

Zur Bedeutung der Validität und der sozialen Akzeptanz in der Berufseignungsdiagnostik

The relevance of validity and applicants' acceptance of tests in personnel selection and training

Martin Kersting

In seinem erfrischend geschriebenen Beitrag widmet sich Marcus (2003) dem „*Wunder der sozialen Erwünschtheit in der Personalauswahl*“ und kritisiert zu Recht den unreflektierten Umgang mit diesem Thema. Gegen sozial erwünschtes Antwortverhalten getroffene Maßnahmen erbringen hinsichtlich der Kriteriumsvalidität keinen Vorteil und sind möglicherweise sogar schädlich oder unfair. Für Marcus (2003, S. 130) steht deshalb fest, dass es in Bewerbungssituationen „*zwar das Phänomen veränderten Antwortverhaltens gibt, daraus aber keine negativen Konsequenzen erwachsen (...)*“. Gemeint sind Konsequenzen für die Berufseignungsdiagnostik. Hier stellt sich die Frage, welche Konsequenzen Marcus in seinem Beitrag bedacht hat und welche er übersehen hat. Bedacht hat Marcus vor allem zweierlei: die Organisation, die eine Personalentscheidung trifft, einerseits und die Kriteriumsvalidität als handlungsleitenden Maßstab andererseits. Der Bewerber kommt nur am Rande vor, nämlich als eine von der Entscheidung betroffene Person. Damit ist die von Marcus skizzierte berufseignungsdiagnostische Welt beschrieben. In dieser kleinen Welt spielt das veränderte Antwortverhalten des Bewerbers vielleicht keine große Rolle. De facto ist die Welt der Berufseignungsdiagnostik aber wesentlich komplexer. Aus Platzgründen beschränkt sich der vorliegende Diskussionsbeitrag auf zwei der von Marcus übersehenen Aspekte: die soziale Akzeptanz und die Bedeutung der Validität jenseits der Korrelation zwischen Prädiktor und Kriterium. Auf einige andere Probleme der Position von Marcus hat Kanning (2003) bereits hingewiesen.

Wie wirkt der Einsatz verfälschbarer Verfahren auf die Bewerber?

Der Bewerber ist im Marcusschen „*Funktionsmechanismus Eignungsdiagnostik*“ ein passives Objekt, er ist von den Entscheidungen der Organisation betroffen, nimmt diese vielleicht als fair oder unfair wahr,

vor allem aber nimmt er die Entscheidung hin. Demgegenüber begreifen Konzepte der sozialen Akzeptanz (z. B. Herriot 1989; Schuler, 1990) Personalauswahl als einen interaktiven Prozess und eine Kommunikationssituation der Organisation mit den Bewerbern. Kersting (1998) fasst die Wirkung der sozialen Akzeptanz von Auswahlverfahren zusammen: (1) Bereits die Ankündigung bestimmter Verfahren kann potenzielle Bewerber im Vorfeld abschrecken. (2) Inakzeptable Verfahren können die Teilnehmer dazu veranlassen, Stellenangebote der Organisation auszuslagen. (3) Das Auswahlverfahren vermittelt ein Organisationsimage und wird als Ausdruck der organisationspolitischen Prinzipien und der Organisationskultur angesehen. Über diesen Mechanismus kann die Auswahlverfahren (4) die zukünftigen (zumindest anfänglichen) Einstellungen und Verhaltensweisen der ausgewählten Personen in der Ausbildung bzw. im Berufsleben ebenso beeinflussen wie (5) die allgemeine Einstellung sowohl der ausgewählten als auch der abgelehnten Bewerber gegenüber der Organisation. Marcus' Bewertung des Phänomens der sozialen Erwünschtheit berücksichtigt die hier genannten Wirkungen nicht. Er geht anscheinend von einer Welt aus, in der sich Organisationen aus einer unbegrenzten Anzahl an „Objekten“ das günstigste wählen und sich um ihr Image nicht zu sorgen brauchen. Die Selbstselektion der Bewerber, ihre Reaktionen auf den Einsatz verfälschbarer diagnostischer Verfahren und ihre Entscheidungen bleiben in dieser organisationszentrierten Sicht unberücksichtigt. Die von Marcus skizzierte eignungsdiagnostische Welt hat wenig mit der tatsächlichen aktuellen Situation (z. B. bei der Suche nach Spezialisten oder High-Potentials, bei der Besetzung interner Positionen) zu tun und sie hat noch weniger mit der Welt von morgen zu tun, die sich aufgrund des demographischen Wandels zu Gunsten der Bewerber ändern wird. Bislang wissen wir zu wenig darüber, wie sich die von den Bewerbern oder anderen relevanten Gruppen (z. B. Personalvertretung) wahrgenommene Möglichkeit zum „Faken“ auf die Akzeptanz von verfälschbaren Persön-

