

Alles bestätigt? Anmerkungen zum Kommentar von Wiebke Putz-Osterloh (1983)

Joachim Funke

Als Antwort zur Replik von Putz-Osterloh (1983) auf die vorangegangene Arbeit von Funke (1983) werden zwei nach wie vor kritische Aspekte hervorgehoben: (a) es wird eine Präzisierung des Transparenzkonzepts vorgeschlagen, (b) die Reliabilitäts-/Validitätsproblematik wird unter Hinzuziehung einer kleinen Studie erneut aufgezeigt. Auch am Argument bisher nicht falsifizierbarer Theorien im Bereich des komplexen Problemlösens wird festgehalten.

As a response to the reply of Putz-Osterloh (1983) on Funke's (1983) article two critical aspects of previous work are demonstrated: (a) a more precise formulation of transparency is given, (b) the problem of reliability/validity is demonstrated in the context of a small study. The argument of non-falsifiable theories about complex problem solving is renewed.

Mein Beitrag über die Bedeutung der Testintelligenz beim komplexen Problemlösen (Funke 1983) führte zu einem Kommentar (Putz-Osterloh 1983a), zu dem ich ein paar kurze Anmerkungen machen möchte. Die Replik vermittelt den Eindruck, es handele sich bei dem von mir gefundenen Zusammenhang zwischen Intelligenz und komplexem Problemlösen um nichts Überraschendes, sondern „bestätigte“ frühere Befunde. Warum ich anderer Ansicht bin, soll an zwei wichtigen Aspekten dieser Studien veranschaulicht werden:

- (1) Die Definition von Transparenz bzw. Intransparenz scheint mir unpräzise und ermöglicht daher eine relativ großzügige Begriffsverwendung; ich werde eine präzisere Charakterisierung vorschlagen.
- (2) Die Bestimmung des Gütekriteriums und dessen Validierung durch Selbst- bzw. Fremdbeurteilungen der Vpn überzeugt nicht. Neben testtheoretischen Überlegungen zeige ich an einer kleinen Untersuchung die Bedenklichkeit dieses Vorgehens.

Schließlich möchte ich dafür plädieren, angesichts der vielen und m.E. noch längst nicht gelösten Probleme theoretischer wie methodologischer Natur keine unnötigen „Scheingefechte“ zu führen, sondern mit konstruktiven Ideen zu versuchen, verschiedene Forschungszweige zu integrieren und zu einer Systematisierung vorliegender Befunde zu kommen, wie sie von Dörner (1983) und Jäger (1983) gefordert bzw. versucht wird.

Zur Frage der *Transparenz/Intransparenz* einer Problemstellung schreibt Putz-Osterloh (1981, S. 84): „... müssen die Pbn unter der intransparenten Bedingung den Zustand der Systemvariablen ... erfragen. Dabei können sie allerdings aus dem Katalog der möglichen Maßnahmen alle Variablen, über deren Zustand sie Informationen benötigen, ableiten.“

Wesentliches Merkmal ist hier die aktive Informationsbeschaffung. Ähnlich Putz-Osterloh & Lüer (1981, S. 312): „Als intransparent gelten ... Probleme, deren Variablen erst durch aktives Eingreifen des Problemlösers für ihn erkennbar und analysierbar sind, da ihm zu Beginn nicht genügend Informationen über die Variablen vorliegen“. In der Kurzform: „Ein Problem ist für einen Problemlöser dann intransparent, wenn ihm nicht alle relevanten Daten vorgegeben sind.“ (Putz-Osterloh 1983b, S. 101). Diese Formulierungen machen die Notwendigkeit einer Präzisierung deutlich, denn ab welchem Zeitpunkt will man etwa Transparenz in Bezug auf die für die Schneiderwerkstatt wichtige Relation zwischen Werbung und Nachfrage gelten lassen:

- (a) wenn die Vp global erfährt, Werbung steigere die Nachfrage,
- (b) wenn die Vp die genaue Relation erfährt, wieviel Nachfrage durch Erhöhung der Werbungskosten um eine Einheit zusätzlich erzeugt wird,
- (c) wenn die Vp darüberhinaus den Grenzwert erfährt, von dem ab die Relation nicht mehr gilt?

