedingungen kann als gleichmäßig bezeichnet werden (Chi-Quadrat = 1.61 bei $df = 3$, n.s.). Hinsichtlich des gruppenbildenden Merkmals „APM-Score“ finden sich innerhalb jeder Gruppe von I keine Unterschiede für die zwei T-Bedingungen, jedoch liegt für I- eine insgesamt erhöhte Streuung im Vergleich zu I+ vor. Gemäß den Angaben von R a v e n (1965) darf man die Mittelwerte der beiden Gruppen mit Perzentilwerten von ungefähr 75 bzw. oberhalb von 95 gleichsetzen.

Die nächsten vier Angaben in Tabelle 3 beziehen sich auf den bereits erwähnten FKP von K r e u z i g (1981) mit den Subskalen KDD sowie STA, wobei der Index den Zeitpunkt der Vorgabe (1 = vor, 2 = nach Systembearbeitung) markiert. Weitere Details über die prä- bzw. postdiktive Valenz dieses Tests finden sich im Ergebnisteil.

3.4 Operationalisierung der AV

Um die Kritik an der bisherigen Art der Erfolgsmessung konstruktiv weiterzuführen, wird als verbessertes und dennoch vergleichbares Erfolgsmaß die Anzahl der Monate mit einem Aufwärtstrend im GESAMT-Kapital der SWS herangezogen (nach L ü e r 1983 wird in neueren Arbeiten ähnlich vorgegangen). Im Unterschied zum bereits bekannten Flüssigkapitaltrend liegen der Berechnung des Gesamtkapitals pro Monat folgende Bestimmungsgrößen zugrunde: (1) das Flüssigkapital, (2) der Zeitwert vorhandener Maschinen, wofür der Zustand der Maschinen (MA/MM) ausschlaggebend ist, (3) der laufzeitabhängige Buchwert von Lieferwagen sowie (4) das in Rohwaren- und Fertigteillager vorhandene Material zu einem Konkurswert von 2.- bzw. 20.- pro Stück.

Die Gleichung zur Bestimmung des Gesamtkapitals GK lautet also unter Verwendung der aus Tabelle 1 bekannten Kürzel:

$$GK = KA + A1 \cdot X \cdot 8000 + A2 \cdot X \cdot 16000 + LW \cdot (8000 - ND \cdot 100) + RL \cdot 2 + HL \cdot 20 \quad (1)$$

wobei: $X = MA/MM$.

Natürlich sind auch andere Formen der Bilanzierung denkbar, in denen beispielsweise die Zufriedenheit der Arbeiter oder die Anzahl geschaffener Arbeitsplätze oder noch weitere Variablen einfließen. Bei dem hier vorgestellten Maß GK handelt es sich um eine Art „monatlicher Bilanz“, in die wesentliche kostenintensive Größen des Systems einfließen. Ein Vorteil dieser Bewertung besteht darin, daß unterschiedliche Investitionstätigkeiten nicht mehr auf die AV einwirken: selbst wenn eine Vp monatlich eine Maschine zu 20000,- erwirbt und damit im Flüssigkapital kontinuierlich Verluste macht, wird diese Investition im bilanzierten Gesamtkapital verrechnet. Ein Auf- oder Abwärtstrend in der AV entsteht also nicht mehr durch unterschiedliches Investitionsverhalten, sondern im wesentlichen durch Effektivität in Produktion und Verkauf.

4. Untersuchungsergebnisse

Nach diesen umfangreichen Erläuterungen des „Reiz“materials und seiner Auswertung sollen nun die Befunde der Hypothesentestung berichtet werden. Neben dem (o.g.) neuen Erfolgsmaß „Trend im Gesamtkapital“ soll zum Vergleich auch die alte AV „Trend des Flüssigkapitals“ herangezogen werden (die Korrelation beider AVn beträgt übrigens 0.66). Zunächst erfolgt die Darstellung der hypothesenprüfenden und dann der explorativen Datenanalyse.

