

Zur Akzeptanz von Intelligenz- und Leistungstests

Martin Kersting

Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420-433.

1. Einleitung

1.1 Theoretischer Hintergrund und praktische Relevanz

Das Thema »Akzeptanz der Diagnostik« hat seit einigen Jahren zumindest im Bereich der Personalpsychologie Konjunktur. Spätestens seitdem (wieder) ein Mangel an qualifiziertem Personal besteht, ist man besorgt, wie die Diagnostik und die diagnostisch begründete Personalentscheidung bei den Bewerbern ankommen. Durch den aktuellen Fachkräftemangel, der im Vergleich zu dem zukünftigen demografisch bedingten Personal-mangel harmlos ist, verlagert sich die »Macht« vom Anbieter auf die Nachfragerseite, sodass man aus strategischen Gründen – nicht nur des Personalmarketings – versucht, die Perspektive der »anderen Seite« (der Bewerber) einzunehmen. Hausknecht, Day und Thomas (2004) haben metaanalytisch nachgewiesen, dass sich eine positive Wahrnehmung des Auswahlprozesses durch die (ausgewählten *und* abgelehnten) Bewerber u. a. günstig auf das Ansehen der Organisation sowie auf die Bereitschaft auswirkt, ein Stellenangebot anzunehmen und die Organisation anderen potenziellen Bewerbern zu empfehlen. Ist die Personaldiagnostik und Entscheidung hingegen im Urteil der Bewerber unzumutbar, könnte sich dies – so die Befürchtung – nicht nur auf die Entscheidungen der Bewerber, sondern sogar ungünstig auf den gesamten Absatz von Dienstleistungen und/oder Produkten des jeweiligen Unternehmens auswirken. Zumindest kann eine Diagnostik mit mangelhafter Akzeptanz den Auftritt als attraktiver Arbeitgeber (»employer branding«) beschädigen. Aus diagnostischer Perspektive ist u.a. zu prüfen, ob mangelnde Akzeptanz auch zu einem (Test-)Verhalten führt, welches die Validität der Messung beeinträchtigt.

Die soziale Akzeptanz von Auswahlverfahren ist Gegenstand einer Fülle von Theorien und Forschungsstudien. Der Übersichtsarbeit von Ryan und Ployhart (2000) zufolge widmet sich ein Teil der Arbeiten vorrangig den Einstellungen und Motiven der Bewerber (z.B. Arvey, Strickland, Drauden & Martin, 1990), während ein anderer Teil der Arbeiten die Wahrnehmung des gesamten Auswahlprozesses (einschließlich der Auswahlentscheidung) fokussiert. Die zuletzt genannten Studien und Modelle sind vor dem Hintergrund der Theorien zur Austausch- und Verteilungsgerechtigkeit zu sehen, einem virulenten Forschungszweig der Psychologie sowie benachbarter Disziplinen (siehe beispielsweise die Metaanalyse von Colquitt, Conlon, Wessen, Porter & Ng, 2001). Der Ansatz von Gilliland (1993) baut unmittelbar auf der Gerechtigkeitstheorie auf. Auch das Konzept der sozialen Validität von Schuler und Stehle (1983, siehe auch Schuler, 1990) betont die Interaktion zwischen den Bewerbern und den Vertretern der auswählenden Organisation. Ein integrativer Charakter kommt den Modellen von Ryan und Ployhart (2000) sowie Hausknecht et al. (2004) zu.

Während die meisten der genannten Modelle *alle* bewerberseitig wahrgenommenen Prozessphasen einer Personalauswahlentscheidung berücksichtigen, konzentriert sich das Forschungsprojekt AKZEPT!, über das hier berichtet wird, auf die Akzeptanz von Tests, Persönlich-