lichkeitstests in der Personalauswahl auswirkt. Im Sinne von Marcus davon auszugehen, es gäbe keine negativen Konsequenzen, ist weder plausibel noch empirisch fundiert.

Warum man beim Einsatz konstruktorientierter Verfahren nicht auf die Konstruktvalidität verzichten kann

In der gleichen atomistischen Weise, wie Marcus in der Welt der handelnden Personen nur die Organisation betrachtet, reduziert sich sein Blick bei der Validitätsfrage auf die Kriteriumsvalidität. Obwohl es sich bei Persönlichkeitstests um konstruktorientierte Verfahren handelt, erachtet Marcus die Frage der Konstruktvalidität für weitgehend irrelevant. Die Gewissheit, dass aus dem veränderten Antwortverhalten keine negativen Konsequenzen erwachsen, bezieht er allein auf die Kriteriumsvalidität. Diese stellt aber nur einen Validitätsaspekt dar. Auf die Konstruktvalidität wirkt es sich sehr wohl aus, wenn das Antwortverhalten nicht nur von dem interessierenden Merkmal, sondern auch von weiteren Einflussgrößen (z. B. der Absicht, sich positiv darzustellen) abhängt. Die Eigenschaftsmessung ist dann – wie Marcus (2003, S. 130) selbst schreibt – konfundiert. Der Autor fährt fort: „Wenn sich jetzt aber herausstellt, dass trotz dieser Konfundierung die Vorhersagegüte mindestens ebenso gut ist wie ohne oder mit geringerer Konfundierung, kann das nur bedeuten, dass die Beeinträchtigung der Konstruktvalidität auf der einen Seite ausgeglichen wird durch einen mindestens ebenso starken positiven Einfluss von 'faking' auf die Kriteriumsvalidität auf der anderen Seite.“ Hier trennt Marcus, was unverrückbar zusammengehört. Er nimmt an, Konstrukt- und Kriteriumsvalidität stünden sich auf verschiedenen Seiten in einem kompensatorischen Verhältnis gegenüber. Und er geht offensichtlich davon aus, dass in der Berufseignungsdiagnostik die Hegemonie der Kriteriumsvalidität gilt und die Konstruktvalidität vernachlässigbar ist. Beide Annahmen sind anzuzweifeln.

Entscheidend für die Diskussion um die Bedeutung der sozialen Erwünschtheit beim Einsatz von Persönlichkeitsfragebogen in der Eignungsdiagnostik ist die Frage der Validität, wobei es grundsätzlich nicht um die Validität eines Verfahrens, sondern um die Validität der Interpretation der Daten geht, die mit dem Verfahren gewonnen werden (Cronbach, 1971, p. 447). Die Reaktionen auf die Items eines Persönlichkeitstests werden als Indikator für ein (oder mehrere) Personmerkmal(e) interpretiert. Der Diagnosti-