4.1 Hypothesenprüfende Datenanalyse

Zur Überprüfung der drei Hypothesen wurde eine 2x2-Varianzanalyse mit den AVn „Kapitaltrend“ nach P u t z - O s t e r l o h (TRENDPO) sowie „Kapitaltrend“ nach F u n k e (TRENDFU) durchgeführt. In beiden Fällen ergab eine Überprüfung auf

Varianzhomogenität mittels F-Max-Tests keine Abweichungen von dieser Annahme. Die Ergebnisse beider Analysen sind in Tabelle 4 dargestellt.

Tabelle 4:
Ergebnisse der 2x2-Varianzanalyse mit der AV „Aufwärtstrend“ nach den
Auswertungsvorschlägen von Putz-Osterloh (TRENDPO) und
Funke (TRENDFU)

Quelle der Variation	df	AV: TRENDPO				AV: TRENDFU			
		MS	F	p	η	MS	F	p	η
A (Testintelligenz)	1	7.52	0.96	0.33	0.14	35.02	3.92	0.054	0.28
B (Transparenz)	1	0.19	0.02	0.88	0.02	15.19	1.70	0.20	0.18
A x B	1	11.02	1.41	0.24	0.17	11.02	1.23	0.27	0.16
Fehler	44	7.81				8.94			

Neben den Angaben zu den „mean squares“ (MS) sowie dem F-Wert gibt Eta die Korrelation zwischen UV und AV an.

Die Analyse mit der AV „TRENDPO“ liefert keinen bedeutsamen Effekt: weder Haupteffekte noch die Interaktion lassen einen Einfluß der Faktoren auf die AV vermuten. Die Nullhypothesen zu H1, H2 und H3 können nicht verworfen werden. Das multiple R zwischen allen UVn und der AV liegt mit 0.227 ziemlich niedrig.

Bei der AV „TRENDFU“ muß die Nullhypothese zu H1 über einen ausbleibenden Intelligenzeffekt verworfen werden. Die niedrig testintelligenten Vpn sind mit 2.25 Monaten Aufwärtstrend weniger erfolgreich als hochintelligente mit einem Wert von 3.96. H2 (Transparenzeffekt) und H3 (Interaktion) erhalten dagegen in bezug auf das Maß der Problemlösegröße keine Stützung aus den Daten. Das multiple R zwischen allen Prädiktoren und der AV beträgt 0.367 für diese Analyse.

Eine post-hoc-Bestimmung der Teststärken ergibt für die angenommenen Effektgrößen (H1, H2 groß; H3 mittel), einem Alpha = 0.10 sowie einem $n' = (48-4)/2+1=23$ eine Power von 0.86 für H1 und H2 sowie von 0.53 für H3. Daraus ist zu folgern, daß mit ziemlicher Sicherheit (nämlich Beta = 0.14) kein Haupteffekt Transparenz in der angenommenen Größe in der Population vorliegt, während über die Gültigkeit von H3 wegen Beta = 0.47 erst nach einer Vergrößerung der Stichprobe entschieden werden sollte (vgl. Westermann & Hager 1982).

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. Daß mit TRENDPO keine ideale Bewertung möglich ist, wurde schon gezeigt. Weiteres Material für die Behauptung, daß TRENDFU zu favorisieren ist, wird nachfolgend geliefert.

Unabhängig von der Tatsache, daß TRENDPO andere Gütekriterien als TRENDFU anlegt, stellt sich nämlich die Frage nach der *Reliabilität der AV*. Zu diesem Zweck wurden 14 Vpn nach einem Intervall von zwei Wochen zu einer Wiederholungssitzung gebeten. Es handelte sich dabei allerdings nicht um eine echte Zufallsstichprobe,

sondern um diejenigen, die erstens zum angegebenen Termin Zeit und zweitens Interesse dafür hatten. Die Verteilung auf die vier experimentellen Gruppen war nicht ganz ausgeglichen: aus der Gruppe „-“ (erste Angabe: Intelligenz, zweite Angabe: Transparenz) kamen 2, aus „- +“ 5, aus „+ -“ 4 und aus „+ +“ 3 Vpn.