keitsfragebogen und Assessment-Centern. Um den Beitrag eines derartigen Verfahrens zur sozialen Akzeptanz einer diagnostischen Situation ermeszen zu können, müssen zunächst Parameter identifiziert werden, die bei sonst weitgehend gleichen Situationsbedingungen in Abhängigkeit vom Verfahren variieren. Zahlreiche Anwender sind offensichtlich überzeugt, dass die Verfahren (z.B. Tests) eine unterschiedliche Akzeptanz bei den Teilnehmern finden, und berücksichtigen bei der Auswahl diagnostischer Verfahren Akzeptanzgesichtspunkte mit einer hohen Priorität (siehe z.B. Steck, 1997). Sie sind dabei aber weitgehend auf Spekulationen angewiesen. Empirische Ergebnisse finden sich lediglich zum Vergleich verschiedener diagnostischer Verfahrensklassen. So gilt es beispielsweise als ausgemacht, dass Interviews und Assessment-Center sich bei Bewerbern einer größeren Beliebtheit erfreuen als Tests (Hausknecht et al., 2004). Die zugrunde liegenden Untersuchungen kranken allerdings häufig daran, dass die Beurteiler den Beurteilungsgegenstand nicht hinreichend kennen. In der Mehrheit der Studien wurden die Akzeptanzurteile auf der Basis von Kurzbeschreibungen (sogenannte Vignetten) der zu beurteilenden Verfahren abgegeben. Marcus (2004) hat gezeigt, dass die Akzeptanzurteile, die nach der Lektüre einer solchen Kurzbeschreibung gegeben werden, nur gering mit den Akzeptanzurteilen korrelieren, die nach intensiverer Auseinandersetzung mit den Verfahren erfolgen. Eine angemessenere Operationalisierung würde allerdings die Schwächen der Fragestellung nicht kompensieren: Ob man ein Interview oder einen Test einsetzt, sollte von der diagnostischen Fragestellung, von den Hauptgütekriterien und von den weiteren Rahmenbedingungen abhängen. Der Akzeptanz kommt hingegen dann auswahlrelevante Bedeutung zu, wenn mehrere Verfahren unter den genannten Gesichtspunkten als weitgehend äquivalent erscheinen. Dies wird eher beim Vergleich von Verfahren mit überlappenden Geltungsansprüchen – beispielsweise beim Vergleich zweier Intelligenztests – der Fall sein als beim Vergleich so disparater Verfahren wie Interview und/oder Assessment-Center einerseits und Tests andererseits. Notwendig ist es daher, für vergleichbare Verfahren einen in standardisierter Form bestimmten Akzeptanzwert zu ermitteln und den Anwendern als Information zur Verfügung zu stellen. Aktuell nehmen zwar etliche Autoren für ihre Tests oder Fragebogen eine hohe Akzeptanz in Anspruch, die Deziertheit dieser Aussagen steht aber meist weit hinter der empirischen Grundlage zurück. Die Akzeptanz von Tests und Fragebogen wird von Autoren gerne behauptet, aber entweder überhaupt nicht oder nur mit unzureichenden Mitteln (z.B. mit Single-Item-Messungen) empirisch untersucht. Damit besteht die Gefahr, das Wünschenswerte nicht ausreichend vom Faktischen zu unterscheiden. Die für jede Wissenschaft konstitutive Wissensakkumulation bleibt aus, solange jeder Autor eine andere Form der Akzeptanzmessung wählt. Das Projekt AKZEPT! soll hier Abhilfe schaffen. Im Folgenden wird zunächst der Begriff »Akzeptanz« definiert, bevor fünf Postulate zur Messung der Akzeptanz formuliert werden. Anschließend wird der Akzeptanzfragebogen vorgestellt. Im empirischen Teil wird über drei Studien berichtet, in

DER AUTOR

MARTIN KERSTING ist Professor am Bildungs- und Wissenschaftszentrum der Bundesfinanzverwaltung in Münster. Er ist Mitautor der DIN 33430 sowie Autor von Tests und zahlreichen wirtschaftspsychologischen Publikationen. Für mehr Informationen siehe: <http://kersting-internet.de>

denen zum einen die psychometrische Qualität des Akzeptanzfragebogens ermittelt wurde und in denen zum anderen die Akzeptanzurteile über insgesamt fünf Intelligenz- oder Leistungstests erhoben wurden. Darüber hinaus wird der Einfluss der einzelnen Akzeptanzdimensionen sowie der eigenen (selbst eingeschätzten) Leistung auf das Akzeptanzurteil untersucht. Ein Fazit und ein Ausblick runden den Artikel ab.

1.2 Definition und Diagnose der Akzeptanz diagnostischer Verfahren

Zunächst wurde der Begriff »Akzeptanz« in Bezug auf diagnostische Verfahren definiert: Das Akzeptanzurteil stellt die Summe der auf das individuelle Erleben und Bewerten diagnostischer Verfahren gerichteten Aspekte dar, wobei deren Wechselwirkungen untereinander sowie mit Situationsparametern, Personvariablen und dem subjektiv wahrgenommenen diagnostischen Ziel berücksichtigt werden müssen.