Dieses Beispiel macht deutlich, daß es offenkundig ein Kontinuum der Transparenz und damit des zur Problemlösung notwendigen Wissens gibt. Dieses Kontinuum kann nun nach mehreren Ebenen aufgebrochen werden. Transparenz eines Systems kann – zu einem gegebenen Zeitpunkt – hinsichtlich folgender Aspekte bestehen: (a) Zielkriterium bzw. -kriterien, (b) Variablenanzahl, (c) Variablenzustand und (d) Variablenverknüpfung. Intransparenz bezüglich eines oder mehrerer dieser Aspekte kann zudem temporärer oder permanenter Natur sein: im ersten Fall besteht die Möglichkeit, während des Experiments die notwendigen Informationen einzuholen, im zweiten Fall werden bestimmte Systemdetails dem Probanden gegenüber auf Dauer geheimgehalten. Neben dem objektiven Grad an Transparenz (wie weit also der VI seine Karten aufdeckt) gilt es, auch den subjektiven Grad an Transparenz (also die Nutzung der objektiv verfügbaren Daten) zu berücksichtigen. Diesen sollte man der Präzision wegen besser als „Grad des Systemwissens“ bezeichnen und die Bezeichnung „Transparenz“ für objektive Charakteristika der Problemsituation reservieren.

Daß die so bestimmte Transparenz nicht ausschließlich durch eine Vp-VI-Interaktion vermittelbar ist, wie sie von Putz-Osterloh unter Hinweis auf einen „Bewegungsfreiraum für die Vp“ gefordert wird, dürfte klar sein: auch ohne sozialpsychologische Störgrößen kann sich die Vp interaktiv mit einem Rechner Informationen erfragen und sich durch ein Datennetz „hangeln“, ähnlich der Art, wie sie sich vom VI Auskünfte erbittet. Studien über Informationssammlung bei komplexen Problemen haben übrigens schon Glaser, Damrin & Gardner (1954) mit ihrer „tab item“-Methode durchgeführt (siehe auch Rimoldi 1960, McGuire & Babbott 1967, Karlins et al. 1967). Auch in jüngster Zeit gibt es Arbeiten, in denen die Diagnose von Fehlern in komplexen Systemen wie Schiffen, Flugzeugen oder Atomkraftwerken untersucht wird (Rasmussen & Rouse 1981). Die Kommunikation mit einem VI ist dort in der Regel ersetzt durch die (realistischere) Kommunikation mit dem System selbst. Dies scheint mir den gleichen Bewegungs-

freiraum für die Vp zu gewähren, ohne daß VI-Effekte massiv wirksam werden könnten wie im bisherigen Vorgehen, wo gleichzeitig dem VI großer Spielraum für sein Verhalten gelassen wird (sog. Effekt des „laxen“ Vorgehens; nach Bredenkamp 1980, S. 41).

Tabelle 1: Vergleich der Studien von Putz-Osterloh (1981), Putz-Osterloh & Lüer (1981) und Funke (1983) hinsichtlich der Transparenzmerkmale unter zwei verschiedenen experimentellen Bedingungen; zur Ergänzung die Studien von Putz-Osterloh (1983b) und Dörner et al. (1983)

Transparenz hinsichtlich	„Transparenz“			„Intransparenz“			PO 1983b	DÖ et al. 1983
	PO 1981	PO&L 1981	FU 1983	PI 1981	PO&L 1981	FU 1983		
a) Zielkriterium	+ ¹	+	∅	+	+	∅	+	--
b) Variablenanzahl	+	+	++	∅ ²	∅ ²	++	∅ ²	-
c) Variablenzustand	+	+	++	-	-	++	-	- ³
d) Variablenverknüpfung	+	+	+	--	--	--	--	--

¹ - -: sehr niedrig, -: niedrig, ∅: mittel, + hoch, ++: sehr hoch

² es werden alle Eingriffsmöglichkeiten genannt, Katalog möglicher Maßnahmen liegt vor

³ automatische Mitteilungen über Zustand der „Alarmmatrix“ (Dörner et al. 1983, S. 130)