Die Resultate dieser Test-Retest-Studie sprechen nicht für TRENDPO: die Rangkorrelation zwischen dem ersten und zweiten Trendwert beträgt lediglich 0.199, während sich für TRENDFU eine Test-Retest-Korrelation von 0.798 ergibt. Auf dem Hintergrund dieser – allerdings nur an einer kleinen Stichprobe gewonnenen – Werte überraschen ausbleibende Effekte in bezug auf das Gütemaß TRENDPO wenig. Je unreliabler ein Maß ist, der Meßfehler also sehr groß wird, umso schwächer lassen sich Effekte nachweisen, selbst wenn sie tatsächlich vorliegen. Wenn – wie im Fall des Intelligenztests – auch dort bereits Meßfehler vorliegen, kann nur durch eine zweiseitige (und in diesem Fall multiplikative) Attenuationskorrektur der wahre, „unbiased“ Zusammenhang geschätzt werden (vgl. Kessler & Greenberg 1981, Kap. 10).

4.2 Heuristische Datenanalyse

Unter diesem Stichwort berichten wir zwei Ergebnisse, zu denen zwar keine expliziten Hypothesen vorliegen, die aber von heuristischem Wert sein können: (1) Gewichtung der verschiedenen Eingriffsmöglichkeiten sowie (2) die Prädiktionsgüte des FKP für den Lösungserfolg.

Unterschiedliche Gewichtungen der insgesamt 10 Eingriffsbereiche darf man sicher erwarten. Abbildung 5 zeigt den auf die jeweilige Gesamtzahl an Eingriffen relativierten Anteil der 10 Bereiche für die vier Gruppen.

Daraus kann man entnehmen, daß den Größen Rohmaterial-Einkauf, Hemdenpreis, Werbung sowie Reparaturkosten bevorzugte Aufmerksamkeit zuteil wurde, während Veränderungen bei Lieferwagen und Geschäftslage selten vorgenommen wurden. Von insgesamt 5760 möglichen Eingriffen (= alle 48 Vpn intervenieren in allen 12 Monaten in allen 10 Bereichen) wurden 2471 (= 42.9%) genutzt. Getrennt nach Gruppen ergeben sich bei I- T- 36.11%, bei I- T+ 45.62%, bei I+ T- 40.07% und bei I+ T+ 49.79% genutzte Eingriffsmöglichkeiten.

Um nun zu einer genaueren Aussage darüber zu kommen, welche Beziehungen zwischen den beiden Faktoren und dem Eingriffsverhalten vorliegen, wurde eine multivariate 2x2-Varianzanalyse mit den 10 Eingriffshäufigkeiten als AVn berechnet (via SPSS-Prozedur REGRESSION; Nie et al. 1975). Wir unterstellen dabei ungeprüft multivariat normalverteilte Residuenvariablen sowie homogene Varianz-Kovarianzmatrizen der Kriteriumsvariablen für jede Kombination von Prädiktoren (vgl. Erdfelder 1983, dem ich nicht nur für seine diesbezügliche Beratung ganz herzlich danke). Da die Inspektion der Daten eher eine diffuse Nonzentralitätsstruktur vermuten läßt, wird als Prüfgröße das PILLAI-BARTLETT-V herangezogen, das sich nach Untersuchungen von Olson (1976) als robustes Maß herausgestellt hat (vgl. Bredenkamp 1980, S. 84).