Für die Akzeptanzurteile der diagnostizierten Personen über die diagnostischen Verfahren sollte der Begriff »Akzeptanz« und nicht der Begriff der »sozialen Validität« gewählt werden. Der Begriff »soziale Validität« impliziert Jäger (1986) zufolge eine unzweckmäßige Überdehnung eines der psychologischen Diagnostik angemessenen Validitätskonzepts. Die Verwendung des Begriffs »soziale Validität« suggeriert, dass es sich bei dem Akzeptanzurteil um eine Komponente der Validität handeln würde. Wenn dies der Fall wäre, dann wäre sie von anderen Reliabilitäts- oder Validitätskomponenten abhängig, de facto besteht aber keine sachlogische Verknüpfung des Nebengütekriteriums Akzeptanz mit dem Hauptgütekriterium der Validität (siehe Kersting, 1998). Es ist allerdings einzuräumen, dass die fachlich problematische, aber attraktive Bezeichnung »soziale Validität« möglicherweise wesentlich zum Aufschwung des Themas beigetragen hat. Die kontrafaktische Reinterpretation der Akzeptanz als Validitätsaspekt schafft ein leicht zu beschaffendes Surrogat für den Qualitätsstandard der Validität, der ungleich schwerer zu erfüllen ist.

Die Messung der Akzeptanz diagnostischer Verfahren sollte den folgenden fünf Postulaten gerecht werden.

(1) Akzeptanzmessungen zu diagnostischen Verfahren setzen voraus, dass die befragten Personen das in Frage stehende diagnostische Verfahren bearbeitet haben.

(2) Zur Akzeptanzmessung sollte ein nachweislich (2a) psychometrisch befriedigendes, (2b) theoriegeleitetes und (2c) mehrdimensionales Instrument eingesetzt werden:

(3) Situations- und Kontextvariablen (z.B. Freiwilligkeit, Anonymität, Ziel und Bedeutsamkeit der Diagnose) sowie

(4) Personvariablen (z.B. demografische Variablen, Fähigkeit, Persönlichkeit, Interesse, Motivation) sollten bei der Akzeptanzmessung berücksichtigt werden. Von besonderer Bedeutung ist die Erhebung und Berücksichtigung des Ergebnisses, das die beurteilende Person in dem beurteilten Verfahren erzielt hat, und/oder die Erhebung der diesbezüglichen Selbsteinschätzung.

(5) Sofern verschiedene Verfahren unter Akzeptanzgesichtspunkten miteinander verglichen werden sollen, müssen die Verfahren unter annähernd gleichen situativen Rahmenbedingungen eingesetzt und beurteilt werden.

Der Erkenntniswert von Studien zur Akzeptanz diagnostischer Verfahren ist in dem Maße eingeschränkt, in dem die hier genannten Postulate nicht erfüllt werden.

2. Der Akzeptanzfragebogen AKZEPT!

Empirischer Ausgangspunkt des Projekts AKZEPT! war die Entwicklung und langjährige empirische Evaluation des AKZEPT!-Fragebogens. Die vorhandenen Messinstrumente zur Erfassung der sozialen Akzeptanz eignungsdiagnostischer Verfahren (u.a. von Arvey, Strickland, Drauden & Martin, 1990; Bauer, Truxillo, Sanchez, Craig, Ferrara & Campion, 2001; Lounsbury, Bobrow & Jensen, 1989; sowie Nevo & Sfez, 1986) lagen nicht auf Deutsch vor und wurden nicht allen psychometrischen Kriterien und Projektzielen gerecht. Bislang wurden drei Varianten des AKZEPT!-Fragebogens entwickelt: für Intelligenz- und Leistungstests (AKZEPT!-L), für Persönlichkeitsfragebogen (AKZEPT!-P) und für Assessment-Center (AKZEPT!-AC). Die drei Varianten unterscheiden sich zunächst in den Itemformulierungen: Während im AKZEPT!-L beispielsweise von »Testaufgaben« gesprochen wird, geht es im AKZEPT!-P um »Verfahren«. Darüber hinaus gibt es neben Skalen, die in allen drei Varianten vorkommen, auch einige verfahrensspezifische Skalen. So wird z.B. nur bei Persönlichkeitsfragebogen die »Wahrung der Privatsphäre« thematisiert. Eine vierte Variante zur Messung der Akzeptanz von Interviews befindet sich in Vorbereitung. Informationen zu allen Varianten des AKZEPT!-Fragebogens finden sich unter www.kersting-internet.de/Akzeptanzfragebogen.html. Der AKZEPT!-Fragebogen steht allen Personen kostenlos zur Verfügung, die einen Rücklauf von aussagekräftigen (anonymisierten) Daten gewährleisten. Im vorliegenden Artikel geht es ausschließlich um die Akzeptanz von Intelligenz- und Leistungstests, die mit dem AKZEPT!-L erfasst werden kann.