Vergleicht man einmal einige Studien miteinander, in denen die „Schneiderwerkstatt“ eingesetzt wurde, ergibt sich hinsichtlich der vier genannten Ebenen von Transparenz folgendes Bild. Direkt vergleichbar sind die Studien von Putz-Osterloh (1981), Putz-Osterloh & Lüer (1981) und Funke (1983), da dort jeweils Transparenz- und Intransparenzbedingungen realisiert wurden. Bei Putz-Osterloh (1983b) wie bei Dörner et al. (1983) lagen jeweils „intransparente“ Bedingungen vor. – Die in Tabelle 1 eingetragenen Ausprägungsgrade beruhen auf Einschätzungen, da in den genannten Publikationen nicht alle Details aufgeführt sind. Deutlich sichtbar wird der Unterschied zwischen Funke (1983) und den übrigen Arbeiten auf der Ebene „Intransparenz“: die sehr hohe Transparenz bezüglich des Variablenzustands bei Funke durch das monatliche Display der Spielresultate zeigt die Richtigkeit des Arguments von Putz-Osterloh (1983b), soweit es diesen Aspekt der Transparenz betrifft. Bezüglich der Intransparenz der Variablenverknüpfung jedoch zeigen sich keine Unterschiede bei den fünf genannten Studien. Funke's Arbeit prüft somit den reinen Effekt der Verknüpfungskennntnis, die beiden anderen Studien dagegen konfundieren die Aspekte (c) und (d).

Nun zum Problem der *Validität der Gütekriterien*. Putz-Osterloh (1983a) schreibt: „Es schwebt ihm vor, die Güte des Verhaltens einer Vp durch Vergleich mit einem Optimalverhalten zu messen.“ Dies trifft korrekt meine Zielvorstellungen (vgl. Funke 1981) und es gibt Studien, in denen dies erfolgreich gelungen ist (z.B. die ausgezeichnete, wenn auch für Psychologen teilweise schwer zu lesende Arbeit von Thalmair 1979). Dies setzt jedoch ein tieferes Eindringen in die systemtheoretische Materie voraus, als es z.B. bei der von Putz-Osterloh (1983a) als Beleg angeführten Stelle bei Dörner et al. (1983, S. 44–47) anzutreffen ist. Blickt man auf das umfangreiche Rüstzeug, das von der Systemtheorie unter den Stichworten „complexity“, „connectivity“ und „catastrophe“ zur Beschreibung und Analyse hochkomplexer Systeme angeboten wird (zur Einführung: Casti 1979 sowie die ersten drei Kapitel aus Postle 1980 über die höchst interessante mathematische Katastrophentheorie), wundert man sich über deren geringe Rezeption bei psychologischen Systemforschern. Dies allein mag erklären, warum ein m. E. für die Interpretation des Geschehens so wichtiger Tatbestand wie der der Nichtoptimierbarkeit eines Systems zwar seit 1973 bekannt, aber bis zu meiner Kritik nicht mitgeteilt wurde. Es stellt sich ja schließlich die Frage, ob man mit den ausgewählten unlösbaren Problemstellungen nicht Experimente zur Frustrationstoleranz macht, ohne sie als solche zu kennzeichnen! Vielleicht wäre es geschickter, zunächst mit komplexen Problemtypen zu arbeiten, deren Lösbarkeit garantiert ist. Oftmals reicht schon die Änderung einer Gleichung aus, um aus einem vordem nichtoptimierbaren ein optimierbares System zu machen. Die für „Tanaland“ behauptete Nichtoptimierbarkeit kann daher nicht auf andere, ähnliche Systeme generalisiert werden.