Zwar scheiterte der F-Test der globalen Hypothese, daß zwischen den „sets“ aus Prädiktoren und Kriterien ein linearer Zusammenhang besteht, knapp am 10%-Niveau, was im Normalfall zum Abbruch der Analyse führt; dennoch soll infolge der heuristischen

Absicht die Prüfung von Haupteffekten und Interaktion vorgenommen werden. Dabei zeigt sich für den Faktor Transparenz ein markanter Haupteffekt ($V' = .439$, $F = 2.738$ bei $df\text{-Zähler} = 10$, $df\text{-Nenner} = 35$ und $\text{Alpha} = 0.05$ signifikant). Faktor A sowie die Wechselwirkung erzielen ein F kleiner Eins. Dieses Ergebnis deutet an, daß zwischen transparenter und intransparenter Bedingung bedeutsame Unterschiede in bezug auf die Eingriffshäufigkeit bestehen.

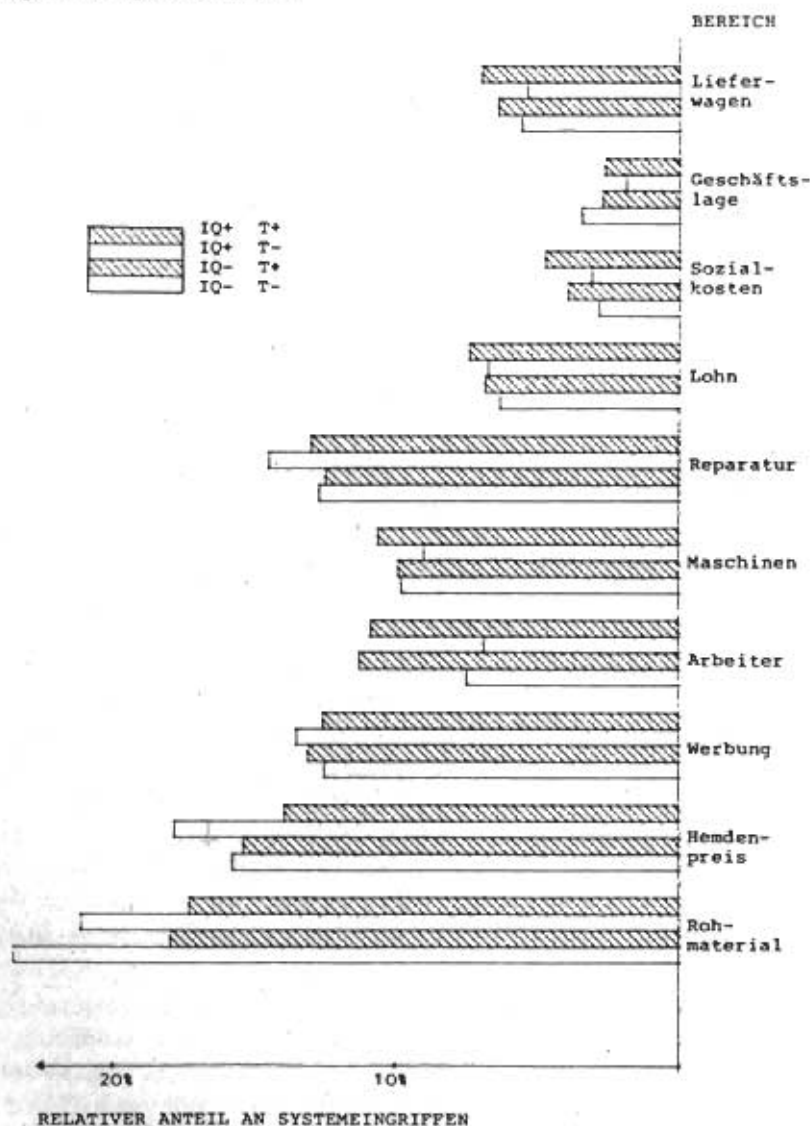


Abb. 5: Die Bedeutung der Eingriffsvariablen aus der Sicht der Probanden, geschätzt aus dem relativen Anteil an den gesamten Systemeingriffen (getrennt für vier Gruppen).