Für die Bearbeitung der Items wird eine sechsstufige Zustimmungsskala genutzt, deren Extremausprägungen mit »trifft nicht zu« und »trifft genau zu« bezeichnet sind. Bei der Auswertung wird ein extrem negativer Wert mit »1«, ein extrem positiver Wert mit »6« codiert. Durch nachträgliche Umpolung negativ formulierter Items wird erreicht, dass hohe Werte einem positiven Akzeptanzurteil entsprechen.

Der AKZEPT!-L umfasst die vier Skalen Messqualität, Augenscheinvalidität, Kontrollierbarkeit und Belastungsfreiheit (siehe Tab. 1). Jede Skala umfasst vier Items. Drei der vier Skalen haben ihr theoretisches Fundament in denjenigen Gerechtigkeitsregeln nach Gilliland (1993), die sich auf Messinstrumente beziehen lassen (siehe Spalte 3 in Tabelle 1). Beispielitems für die Skalen sind ebenfalls in Tabelle 1 abgebildet. Bei der Messung der Augenscheinvalidität (»face validity« oder treffender: »faith validity«, nach Cattell, Eber & Tatsuoka, 1970) muss das diagnostische Ziel expliziert werden, sodass unterschiedliche Zielstellungen unterschiedliche Itemformulierungen bedingen. Einem Verfahren wird eine hohe Augenscheinvalidität zugesprochen, wenn die subjektiv wahrgenommene Messintention mit der seitens des Diagnostikanden vermuteten oder explizit genannten diagnostischen Fragestellung übereinstimmt, zu deren Beantwortung das be-

PROF. DR. MARTIN KERSTING Bildungs- und Wissenschaftszentrum der Bundesfinanzverwaltung Gescherweg 100 48161 Münster T+49 251 – 86 70 – 166 E martin@kersting-internet.de

HO-
GRE-
FE

HO-
GRE-
FE

Tabelle 1: Beispielitems, theoretischer Bezug und Reliabilität (interne Konsistenz) der AKZEPT!-L-Skalen

Skala (jeweils 4 Items)	Beispielitem	G-R ¹	Interne Konsistenz in Studie Nr.		
			1	2	3
Messqualität	Der Test ermöglicht es, die zwischen verschiedenen Menschen bestehenden Leistungsunterschiede in der vom Test gemessenen Fähigkeit exakt zu messen.	2,3	.66	.71	.77
Augenscheinvalidität	Dass man mit dem Test geeignete Personen für einen Job herausfinden kann, ist zu bezweifeln. (Negativ gepoltes Item)	1	.73	.67	.76
Kontrollierbarkeit	Bei der Bearbeitung der Testaufgaben wusste ich jederzeit, was ich tun muss.	8	.65	.75	.82
Belastungsfreiheit	Die Bearbeitung der Testaufgaben ist belastend. (Negativ gepoltes Item)	–	.67	.68	.75

¹ Bezug zu den prozeduralen Regeln, deren Beachtung oder Verletzung nach Gilliland (1993) die Akzeptanz beeinflussen. Eintrag in der Spalte: Nr. der Regel nach Gilliland (1993)

treffende Verfahren eingesetzt wird. Augenscheinvalidität ist somit ein relationaler Begriff, der die Wahrnehmung des Verfahrens mit der Wahrnehmung der diagnostischen Fragestellung verbindet (Kersting, 2003). Ein Test hat nicht eine Augenscheinvalidität, sondern für verschiedene (wahrgenommene) Anwendungszwecke verschiedene Augenscheinvaliditäten. Die bislang empirisch erprobten Items des AKZEPT!-L beziehen sich auf die Augenscheinvalidität für die Eignungsdiagnostik.

Darüber hinaus gibt es zwei Zusatzitems. Mit einem Item »Akzeptanz-Gesamtbeurteilung« wird gefragt: »Welche Schulnote würden Sie dem soeben bearbeiteten Test geben?« Zur Beantwortung steht die Schulnotenskala (von »sehr gut« bis »ungenügend« zur Verfügung). Außerdem wird abschließend – ebenfalls auf der Schulnotenskala – um eine Selbsteinschätzung zum eigenen Abschneiden im Test gebeten, wobei die eigene Bildung und das Alter berücksichtigt werden sollen: »Im Vergleich mit anderen Personen meiner Altersgruppe (mit gleicher Schulbildung) denke ich, dass ich im Test ... abgeschnitten habe.« Um das Verständnis der nachfolgend berichteten Ergebnisse und korrelativer Beziehungen zu erleichtern, wurden die beiden Zusatzitems vor der Auswertung umgepolzt, sodass – anders als bei Schulnoten – auch hier ein hoher Wert einer positiven Ausprägung entspricht. Somit sind alle im Folgenden berichteten positiven Korrelationen zwischen den AKZEPT!-L-Skalen und den Zusatzitems auch logisch (inhaltlich) positiv, hohe Werte stehen grundsätzlich für eine positive Akzeptanzbeurteilung.

