

Vorbemerkung des Autors:

Ich gratuliere Ihnen zu Ihrer Zeitschrift, die für mich als Auslandsösterreicher eine ganz besondere Funktion erfüllt: den Kontakt mit der österreichischen Psychologie zu vertiefen. Dafür bin ich Ihnen zu Dank verpflichtet.

Jetzt entsteht aber bei mir auch das Bedürfnis, mich ein wenig „einzumischen“: der diskussionswürdige Aufsatz von R. Kisser zur „Kritik an der Kritik psychologischer Tests“ hat mich zu den folgenden Anmerkungen verleitet.

Kritische Anmerkungen zu R. Kisser's „Kritik an der Kritik psychologischer Tests“

Von W. Kempf

Kritik an der Testpsychologie zu üben, ist standespolitisch nicht ungefährlich. Denn die Testpsychologie trägt wesentlich mit zum Ansehen der Psychologie in der Öffentlichkeit bei. Es ist daher zu begrüßen, daß *Rupert Kisser* gegen eine pauschale Verurteilung der Testpsychologie ins Feld zieht. Es ist dies umso mehr zu begrüßen, als sich *Kisser* offensichtlich um eine ausgewogene und dem heutigen Stand der Testtheorie entsprechende Argumentation bemüht.

Gleichzeitig werden an *Kisser's* Argumentation jedoch einige Voraussetzungen deutlich, die – als selbstverständlich hingenommen – die Gefahr heraufbeschwören, daß Leistungsfeststellungen mit Hilfe von Tests, von denen R. Kisser sagt, daß sie „eine durchaus scharfe Waffe“ gegen soziale Ungerechtigkeit sein können, sich am Ende als Bumerang erweisen.

So schreibt *Kisser* u.a.: „Es steht . . . außer Zweifel, daß jede entwickelte Gesellschaft Entscheidungen, wie z.B. über Zulassung und Nichtzulassung zu einem bestimmten Bildungsweg, über Anstellung oder Nichtanstellung eines Stellenbewerbers, über Entzug bzw. Nichtentzug einer Fahrerlaubnis usw. treffen muß.“ Ich stimme da mit *Kisser* voll überein: es sind erst die Bedingungen einer entfremdeten Gesellschaft, unter denen wir über Menschen verfügen müssen, die wir gar nicht kennen, und aus welcher das Bedürfnis nach einer Testpsychologie erwächst.

Es stellt sich aber die Frage, ob wir nicht am Ende der weiteren Entfremdung in unserer Gesellschaft Vorschub leisten, wenn wir diesem Bedürfnis nachkommen, bzw. den Eindruck erwecken, daß solche Entscheidungen mit Hilfe der Testpsychologie in qualifizierter Weise getroffen werden können?

Ich möchte damit nicht die psychologische Diagnostik an und für sich in Frage stellen, wohl jedoch gewisse Tendenzen innerhalb der Diagnostik, die darauf hinauslaufen, daß den vorgeblich „objektiven“ und „wissenschaftlich fundierten“ Tests immer mehr Gewicht beigemessen wird. Denn, wie ich mir als Testtheoretiker zu beurteilen zutraue, ist die theoretische Fundierung unserer Tests keineswegs so tragfähig wie das wissenschaftliche Ansehen, zu dem die Testtheorie der Testpsychologie verhilft:

1. Wenn *Kisser* schreibt, daß es „unser aller Erfahrung“ entspricht, „daß verschiedene Menschen über ‚mehr‘ oder ‚weniger‘ Klugheit, Ängstlichkeit oder Fingerfertigkeit verfügen, und daß „dieses ‚mehr‘ oder ‚weniger‘ prinzipiell auch einer Quantifizierung zugänglich“ ist, so ist dies sicherlich zutreffend. Aber es wird dabei die Frage übersehen, ob eine solche Quantifizierung den Aufgaben der psychologischen Diagnostik angemessen ist.