Zur Klärung der Frage, welche Eingriffsbereiche für diesen Effekt verantwortlich sind, dient die Diskriminanzfunktion einer Diskriminanzanalyse zwischen Transparenz- und Intransparenz-Gruppe. Danach schenken Vpn unter transparenten Bedingungen der Arbeiterzahl fast doppelt soviel Aufmerksamkeit wie unter Intransparenz: 6.3 Interventionen hier stehen 3.3 dort gegenüber. Die genannte Variable ist mit deutlichem Abstand die höchstladende auf der erhaltenen Diskriminanzfunktion.

Eine zweite Frage beschäftigt sich mit der Prädiktionskraft des FKP von Kreuzig (1981). Unsere Vermutung, die FKP-Werte seien nur dann gute Prädiktoren des Erfolgs (wenn überhaupt), sofern sie *posthoc* erhoben werden, verlangt die Ermittlung zweier Korrelationskoeffizienten, nämlich des Vorher- bzw. Nachher-Wertes mit dem Erfolgsmaß.

Vorab jedoch noch einige zusätzliche Informationen: die Test-Retest-Korrelation beträgt für die Skala KDD 0.867, für STA 0.849; die Korrelation zwischen beiden Skalen beträgt vorher 0.323, nachher 0.329 (Kreuzig gibt eine Korrelation von 0.475 bei $N = 42$ an).

Das Ergebnis der multiplen Regression von KDD1 und STA1 auf TRENDPO liefert ein multiples R von 0.088, für KDD2 und STA2 einen Wert von 0.159. Bezieht man sich auf die AV TRENDFU, ergeben sich multiple R 's von 0.179 (vorher) bzw. 0.271 (nachher; alle genannten Werte nicht signifikant). Diese Befunde stärken unsere Vermutung: bei beiden AVn liefern die nachher erhobenen FKP-Daten, bei denen der Proband seine gemachten Erfahrungen verwerten kann, höhere Korrelationen als die vorher ermittelten, bei TRENDFU fallen die multiplen Korrelationen – wegen der besseren Reliabilität nicht verwunderlich – zwar höher aus als bei TRENDPO, insgesamt jedoch wesentlich niedriger als der von Kreuzig (1981) berichtete signifikante Zusammenhang von 0.571 zwischen Summenscore und Gütemaß.

5. Diskussion der Ergebnisse

Die berichteten Ergebnisse stehen nicht in Einklang mit wiederholt vorgelegten Befunden. Vielmehr zeigt sich ein deutlicher und vor allem genereller Zusammenhang zwischen Testintelligenz und Problemlösegröße.⁵⁾ Zwei mögliche, voneinander unabhängige Ursachen dafür könnten sein: (1) durch die mangelnde Reliabilität der bisher verwendeten abhängigen Variable „Lösungsgüte“ ist kein Effekt aufzeigbar gewesen, (2) die für diese Arbeit vorgenommenen Veränderungen (z.B. hinsichtlich der Offenlegung bislang geheimgehaltener Variablen oder der Ausschaltung des VI-„Filters“) sind so gravierend, daß kein echter Vergleich möglich ist.

Über das erste Argument braucht man nicht viel Worte zu verlieren. Für zukünftige Untersuchungen mit diesem System liegt eine alternative Operationalisierung mit besseren Eigenschaften vor. Das zweite Argument ist insofern schlagkräftig, als bewußt Abweichungen von der Vorlage in Kauf genommen wurden. Daß durch VI-Einflüsse wesentliche Effektverschiebungen zustande kommen, konnte etwa Klüwe (1983)

⁵⁾ Dessen praktische Bedeutsamkeit ist jedoch mit einem Eta von 0.28 (vgl. Tabelle 4) nicht gerade „berauschend“, wie Jäger (1983, schriftliche Mitteilung) feststellt. Es gibt allerdings Grund zur Annahme, daß der Zusammenhang tatsächlich größer ist, wenn man etwa die unzulängliche Operationalisierung der Intelligenz via APM bedenkt.