ker schlussfolgert von dem Persönlichkeitstest auf das Konstrukt. Das Konstrukt selbst muss unabhängig sein von dem jeweils eingesetzten Messinstrument und der Situation der Messung. Die konstruktorientierte Interpretation der Daten ist aber durch das veränderte Antwortverhalten betroffen, was Marcus entgeht, da er nur die Kriteriumsvalidität betrachtet. Konstrukt- und Kriteriumsvalidierung stehen aber nicht, wie Marcus es suggeriert, in einem kompensatorischen Gegensatzverhältnis, sondern sind komplementäre Erkenntniszugänge zu ein und derselben Fragestellung. Es ist nicht möglich, einen Mangel an Konstruktvalidität durch ein Mehr an Kriteriumsvalidität auszugleichen, wie Marcus (2003, S. 130) es formuliert. Die Validität ist eine Einheit. „(...) [C]onstruct validity is indeed the unifying concept of validity that integrates criterion and content consideration into a common framework for testing rational hypothesis about theoretical relevant relationships“ (Messick, 1980, p. 1015). Marcus skizziert eine Situation, in der die Konstruktannahmen sich nicht bestätigen, ein Verfahren aber eine Kriteriumsvalidität aufweist. Ein Verfahren V indiziert also nicht das intendierte Konstrukt x, korreliert aber mit dem Kriterium K. Marcus nimmt diese Situation hin und erfreut sich an der Kriteriumsvalidität, die seiner Ansicht nach den Mangel an Konstruktvalidität kompensiert. Das Validitätsproblem besteht aber in diesem Fall darin, dass das Verfahren V etwas anderes indiziert als angedacht, also nicht das Konstrukt x, sondern ein anderes Konstrukt, z. B. y. Möglich und notwendig ist in diesem Fall die Suche nach einer anderen Theorie (einem anderen Konstrukt), das die gegebene Kovariation in Prädiktor und Kriterium erklärt. Auf der Grundlage gültiger Annahmen können wir das Verfahren V dann als Indikator für y interpretieren (und korrigieren unsere irriige Annahme, es sei ein Indikator für x). Nach dieser notwendigen Arbeit stimmen Konstrukt- und Kriteriumsvalidität überein.

Ohne gültige Konstruktannahmen ist die Kriteriumsvalidität wenig nützlich. Erst die Konstruktannahmen stellen einen Rahmen zur Integration verschiedener Befunde (zu einem Konstrukt, nicht etwa nur zu einem Verfahren) dar und sind somit das Fundament eines kumulativen Erkenntnisgewinns. Eine Beschränkung der Berufseignungsdiagnostik auf Kriteriumsvalidierungen ist „blinde Technologie, solange die Zusammenhänge zwischen Test- und Kriteriumsverhalten nicht psychologisch erhellt, und das heißt theoretisch auf den Begriff gebracht sind“ (Jäger, 1986, S. 284). Solch eine technokratische Sichtweise der Berufseignungsdiagnostik transportiert der Beitrag von Marcus. Zentrale diagnostische Fragen bleiben unberührt, wenn man lediglich (signifikante) Zusammenhangsmuster („miscellaneous correlations“, Cronbach, 1989, S. 155) zwischen Prädiktoren und

Kriterien betrachtet. Erst eine theoretische Einordnung dieses Zusammenhangs schöpft den Nutzen von Studien zur Kriteriumsvalidität aus. Für diese Einordnung ist es unabdingbar notwendig zu wissen, was gemessen wird. Die fundamentale Bedeutung von Konstruktannahmen für die Kriteriumsvalidierung wird im Folgenden für fünf Bereiche erläutert. (Die nachstehenden Ausführungen zur Bedeutung von Konstruktannahmen gelten für die von Marcus thematisierten konstruktorientierten Persönlichkeitstests. Für simulations- oder biographieorientierte Verfahren bedürfte es einer gesonderten Darstellung.)