Tatsächlich hatte noch *William Stern* z.B. Intelligenz als die Fähigkeit definiert, sich unter zweckmäßiger Verfügung über Denkmittel auf neue Forderungen einzustellen. Und ersichtlich ist unter einem solchen Verständnis von „Intelligenz“ nicht in erster Linie nach einer Quantifizierung der Intelligenz gefragt, sondern nach einer qualitativen Bestimmung eben der Denkmittel, über die ein Proband verfügt. Es läßt sich auch psychologiegeschichtlich zurückverfolgen (vgl. *Grünwald* 1978), wie der ursprüngliche qualitative Ansatz von *Binet & Simon* erst durch die Abkopplung bestimmter Methoden von ihrem theoretischen Hintergrund in die heute übliche Praxis der Intelligenzmessung gemündet hat.

In Untersuchungen zur Intelligenzentwicklung bei Kindern ist man inzwischen von einer quantifizierenden Intelligenzdiagnostik wieder abgegangen (vgl. *Klahr* 1978, *Scandura* 1973). Nur die Testpsychologie hinkt da noch nach.

2. Wenn *Kisser* schreibt, die klassische Testtheorie stelle „ein in sich relativ geschlossenes System bereits vorwissenschaftlich plausibler Annahmen dar“, so hat er auch damit nicht ganz unrecht. Er übersieht dabei aber, daß diese Plausibilität zwar vorwissenschaftlich, aber doch durch eine ganz bestimmte – an den Naturwissenschaften orientierte – Denkungsweise bestimmt ist. Wie sich z.B. zeigen läßt, hat der von der klassischen Testtheorie definierte „True-Score“ nur wenig mit dem zu tun, was wir als Alltagsmenschen (und auch als psychologische Diagnostiker) meinen, wenn wir von einer „wahren Leistungsfähigkeit“ sprechen (vgl. *Kempf* 1982). Insbesondere ist die (durch die True-Score-Definition der klassischen Testtheorie implizierte) Annahme, daß der Erwartungswert der Meßfehler gleich Null sei, nicht psychologisch begründbar, sondern entspricht nur unserem Wunsdenken: von einem guten Test würden wir fordern, daß er diese Eigenschaft haben soll.

Wie aber wollen wir einen Test konstruieren, der diese Eigenschaft hat, wenn wir sie schon vor Konstruktion des Tests per Annahme voraussetzen?

3. Wenn *Kisser* schreibt, daß neuere testtheoretische Ansätze „dem wesentlichen Mangel der klassischen Testtheorie“ begegnen, so übersieht er, daß es sich dabei eben nur um jenen Mangel handelt, der von diesen neueren Ansätzen als wesentlich angesehen wird. Tatsächlich geht es doch in der klassischen Testtheorie und in der sogenannten probabilistischen Testtheorie um ganz verschiedene Fragestellungen. Während sich die klassische Testtheorie vor allem um die Reliabilität und Validität von Tests bemüht, geht es der probabilistischen Testtheorie in erster Linie um die Homogenität von Tests. In der klassischen Testtheorie werden die Worte „testen“ und „messen“ einfach synonym verwendet. In der probabilistischen Testtheorie geht es darum, Tests so zu konstruieren, daß wir Testergebnisse zu Recht als „Messungen“ bezeichnen können. Ob sich der Aufwand praktisch lohnt, ist eine Frage, die wir bedenken sollten.

4. *Kisser* bezeichnet Tests als „standardisierte Verfahren zur Entnahme von Verhaltensstichproben“. Als solche haben sie zweifellos ihren Wert. Wenn wir schon Urteile über Menschen fällen müssen, die wir nicht (oder kaum) kennen, dann ist es nur angemessen, unser Urteil wenigstens auf eine – nicht beliebige – Stichprobe ihres Verhaltens gründen zu wollen. Die Frage ist nur, ob es tatsächlich die Standardisierung der Verhaltensstichproben ist, welche sie aus der Beliebigkeit herauszuheben vermag. Wenn wir z.B. herausfinden wollen, über welche Denkmittel jemand verfügt, so scheint mir doch das gezielte, hypothesengeleitete Stellen von Aufgaben (die gegebenenfalls ad hoc zu erfinden sind, wie dies manche Diagnostiker ja auch tatsächlich tun) weit angemessener, als wenn jedem Probanden dieselben „Items“ vorgelegt werden. Bedauerlicherweise hat hier gerade die Testtheorie einen hemmenden Effekt: denn ihr ist es zu verdanken, daß standardisierten und damit pauschalisierten Aufgabensammlungen, wie sie unsere Tests darstellen, ein vergleichs-