in eigenen Studien feststellen. Daneben dürfte die Tatsache von Bedeutung sein, daß die Komponente der aktiven Informationsbeschaffung durch die Präsentation des Systems geringeren Einfluß besitzt. Akzeptiert man die von Putz-Osterloh & Lürer (1981, S. 331) vorgenommene Charakterisierung von Testintelligenz als „Fähigkeit zur Analyse von transparenten Informationen und dem erfolgreichen Einsetzen dieser Analyseergebnisse zur Lösung einer transparenten Problemstellung“, muß man zugestehen, daß unsere Problempräsentation sich auf die in Tests verlangten Fähigkeiten zubewegt (obwohl hinsichtlich Komplexität – und vor allem Lösbarkeit – beträchtliche Unterschiede bestehen). Ein Einwand gegen diese Argumentation liegt jedoch in der Tatsache, daß bei Erhöhung der Transparenz durch Vorgabe des Schaubilds Intelligente nicht den von Putz-Osterloh & Lürer (1981) prognostizierten Vorteil gegenüber Niedrigintelligenten erzielen. Die mit dieser These behauptete Interaktion zwischen den Faktoren Intelligenz und Transparenz wird nicht deutlich, wengleich aufgrund der geringen Teststärke diese Aussage weiterer Untersuchung bedarf.

Eine mögliche Erklärung für das Ausbleiben eines Haupteffekts Transparenz (vgl. Tabelle 4) könnte darin gesehen werden, daß die hier operationalisierte Stufung dieses Faktors keine Wirkung zeigt. Diesem Argument kann insofern nicht widersprochen werden, als z.B. die tatsächliche Nutzung des Schaubilds nicht erfaßt wurde. Dies sollte in zukünftigen Untersuchungen auf jeden Fall versucht werden. Ideal wäre hier der Einsatz von Blickerfassungsgeräten. Allerdings lassen die Ergebnisse in den Verhaltensmaßen (Eingriffspräferenzen) sehr wohl den Schluß zu, daß der Faktor „gegriffen“ hat. Die Ergebnisse der multivariaten Varianzanalyse weisen ein – wohl plausibles – häufigeres Eingreifen der Transparenz-Gruppe nach, was eine lebhaftere Auseinandersetzung, vielleicht auch ein stärkeres „Probierverhalten“ mit einzelnen Eingriffsbereichen vermuten läßt.

Daß in bezug auf das Gütekriterium kein Transparenzeffekt auftritt, wohl aber in bezug auf das Eingriffsverhalten, ist bei der erzielten hohen Teststärke und einem angenommenen „großen Effekt“ ein bedeutsamer Hinweis auf die Notwendigkeit der mehrschichtigen Analyse: offensichtlich sind Leistung und Verhalten in diesem System unabhängige Aspekte der Problemlösung. Die Tatsache, daß eine Gruppe mit der SWS anders umgeht als eine zweite, führt nicht zu Unterschieden in der Güte der Bearbeitung. Dieses Resultat ist vergleichbar der weitgehenden Unabhängigkeit der Dimensionen „Fehler“ und „Geschwindigkeit“ in Konzentrationstests. Es zeigt die Notwendigkeit, komplexen Phänomenen komplexe Konstrukte gegenüberzustellen.