(1) Um die Validität eines Verfahrens erschließen zu können, bedarf es Annahmen darüber, was der Prädiktor misst. Messick (1988, S. 37) hat dies anhand von Beispielen verdeutlicht. So ergibt sich z. B. eine andere Bewertung des nachweislichen Prädiktor-Kriteriums-Zusammenhangs, je nachdem, ob man ein und denselben Testwert als Ausprägung auf der Dimension „Flexibilität“ oder als Ausprägung der Dimension „Zerstreuung“ interpretiert. (2) Auch die Frage, ob man die Kriterien als solche akzeptiert, kann erst aufgrund der Interpretation der gemessenen Leistungsdimension entschieden werden. Eine unreflektierte Fixierung auf Prädiktor-Kriteriums-Zusammenhänge ist reaktionär. Während es sich bei Persönlichkeitstests als Prädiktoren um psychologisch definierte Konstrukte handelt, stellen Außenkriterien sozial definierte Konstrukte dar (Wiggins, 1973). Diese sind starken Veränderungen (z. B. den jeweiligen Werten und Normen sowie dem technologischen Wandel) ausgesetzt. Ein „guter“ Polizist war beispielsweise früher eine Respektsperson und ist heute „Bürger in Uniform“ – entsprechend ändern sich die Kriterien als Indikatoren für den Erfolg eines Polizisten. Eine allein an – zwangsläufig zeitlich *zurückliegenden* – Studien zur prädiktiven Kriteriumsvalidität orientierte Berufseignungsdiagnostik bezieht sich auf Verfahren, die *vor* einem möglichen Wandel der sozial definierten Kriterien oder *vor* einem möglichen technologischen Wandel zur Prognose eingesetzt wurden. Eine theorieleose Berufseignungsdiagnostik ist stets rückwärtsorientiert. (3) Konstruktannahmen ermöglichen es, Prädiktoren und Kriterien systematisch aufeinander zu beziehen. Komplexe Kriterien wie der Berufserfolg können theoriegeleitet in einzelne Facetten zerlegt werden, um dann zu entscheiden, welche Facette durch welchen Prädiktor vorhersagbar ist (siehe z. B. das Symmetrie-Prinzip; Wittmann, 1988). (4) Sofern differentielle Kriteriumsvaliditäten für einzelne Gruppen auftreten, kann ebenfalls nur vor dem Hintergrund theoretischer Annahmen entschieden werden, ob es sich um eine wünschenswerte Differenzierung oder um einen unerwünschten „bias“ handelt. (5) Daten zur Kriteriumsvalidierung wären schließlich von zweifelhaften Wert, wenn ihr Geltungsanspruch sich

ausschließlich auf die den Daten zugrunde liegende Situation beziehen würde. Der gängigen Interpretation von Kriteriumsvaliditäten liegen zumeist Annahmen zur zeitlichen Stabilität des prognostizierten und prognostizierenden Merkmals sowie Annahmen zur Situationsinvarianz – und somit Konstruktannahmen – zugrunde. Ohne Konstruktannahmen wären Generalisierungen von Kriteriumsvaliditäten und somit Absicherungen gegen bestimmte Arten von Stichproben- und Messfehler (siehe z. B. Schmidt, 1992) zumindest deutlich erschwert.

Das für die Berufseignungsdiagnostik notwendige Denken in Konstrukten / Theorien steht bei Marcus nicht hoch im Kurs – zum Schaden der Praxis, denn nichts ist be-KANT-lich praktischer als eine gute Theorie. Wir müssen wissen, *was* wir messen. Wenn wir das wissen, können wir die Messung für die Berufseignungsdiagnostik besser nutzen als ohne ein solches Verständnis. Für das Verständnis des veränderten Antwortverhaltens ist es z. B. notwendig, zwischen (1) einer fehlerhaften Selbsteinschätzung, (2) einer Beschönigung an sich korrekter Einschätzungen und (3) einer bewussten Täuschung zu unterscheiden (siehe z. B. Barrick & Mount, 1996; Mersman & Shultz, 1998; Paulhus, Bruce & Trapnell, 1995). Wir müssen wissen, dass möglicherweise ein und dasselbe Verfahren mit und ohne „verfälschten“ Antworten jeweils etwas anderes misst – wie Schmit und Ryan (1993) am Beispiel von in der Personalauswahl eingesetzten Persönlichkeitstests gezeigt haben. Studien und konzeptionelle Überlegungen zum Effekt eines veränderten Antwortverhaltens auf den Nutzen von Persönlichkeitstests in der Berufseignungsdiagnostik müssen sich daher – wie von Rosse, Stecher, Miller und Levin bereits 1998 gefordert – mit der Konstruktebene auseinandersetzen. Die von Marcus praktizierte Umgehung des eigentlichen Themas durch Fixierung auf Kriteriumsvaliditäten bringt uns erkenntnistheoretisch nicht weiter und erschwert die Erledigung der diagnostischen Arbeit. Dies fällt in dem Diskussionsbeitrag von Marcus kaum auf, da er nur eine der zahlreichen Aufgaben von Berufseignungsdiagnostikern betrachtet. Marcus (2003, S. 130) sieht es offensichtlich als alleinigen Zweck der Berufseignungsdiagnostik an, Unternehmen bei der Auswahl der geeignetsten Bewerber zu unterstützen und dabei Instrumente auszuwählen, deren zusätzlicher Nutzen die zusätzlichen Kosten übersteigt. Für eine derart amputierte Berufseignungsdiagnostik spielt die Frage nach der Bedeutung des veränderten Antwortverhaltens möglicherweise eine geringe Rolle. Tatsächlich sind die Aufgaben und die Möglichkeiten der Berufseignungsdiagnostik aber wesentlich umfassender. Selbst wenn man, wie Marcus, innerhalb der Diagnostik nur die Selektionsaufgabe betrachtet, so zählen zu dieser Aufgabe nicht nur die Auswahl eines Ver-