weise hohes wissenschaftliches Ansehen beigemessen und mehr Vertrauen geschenkt wird. Im Extremfall kann dies bis zur Entmündigung des psychologischen Diagnostikers, bis hin zu automatisierter und computergesteuerter Testvorgabe, -auswertung und -interpretation führen.

5. Gleichzeitig wird von den verbreiteten testtheoretischen Ansätzen auch „Druck“ in der Richtung ausgeübt, Testergebnisse in bestimmten Kennzahlen (Testscores oder Personenparametern) zusammenzufassen und eben nicht danach zu fragen, welche Aufgaben ein Proband wie beantwortet und wie wir das Zustandekommen dieser Antworten verstehen können.

Im Testmodell von *Rasch* wird die Bildung von Summenscores statistisch gerechtfertigt. Doch haben wir dabei nicht den falschen Weg eingeschlagen? Passen wir damit nicht nur die Itemauswahl und Testkonstruktion an eine tradierte Praxis der Scorebildung an?

6. Wie problematisch diese Vorgehensweise in Verbindung mit der Standardisierung der Aufgabenauswahl ist, machen auch *Kissers* Ausführungen zur Validitätsproblematik deutlich. Wenn *Kisser* schreibt, daß die diagnostische Entscheidung am Ende „inhaltlich nach den Anforderungen, die an eine geeignete Person gestellt werden“ getroffen werden muß, so drängt sich mir die Frage auf, warum man dann nicht schon die Auswahl der Testaufgaben und die Auswertung der Testergebnisse an diesen Anforderungen orientiert, warum man denn überhaupt standardisierte Verhaltenstichproben gewinnen und zu Testscores verrechnen will?

Und noch etwas fällt dabei auf: daß die Begründung der diagnostischen Entscheidungen dort beginnt, wo die theoretische Fundierung der psychologischen Tests aufhört. Dies birgt die Gefahr in sich, daß diagnostische Entscheidungen mit einer wissenschaftlichen Autorität versehen werden, die ihnen überhaupt nicht zusteht. Und dies ist es auch, worin ich eine der beiden größten Gefahren der Testpsychologie sehe.

7. Eine nicht mindere Gefahr sehe ich darin, daß durch Tests „gemessene“ psychologische Eigenschaften vorgeblich als wertfrei konzipiert werden. Denn dies bedeutet nichts anderes, als daß die Werturteile, welche sowohl in die Konstruktion bzw. Auswahl von Testaufgaben als auch in die Beurteilung der Antworten der Probanden eingehen, schlichtweg übersehen werden.

Z.B. enthält der Intelligenz-Struktur-Test von *Amthauer* (1970) eine ganze Reihe von Testaufgaben, für die sich mit ein wenig Kreativität – oder, wenn sie auf einem anderen soziokulturellen Hintergrund als dem des Testkonstruktors bearbeitet werden – Antworten anbieten, die laut Auswertungsvorschrift als „falsch“ gewertet werden, obwohl sie bei genauerem Hinsehen sehr wohl erkennen lassen, daß und wie sich der Proband unter zweckmäßiger Verfügung über Denkmittel auf die (von ihm wahrgenommenen) Anforderungen des Tests einzustellen vermag. Dies ist es auch, was die Kritiker psychologischer Tests u.a. meinen, wenn sie davon reden, daß mit der „Allgemeinverständlichkeit menschlicher Eigenschaften . . . meist nur Wertvorstellungen der privilegierten Bevölkerungsschichten gemeint seien“.