Ohne auf die Frage der mir von Putz-Osterloh (1983a) unterstellten „Falschdarstellung“ bzw. „Unterschlagung“ von Information näher eingehen zu wollen (dies kann der Leser selbst beurteilen), soll dennoch unsere Kritik an der Vorgehensweise zur Validierung des Gütekriteriums, die bei „Lohhausen“ und „Schneiderwerkstatt“ ähnlich ausfiel, noch einmal erläutert werden. Es geht also darum, ob man aus den hohen Übereinstimmungskoeffizienten verschiedener Beurteiler und deren Korrelation mit einem „objektiven“ Indikator auf die Validität dieses Indikators schließen darf. Zunächst die Resultate einer kleinen Studie, in der mit einem künstlichen Datensatz von 48 „Vpn“ operiert wurde. Jeder Datensatz bestand aus einer Sequenz von 10 geglätteten Zufallszahlen, die so etwas wie den Erfolg der Vp beschreiben sollte. Wir können nunmehr fragen, ob Rater die hier per Zufall erzeugten „Kurven“ treffsicher beurteilten und wie die Korrelation zu einem „objektiven“ Gütekriterium (nehmen wir dafür die Zahl der Einheiten mit Aufwärtstrend) ausfällt. Aufgrund des erzeugenden Algorithmus wissen wir, daß der Trend zweier aufeinanderfolgender Meßwerte absolut unabhängig voneinander ist, die Reliabilität eines summarischen Trendmaßes also Null beträgt. Vier „Experten“ (Spieler der „Schneiderwerkstatt“) und zwei „Naive“ beurteilen die „Kapital“kurven hinsichtlich ihrer Güte auf einer fünfstufigen Skala. Die Auswertung der erhaltenen Daten erfolgte mittels des Übereinstimmungskoeffizienten r_{cc} von Horst (1949; zitiert nach Sixtl 1982², S. 149f.) ganz analog zu der bei Putz-Osterloh (1981) beschriebenen Prozedur. Das r_{cc} der sechs Beurteiler beträgt 0.941, die Korrelation der durchschnittlichen Beurtei-

lung (schulnotenskaliert) zum Trendmaß beträgt -0.685 , d.h. niedrige gute Noten indizieren einen guten hohen Trendwert. Darf man aus diesen Daten folgern, einen guten Indikator für das Systemverhalten gefunden zu haben? Was bringt das vollkommen „zufällige“ Schwanken der Rohdaten zum Ausdruck? Ich bin der Meinung, daß als einziger Schluß aus diesen Daten folgt, daß Rater den Trendindikator reliabel schätzen, keineswegs aber, daß es sich hierbei um einen validen Indikator des Systemverhaltens handelt! Aus der hohen Interraterkorrelation und der Korrelation mit dem „Außen“-Kriterium folgt nicht die geringste Aussage über die Gültigkeit des gewählten Kriteriums für eine angemessene Systembeschreibung. Eine Validitätsprüfung kann nur über andere Wege erfolgen (vgl. Hand & Sims 1975, Ruben & Lederman 1982), sollte dabei auch die „Fallen“ der klassischen Testtheorie im Auge haben (z.B. bei Verletzungen der Annahme unkorrelierter Fehler; vgl. Zimmerman & Williams 1977). Sternberg's (1982, S. 229) Postulate der Quantifizierbarkeit, Reliabilität, Konstruktvalidität und empirischen Validität bei Intelligenztestaufgaben sollten auch für die Meßinstrumente im Bereich komplexen Problemlösens gelten. Die „augenscheinliche“ Validität mancher Güteindikatoren muß gründlicher belegt sein, bevor weitreichende Schlußfolgerungen daraus gezogen werden. Vor allem sollte man nicht über Validität reden, solange „das Problem der Reliabilitätsüberprüfung des Gütekriteriums noch nicht eindeutig geklärt“ (Putz-Osterloh 1983a) ist. Wenn Vpn von Mal zu Mal andere Strategien verwenden (wie Putz-Osterloh vermutet) und damit andere Gütewerte erzielen, scheint das Konzept eines traits „guter Problemlöser“ doch äußerst problematisch. Wir haben selbst die Situationsspezifität und die jeweils idiosynkratischen Erfahrungen der Vpn im Umgang mit dem System betont; diese sollten bei der Gütebestimmung auch in Rechnung gestellt werden (also Einzelfallbetrachtungen oder Aufnahme eines Situationsparameters in die Erfolgsbestimmung).

Zum Schluß ein Wort zur kritisierten „Theoriearmut“: Putz-Osterloh weist auf Dörner's Theorie, aus der sich „sehr wohl Hypothesen für empirische Untersuchungen ableiten und diese auch *bestätigen* lassen“ (1983a, Hervorhebung von mir). Den heuristischen Wert wollen wir nicht bestreiten, doch läßt die Präzision der Prognosen noch insofern viel zu wünschen übrig, als ihre *Falsifikation* vielfach nicht möglich ist. Hinzu kommt die Tendenz, nichtsignifikante Befunde erst gar nicht oder nur unvollständig mitzuteilen, selbst wenn sie sich auf explizit formulierte Hypothesen beziehen. Herrmann, der sich über den starken Einfluß nur schlecht überprüfbarer Theorien in der Kognitionspsychologie wundert (1982), warnt davor: „Solange man lediglich ‚Fakten‘ sammelt, kann man für eine Weile meinen, irgendwann springe aus den Fakten schon von ganz allein die Wahrheit, der Zugang zur wahren Wirklichkeit, heraus.“ (Herrmann 1983, S. 98). Über die gelegentliche „Nützlichkeit der Fiktionen“ wollen jedoch auch wir nicht hinwegsehen, die „Wende“ allerdings ist von dort nicht zu erwarten.