3. Empirische Studien

3.1 Studie 1

Zunächst wurde der AKZEPT!-L bei 391 Bewerbern eingesetzt, die den neuen Wilde-Intelligenztest (WIT-2, Kersting, Althoff und Jäger, 2008) als Auswahlinstrument im Rahmen einer Ernstfall-Bewerbung um eine Berufsausbildung bei einem Konzern aus der Kommunikationsbranche bearbeiten mussten. Um Effekte sozialer Erwünschtheit zu vermeiden, wurde der Fragebogen unter der Bedingung vollständiger Anonymität bearbeitet. Aus der allgemeinen Bewerberstatistik des Unternehmens ist bekannt, dass die Bewerber mit mittlerer Reife (ca. 45% der Gesamtgruppe) im Durchschnitt 16 Jahre und die Bewerber mit Abitur im Durchschnitt 19 Jahre alt sind. Wie bei technischen Berufsausbildungen üblich, bewerben sich mehr Männer (ca. 65%) als Frauen.

3.2 Studie 2

Zur weiteren Forschung sollten auch Daten von Personen herangezogen werden, die den zu beurteilenden Test – anders als in Studie 1 – nicht unter der Ernstfallbedingung einer Bewerbung bearbeitet haben. Dazu sollte vorab geprüft werden, ob und in welchem Ausmaß die Ergebnisse einer solchen Gruppe mit den Ergebnissen der Ernstfallgruppe vergleichbar sind. Dazu wurde in dem Unternehmen, in dem Studie 1 durchgeführt wurde, eine Gruppe von 78 Personen gewonnen, die das Auswahlverfahren vor einiger Zeit (als noch ein anderer Test zur Eignungsdiagnostik eingesetzt wurde) erfolgreich durchlaufen hatten und sich ebenfalls einer Testung mit dem Wilde-Test (Version 2) unterzogen – allerdings freiwillig. Der Anteil der Personen mit (Fach-)Hochschulreife war hier mit 77% etwas höher als in Studie 1, der Anteil der Frauen mit 28,2% annähernd vergleichbar. Die Personen waren mit durchschnittlich 21,5 Jahren etwas älter als die Bewerbergruppe in Studie 1.

3.3 Studie 3

Teilnehmer der Studie 3 waren Psychologiestudenten im Hauptstudium. Es liegen 95 Akzeptanzfragebogen zum BIS-4 (Jäger, Süß & Beauducel, 1997), 103 Bogen zum IST-2000 R (Amthauer, Brocke, Liepmann & Beauducel, 2001), 66 Bogen zu den Raven-Matrizen (Advanced Progressive Matrices, Bulheller & Häcker, 1998) und 90 Bogen zum d2 (Brickenkamp, 2002) vor. Somit stehen die Daten von 354 Akzeptanzbefragungen zur Verfügung. Dies ist aber nicht gleichbedeutend mit der Stichprobengröße, da einige Personen zwei, drei oder vier Tests bearbeitet und anschließend jeweils einen Akzeptanzfragebogen ausgefüllt haben. In diesen Fällen fand die Testung und Befragung an verschiedenen Terminen (mindestens ein Tag Abstand) statt, die Darbietungsfolge wurde kontrolliert und zeitigte keinen Einfluss auf die nachfolgend berichteten Ergebnisse. Testung und Akzeptanzbefragung erfolgten anonym, lediglich demografische Angaben wurden erhoben und kontrolliert. Da es sich um eine Gruppe von Studenten handelte, waren die Personen vollständig bildungs- und weitgehend altershomogen (Durchschnittsalter: 24,6), der Frauenanteil war, der Studienrichtung entsprechend, mit 82% konstant hoch. Die Variablen Alter und Geschlecht indizieren keine subgruppenspezifische (bezogen auf die »Gruppe BIS«, die »Gruppe IST« usw.) Signifikanz auf die nachfolgend berichteten Ergebnisse.