fahrens, sondern auch die Kombination verschiedener Verfahren sowie die gezielte und rationale Konstruktion von neuen Verfahren. Hierzu bedarf es theoretischer Kenntnisse. Auch für die anderen Aufgaben der Berufseignungsdiagnostik, z. B. die Bedingungsselektion, die Verhaltensmodifikation und die Bedingungsmodifikation (siehe bereits Cronbach & Gleser, 1963), bedarf es des Denkens in Konstrukten. Für die Berufseignungsdiagnostik ist es nicht gleichgültig, ob das Antwortverhalten einer Person die Ausprägung auf einem interessierenden und bekannten Persönlichkeitsmerkmal widerspiegelt oder Ausdruck der (diagnostisch nicht intendierten) Motivation und Fähigkeit zum Faken ist. Die Auswahl von Personen, die Auswahl und Kombination von Verfahren, die Konstruktion von Verfahren, die Zuordnung von Personen zu Bedingungen (z. B. zu Ausbildungen, Trainings, Arbeitsplätzen) oder die Modifikation der Bedingungen oder des Verhaltens (z. B. durch Training, Coaching, Arbeitsplatzgestaltung, Gestaltung von Lehrsystemen) u. v. a. m. variieren in Abhängigkeit von der Interpretation der Daten.

Marcus' Argumentation funktioniert nach dem Muster: Wo es nicht juckt, sollte man nicht kratzen. Aber: es juckt gewaltig. Dazu hat Bernd Marcus noch einmal beigetragen und das verdient unseren Dank.

Literatur

- Barrick, M. & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261–272.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443–507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence. Measurement, theory, and public policy* (pp. 147–171). Urbana, IL: University of Illinois Press.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Chicago, IL: University of Illinois Press.
- Herriot, P. (1989). Selection as a social process. In M. Smith & I. Robertson (Eds.), *Advances in selection and assessment* (pp. 171–187). New York, NY: Wiley.
- Jäger, A. O. (1986). Validität von Intelligenztests. *Diagnostica, 32*, 272–289.
- Kanning, U. P. (2003). Sieben Anmerkungen zum Problem der Selbstdarstellung in der Personalauswahl. *Zeitschrift für Personalpsychologie, 2*, 193–195.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie, 42*, 61–75.
- Marcus, B. (2003). Das Wunder sozialer Erwünschtheit in der Personalauswahl. *Zeitschrift für Personalpsychologie, 2*, 129–132.
- Mersman, J. & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences, 24*, 217–227.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*, 1012–1027.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Erlbaum.
- Paulhus, D. L., Bruce, M. M. & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100–108.
- Rosse, J. G., Stecher, M. D., Miller, J. L. & Levin, R. A. (1998). The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis and cumulative knowledge in psychology. *American Psychologist, 47*, 1173–1181.
- Schmit, M. & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*, 966–974.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie, 34*, 184–191.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 505–560). New York, NY: Plenum.

Dr. Martin Kersting

Institut für Psychologie der RWTH Aachen
 Jägerstr. 17/19
 52056 Aachen
 E-Mail: Kersting@psych.rwth-aachen.de