Literatur

- Bredenkamp, J.: Theorie und Planung psychologischer Experimente. Darmstadt: Steinkopff 1980.
- Casti, J.: Connectivity, complexity, and catastrophe in large-scale systems. New York: Wiley 1979.
- Dörner, D.: Denken, Problemlösen und Intelligenz. In: Lüer, G. (Hrsg.). Band 1. 354–367.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (Hrsg.): Lohhausen. Vom Umfang mit Unbestimmtheit und Komplexität. Bern: Huber 1983.
- Funke, J.: Mondlandung – ein neuer Aufgabentyp zur Erforschung komplexen Problemlösens. Trierer Psychologische Berichte 1981, 8, Heft 9.
- Funke, J.: Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? Diagnostica 1983, 29, 282–302.
- Glaser, R., Damrin, D. E. & Gardner, R. M.: The tab item: A technique for the measurement of proficiency in diagnostic problem-solving tasks. Educational and Psychological Measurement 1954, 14, 283–294.
- Hand, H. & Sims, H.: Statistical evaluation of complex gaming performance. Management Science 1975, 21, 708–717.
- Herrmann, T.: Über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. Sprache & Kognition 1982, 1, 3–14.
- Herrmann, T.: Nützliche Fiktionen. Anmerkungen zur Funktion kognitionspsychologischer Theoriebildung. Sprache & Kognition 1983, 2, 88–99.
- Jäger, A. O.: Intelligenzstrukturforschung: konkurrierende Modelle, neue Entwicklungen, Perspektiven. In: Lüer, G. (Hrsg.). Band 1. 339–353.
- Karlins, M., Coffman, T., Lamm, H. & Schroder, H.: The effects of conceptual complexity in information search in a complex problem solving task. Psychonomic Science 1967, 7, 137–138.
- Lüer, G. (Hrsg.): Bericht über den 33. Kongreß der Deutschen Gesellschaft für Psychologie in Mainz 1982. Göttingen: Hogrefe 1983.
- McGuire, C. H. & Babbott, D.: Simulation technique in the measurement of problem solving skills. Journal of Educational Measurement 1967, 4, 1–10.
- Postle, D.: Catastrophe theory. Glasgow: Fontana 1980.
- Putz-Osterloh, W.: Über die Beziehung zwischen Intelligenz und Problemlöseerfolg. Zeitschrift für Psychologie 1981, 189, 79–100.

- Putz-Osterloh, W.: Kommentar zu dem Aufsatz von J. Funke: Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica* 1983a, **29**, 303–309.
- Putz-Osterloh, W.: Über Determinanten komplexer Problemlöseleistungen und Möglichkeiten zu ihrer Erfassung. *Sprache & Kognition* 1983b, **2**, 100–116.
- 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.
- Rasmussen, J. & Rouse, W. B. (Hrsg.): *Human detection and diagnosis of system failures*. New York: Plenum Press 1981.
- Rimoldi, H. J.: A technique for the study of problem-solving. *Educational and Psychological Measurement* 1960, **20**, 449–460.
- Ruben, B. D. & Lederman, L. C.: Instructional simulation gaming: validity, reliability, and utility. *Simulation & Games* 1982, **13**, 233–244.
- Sixtl, F.: *Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme*. Weinheim: Beltz 1982².
- Sternberg, R. J.: Reasoning, problem solving, and intelligence. In: Sternberg, R. J. (Hrsg.): *Handbook of human intelligence*. Cambridge: Cambridge University Press 1982. 225–307.
- Thalmaier, A.: Zur kognitiven Bewältigung der optimalen Steuerung eines dynamischen Systems. *Zeitschrift für Experimentelle und Angewandte Psychologie* 1979, **26**, 388–421.
- Zimmerman, D. W. & Williams, R. H.: The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology* 1977, **16**, 135–152.

Anschrift des Verfassers:

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

5500 Trier