**AKAD
EMIE**

HUBER

75

X

256

99 x

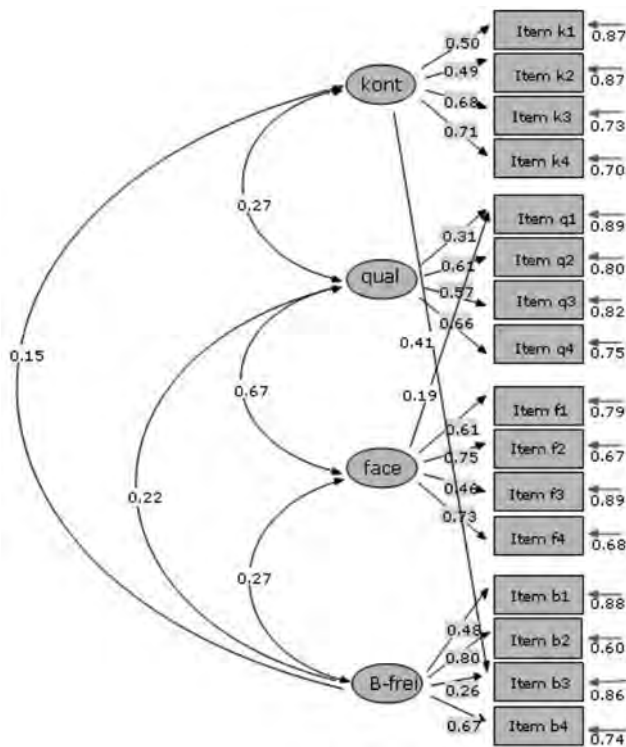


Abb. 1: Konfirmatorische Faktorenanalyse, Studie 1 (N=391) Erläuterung: kont: Kontrollierbarkeit; qual: Messqualität; face: Augenscheinvalidität; B-frei: Belastungsfreiheit. Datenfit: $\chi^2=203.36$, $df=97$, $p=.000$; CFI=.92; SRMR=.064; RMSEA=.053, CI 90% .043-.063.

In allen drei Studien erfolgte die Akzeptanzbefragung im unmittelbaren Anschluss an die Testbearbeitung.

4. Ergebnisse

4.1 Reliabilität

Für die vier Skalen des AKZEPT!-L wurde die Reliabilität über die internen Konsistenzen geschätzt, die Cronbachs-Alpha-Werte sind in Tabelle 1 wiedergegeben, sie können als adäquat bezeichnet werden. Aus einer anderen Untersuchung (Kersting, 1998) ist die Retest-Reliabilität einer auf drei Items verkürzten Form der AKZEPT!-L-Skala-Augenscheinvalidität bekannt. Die im Abstand von durchschnittlich einem Jahr vorgenommene Erst- und Wiederholungsmessung korrelierten zu .85 miteinander, sodass die Retest-Reliabilität deutlich höher ausfiel als die interne Konsistenz, die als Homogenitätsmaß ggf. die Reliabilität der AKZEPT!-L-Skalen unterschätzt.

4.2 Konstruktvalidität

Die Konstruktvalidität des AKZEPT!-L wurde anhand der Daten der 391 Personen aus Studie 1 bestimmt. Studie 2 umfasst zu wenig Personen, um konfirmatorische Faktorenanalysen zu rechnen, bei Studie 3 besteht die Gefahr einer Konfundierung von Inhaltsfaktoren und Methodenfaktoren (aufgrund der unterschiedlichen Test, die dem Akzeptanzurteil zugrunde liegen). In einer konfirmatorischen Faktorenanalyse (ML-Schätzung) konnte die Annahme der vier Akzeptanzdimensionen (siehe Tab. 1) empirisch bestätigt werden (siehe Abbildung 1). Das Modell zeigt einen akzeptablen Datenfit ($\chi^2=203.36$, $df=97$, $p=.000$; CFI=.92; SRMR=.064; RMSEA=.053, CI 90% .043-.063). Mit Ausnahme der Faktoren Kontrollierbarkeit und Augenschein-