Ein letzter Diskussionspunkt bezieht sich auf die generelle Forschungsstrategie. Die vorliegende Arbeit demonstriert erneut, daß auch im Bereich hochkomplexen Problemlösens sehr wohl hypothesenorientierte, experimentelle Untersuchungen unter kontrollierten Bedingungen möglich sind. Eine Reihe bisheriger Schwachpunkte (mangelnde Reliabilität – und damit natürlich mangelnde Validität – des Gütekriteriums, Versuchsleitereffekte, ausschließlich univariate Analysen, keine Kontrolle des Beta-Fehlers etc.; vgl. auch die Argumente von Jäger 1982) können überwunden werden. Problematische Aspekte beim Einsatz derart komplexer, computersimulierter Systeme („Zufalls“-kontrolle und -quantifizierung, Optimierbarkeit der Modellgleichungen, Systemgütebestimmung etc.) sind bekannt, einige Ansätze zur Klärung dieser Aspekte wurden genannt.

Was nach wie vor aussteht, ist – wie auch Krause (1982, S. 163) konstatiert – eine „Theorie kognitiver Prozesse“, auf die sich entsprechende Experimente beziehen müßten, sowie die Integration von multivariater Intelligenzstrukturforschung und experimenteller Denkprozeßforschung (Jäger 1982, S. 30). Modellvorstellungen dazu wurden zwar formuliert (vgl. Hussy 1983), jedoch ist das weitgehende Theoriedefizit und die Unverbundenheit der Forschungszweige noch nicht überwunden. Es nützt auf die Dauer wenig, in heuristischer Absicht enorm viele Daten zu erheben und sich davon Anstöße zur Theoriebildung zu erhoffen. Der Gefahr einer „reduktiven Hypothesenbildung“ – möglicherweise infolge einer „intellektuellen Notfallreaktion“ sensu Dörner (1981) – sollten wir uns nicht aussetzen.

Zusammenfassung

Im Rahmen hochkomplexer Problemlöseforschung wurde wiederholt berichtet über den fehlenden bzw. geringen Zusammenhang zwischen Testintelligenz und Güte der Bearbeitung eines computersimulierten Systems. Am Beispiel der „Schneiderwerkstatt“ werden Probleme dieser Arbeiten aufgezeigt. Mittels eines 2x2-Designs (APM-Testintelligenz: niedrig/hoch, Systemtransparenz: niedrig/hoch) wird der Nachweis geführt, daß allein Testintelligenz wesentliche Bedingung für Lösungsgüte darstellt. Ein Transparenzeffekt schlägt sich lediglich in erhöhtem Eingriffsverhalten nieder. Die Befunde verlangen ein Überdenken der bisherigen Untersuchungsergebnisse.

Summary

Recent research on complex problem solving has found little or no relationship between intelligence and problem solving as measured by the quality of handling a complex computersimulated system. This study shows some of the problems in research with the so-called "Tailor-Shop"-system. In a 2x2-design (APM intelligence: low/high, system's transparency: low/high) it can be shown that only intelligence has a main effect on the dependent variable "quality of problem solving". A transparency effect reveals only changes in system intervention, but without consequences on solving the problem. Therefore it is postulated to reconsider some of the recently published results in this area.

Literatur

- Berkowitz, M.S. & Szabo, M.: Individual and dyadic problem solving as a function of mental ability through computer simulation. *Journal of Educational Research* 1978, 71, 328–333.
- Bredenkamp, J.: *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff 1980.
- Cohen, J.: *Statistical power analysis for the behavioral sciences*. New York: Academic Press 1977.
- Dörner, D.: Kognitive Merkmale erfolgreicher und erfolgloser Problemlöser beim Umgang mit sehr komplexen Systemen. In: Ueckert, H., Rhenius, D. (Hrsg.): 1979a, 185–195.