Eine pauschale Verurteilung der Testpsychologie hat wenig Sinn, aber Fragen wie die oben aufgeworfenen sollten wir uns schon ernsthaft stellen. Wichtiger als ein Schlagabtausch zwischen Befürwortern und Kritikern der Testpsychologie ist, daß wir endlich auch jene Aspekte der Testpsychologie in ihre theoretische Fundierung einbeziehen, die mit den Mitteln von Mathematik und Statistik nicht zu beantworten sind. Solange wir dies nicht

tun, werden wir uns bezüglich der Wissenschaftlichkeit und Objektivität testpsychologischer Gutachten in gefährlichen Illusionen wiegen.

Literatur:

- Amthauer, R.*: Intelligenz-Struktur-Test (IST-70). Hogrefe: Göttingen, 1970.
- Grünwald, H.*: Die sozialen Ursprünge psychologischer Diagnostik. Phil. Diss.: Berlin, 1978.
- Kempf, W.*: Zur Grundlagenkrise der Testpsychologie. In: Aschenbach, G., Kempf, W. (Hrsg.): Psychologie zwischen Positivismus und Hermeneutik. Huber: Bern, 1982 (in Druck).
- Klahr, D.*: Information processing models of cognitive development. In: Scandura, J.M., Brainerd, Ch.J. (Hrsg.): Structural/Process Models of Complex Human Behavior. Sijthoff & Noordhoff: Alphen aan den Rijn, 1978.
- Scandura, J.M.*: Structural Learning. I. Theory and Research. Gordon & Breach: New York, 1973.

Autor:

Prof. Dr. Wilhelm Kempf, Universität Konstanz, D-7750 Konstanz, Postfach 5560.

Nachwort zu Kempfs „Kritische Anmerkungen“

Von Rupert Kisser

Es war für mich sehr erfreulich festzustellen, daß mein Artikel „Kritik an der Kritik psychologischer Tests“ Anstoß zu einer sachlichen und radikalen Diskussion, nicht aber zu einem weiteren „Schlagabtausch zwischen Befürwortern und Kritikern der Testpsychologie“ gegeben hat.

Der vorliegende Diskussionsbeitrag Kempfs zielt zunächst auf den wundesten Punkt psychologischer Diagnostik ab, nämlich auf das Fehlen eines modernen Validitätskonzeptes. Es wäre tatsächlich ein großes Mißverständnis, in den neueren Ansätzen der Testtheorie eine Lösung der Reliabilitäts- und Validitätsproblematik zu sehen. In diesem Zusammenhang wirft Kempf die Frage auf, ob die Quantifizierung menschlicher Eigenschaften den Aufgaben der psychologischen Diagnostik überhaupt angemessen sei. Inwieweit eine qualifizierende Diagnostik mit ad hoc oder ad personam erfundenen Testaufgaben den traditionellen quantifizierenden Verfahren überlegen ist, muß beim derzeitigen Stand der Forschung jedoch als völlig offen angesehen werden. Wie die Diskussion über adaptive Teststrategien in der USA zeigt, erschwert der Verzicht auf die standardisierte Situation jedenfalls die Einsehbarkeit diagnostischer Entscheidungen sowohl für Betroffene wie für Außenstehende erheblich, obwohl bei adaptivem Testen in der Regel bereits sorgfältig analysierte und nicht erst ad hoc erfundene Aufgaben Verwendung finden. Auch kann der verstärkte Einfluß der Individualität des Diagnostikers leicht zu einer Vergrößerung der Fehlervarianz der Urteile führen. Zwischen Mystifikation und Entmündigung des Diagnostikers ist aber genügend Platz für eine vertretbare Vorgangsweise.

Im Zentrum des Kempfschen Beitrages scheint mir aber die Gretchenfrage psychologischer Diagnostik zu stehen, nämlich die Frage, ob psychologische Diagnostik überhaupt eine Antwort auf die konkreten Fragestellungen unserer (entfremdeten) Gesellschaft versuchen soll oder nicht. Ohne diese Frage auf die leichte Schulter zu nehmen, versuche ich mir vorzustellen, wie die psychologische Diagnostik je bessere Beweise ihrer Leistungsfähigkeit erbringen soll, wenn sie es nicht versucht.