validität (kein signifikanter Zusammenhang) sowie Messqualität und Augenscheinvalidität (hoher Zusammenhang) sind alle Faktoren zwar miteinander korreliert, die Korrelationshöhen im Bereich von .15 bis .27 stellen eine Interpretation der einzelnen Dimensionen aber nicht in Frage. Bei der hohen Korrelation der Dimensionen Messqualität und Augenscheinvalidität ist zu bedenken, dass es sich bei dem neuen Wilde-Test (WIT-2, Kersting et al., 2008) um ein Verfahren handelt, welches sich durch Aufgabenmaterial aus dem Berufsalltag und durch die Integration einer Arbeitsprobe explizit um Augenscheinvalidität bemüht; es ist daher nicht verwunderlich, dass bei diesem Test die Urteile über die Validität (Messqualität) und Augenscheinvalidität systematisch kovariieren. Es ist zu erwarten, dass die beiden Faktoren sich bei Tests mit »abstrakterem« Aufgabenmaterial deutlicher trennen lassen, sodass vorerst an der Konzeption der Unterscheidbarkeit der Messqualität und Augenscheinvalidität festgehalten wird. Über die gleichsinnige Kovariation der Urteile über diese beiden Dimensionen darf nicht vergessen werden, dass die Urteile in der absoluten Höhe sehr unterschiedlich ausfallen (siehe Tabelle 2), was gegen eine Zusammenfassung der beiden Dimensionen spricht.

4.3 Vergleichbarkeit der Akzeptanzbeurteilungen bei einem Ernstfalleinsatz und einem Forschungseinsatz des zu beurteilenden Tests

Ein Vergleich der Akzeptanzeinschätzungen in den Studien 1 und 2 ermöglicht es abzuschätzen, ob und in welchem Ausmaß die Situation («Ernstfall» versus »freiwillig») sich auf die Akzeptanzurteile auswirkt, da die übrigen Variablen (insb. der zu beurteilende Test, nämlich der WIT-2) weitgehend konstant gehalten wurden. Trotz der sehr unterschiedlichen Situationen beurteilen beide Gruppen den Test einheitlich, es ergeben sich keinerlei statistisch bedeutsamen Unterschiede im Akzeptanzurteil in den vier Skalen. Wir gehen daher davon aus, dass die nachfolgend anhand von »Freiwilligen«-Daten (Studie 3) gewonnenen Ergebnisse zumindest strukturell auch auf »Ernstfallsituationen« übertragen werden können.

4.4 Die Akzeptanz verschiedener Testverfahren

4.4.1 Deskriptive Statistik

Tabelle 2 sowie Abbildung 2 zeigen die deskriptiven Akzeptanzwerte für die fünf Testverfahren. Für den WIT-2 wurden die Daten der Studie 2 herangezogen, da diese Daten – wie bei den anderen Tests – von »Freiwilligen« stammen, gleichwohl vorab gezeigt wurde, dass dies keinen bedeutsamen Unterschied verursacht. Betrachtet man zunächst das Ergebnis über alle Tests hinweg, zeigt sich ein klares Akzeptanzprofil für Intelligenz- und Leistungstests. Zumindest für die hier befragte leistungsstarke Gruppe stellt die Testung eine sehr gut zu beherrschende, kontrollierte Situation dar, die Tests werden in dieser Hinsicht von den Testteilnehmern sehr positiv beurteilt. Auch die Messqualität der meisten Tests wird positiv eingeschätzt, die Teilnehmer sind überwiegend überzeugt, dass der Test das misst, was er messen soll. Allerdings fällt es den Teilnehmern schwer zu realisieren, dass diese Messung – so gut sie auch ist – relevant ist für die diagnostische Fragestellung

ANZEIGE

DPV

18 28 01

Tabelle 2: Akzeptanzurteile über verschiedene Intelligenz- und Leistungstests (Studie 3)

	WIT-2 (N=78)		IST-2000 R (N=103)		BIS-4 (N=95 ¹)		d2 (N=90)		Raven (APM) (N=66)	
	MW	SW	MW	SW	MW	SW	MW	SW	MW	SW
Messqualität	3.82	.77	4.02	.87	3.93	.78	3.99	.90	4.62	.96
Augenscheinvalidität	3.30	.85	3.26	.90	3.25	1.05	2.75	.96	2.45	.87
Kontrollierbarkeit	5.36	.83	4.79	1.03	5.11	.92	5.78	.44	5.43	.58
Belastungsfreiheit	4.51	.87	3.52	.83	3.62	.97	3.83	.98	2.92	.90
Gesamtbeurteilung	3.73	.62	3.33	.76	3.24 ¹	.94	2.78	1.13	2.36	1.00

Erläuterung: MW=Mittelwert; SW=Standardabweichung;
Skala von 1 (schlechtester Wert) bis 6 (besten Wert)

¹ Beim BIS-4 wurde bei einer Subgruppe die Akzeptanzgesamtbeurteilung nicht eingeholt. Dadurch ergibt sich für die Gesamtbeurteilung beim BIS-4 ein vermindertes N im Umfang von N=59.