- D ö r n e r, D.: Programm TAILORSHOP in der Version für TI-59 mit Drucker PC-100. Modifizierte und kommentierte Fassung von Norbert Streitz, Aachen 1979b. (unveröffentlicht).
- D ö r n e r, D.: Über die Schwierigkeiten des Umgangs mit Komplexität. Psychologische Rundschau 1981, 32, 163–179.
- D ö r n e r, D., K r e u z i g, H.W., R e i t h e r, F. & S t ä u d e l, T. (Hrsg.): L o h h a u s e n. Vom Umgang mit Unbestimmtheit und Komplexität. Bern: Huber 1983.
- E r d f e l d e r, E.: Skriptum zur Vorlesung „Multivariate Datenanalyse“. Trier: Fachbereich I – Psychologie – der Universität (vervielfältigtes Manuskript) 1983.
- F u n k e, J.: Mondlandung – ein neuer Aufgabentyp zur Erforschung komplexen Problemlösens. Trierer Psychologische Berichte 1981, 8, Heft 9.
- H u s s y, W.: Komplexe menschliche Informationsverarbeitung: das SPIV-Modell. Sprache & Kognition 1983, 2, 47–62.
- J ä g e r, A.O.: Intelligenzstrukturforschung. Konkurrierende Modelle, neue Entwicklungen, Perspektiven. In: J ä g e r, A.O. (Hrsg.): Berliner Beiträge zur Intelligenzforschung. Arbeitsbericht 1 aus dem Forschungsprojektschwerpunkt „Produktives Denken/Intelligentes Verhalten“ an der FU Berlin, 1983. 8–34.
- K e s s l e r, R.C. & G r e e n b e r g, D.F.: Linear panel analysis. Models of quantitative change. New York: Academic Press 1981.
- K l u w e, R.: Schriftliche Mitteilung 1983.
- K r a u s e, W.: Problemlösen – Stand und Perspektive. Teil II. Zeitschrift für Psychologie 1982, 190, 141–169.
- K r e u z i g, H.W.: Gütekriterien für die kognitiven Prozesse bei Entscheidungssituationen in sehr komplexen Realitätsbereichen und ihr Zusammenhang mit Persönlichkeitsmerkmalen. In: U e c k e r t, H., R h e n i u s, D. (Hrsg.): 1979, 196–209.
- K r e u z i g, H.W.: Über den Zugang zu komplexem Problemlösen mittels prozeßorientierter kognitiver Persönlichkeitsmerkmale. Zeitschrift für Experimentelle und Angewandte Psychologie 1981, 28, 294–308.
- L ü e r, G.: Mündliche Mitteilung 1983.
- N i e, N.H., H u l l, C.H., J e n k i n s, J.G., S t e i n b r e n n e r, K. & B e n t, D.H.: SPSS. Statistical package for the social sciences. New York: McGraw-Hill 1975.
- O l s o n, C.L.: On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin 1976, 83, 579–586.
- P e l z m a n, L.: Ergebnisse der experimentellen Wirtschaftsforschung zum Optimierungsverhalten. Psychologie und Praxis 1982, 26, 182–191.
- P o o l e, L. & B o r c h e r s, M.: 77 Basic-Programme. München: te-wi Verlag 1980.
- P u t z - O s t e r l o h, W.: Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. Zeitschrift für Psychologie 1981, 189, 79–100.

- Putz-Osterloh, W. & Lüer, G.: Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift für Experimentelle und Angewandte Psychologie* 1981, 28, 309–334.
- Raven, J.C.: *Advanced progressive matrices. Sets I and II. Plan and use of the scale with a report of experimental work.* London: Lewis 1965.
- Schick, K.: *Lineares Optimieren. Einführung in die mathematische Behandlung moderner Probleme in den Wirtschaftswissenschaften.* Frankfurt: Diesterweg 1981.
- Ueckert, H. & Rhenius, D. (Hrsg.): *Komplexe menschliche Informationsverarbeitung.* Bern: Huber 1979.
- Westermann, R. & Hager, W.: Entscheidung über statistische und wissenschaftliche Hypothesen: Zur Differenzierung und Systematisierung der Beziehungen. *Zeitschrift für Sozialpsychologie* 1982, 13, 13–21.

Anschrift des Verfassers:

Dipl.-Psych. Joachim Funke
Fachbereich I – Psychologie
der Universität
Schneidershof
5500 Trier