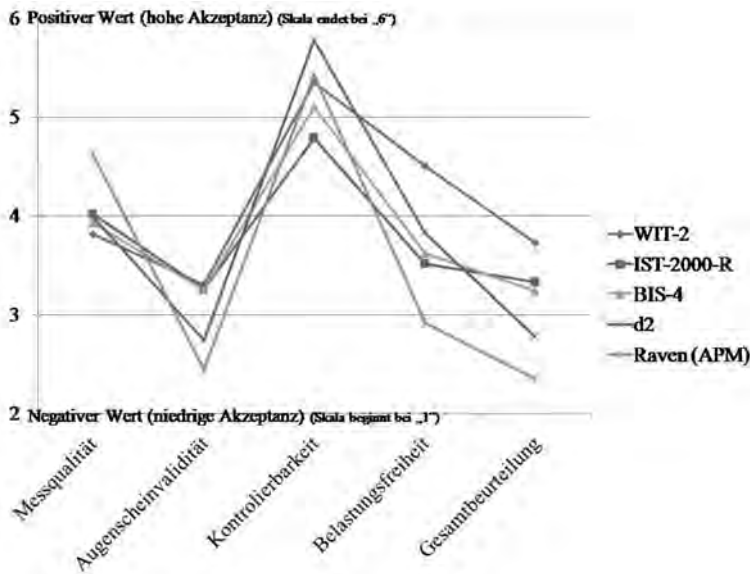


Abb. 2: Akzeptanzurteile über verschiedene Intelligenz- und Leistungstests (Studie 3, N: siehe Tabelle 2)

(hier der zukünftige Ausbildungs- und Berufserfolg). Die Tests werden unter dem Gesichtspunkt der Augenscheinvalidität relativ kritisch bewertet. Eine Benotung im mittleren Bereich gibt es für die Belastungsfreiheit, wobei auch nicht zu erwarten ist, dass ein Intelligenz- und/oder Leistungstest als vollkommen belastungsfrei wahrgenommen wird. Das relativ gute Abschneiden des d2-Tests in dieser Dimension ist vor dem Hintergrund zu sehen, dass dieser Test zwar sehr anstrengend, aber zeitlich deutlich kürzer als die übrigen Verfahren ist. Ungeachtet dieser Gemeinsamkeiten, zeigen sich aber durchaus beachtliche Akzeptanzunterschiede zwischen den Verfahren. So fallen die Höhen und Tiefen der Akzeptanzbewertung für den Raven-Test besonders extrem aus: Die permanente Wiederholung ein und desselben Itemtyps wirkt sich günstig auf die Akzeptanzurteile über die Messqualität und Kontrollierbarkeit aus, die Teilnehmer können aber kaum einen Zusammenhang zwischen den abstrakten Testitems einerseits und Aufgaben aus dem Leben andererseits herstellen. Ähnlich verhält es sich mit dem d2, der ebenfalls schlechte Werte in der Augenscheinvalidität bekommt. Die kritische Beurteilung der Augenscheinvalidität ist besonders bedeutend, da das Gesamt-Akzeptanzurteil wesentlich von der Augenscheinvalidität geprägt wird (siehe unten). Diesbezüglich schneiden Tests mit unterschiedlichen Anforderungen (z.B. schlussfolgerndes

Denken und Merkfähigkeit) und Aufgabenmaterial (z.B. verbal, numerisch und figural) deutlich besser ab.

4.4.2 Mittelwertsvergleiche

Zum IST-2000 R liegen Daten von 50 Personen vor, die zusätzlich den BIS-4, sowie von 86 Personen, die zusätzlich den d2 bearbeitet und unter Akzeptanzgesichtspunkten beurteilt haben. Die hier möglichen zufallskritischen Absicherungen bestätigen die oben bereits dargestellten Beobachtungen. Der Vorsprung des IST-2000 R gegenüber dem d2 in der Augenscheinvalidität lässt sich statistisch absichern ($t = -4.53$), der Vergleich zur Augenscheinvalidität des BIS-4 fällt hingegen unentschieden aus. Interessant ist der relativ schlechte Wert für die Kontrollierbarkeit des IST-2000 R. In dieser Dimension bleibt der Test sowohl hinter dem BIS-4 ($t = 2.19$) als auch vor allem hinter dem d2 ($t = 9.51$) zurück. Möglicherweise trägt neben anderen Gründen vor allem die Aufgaben- gruppe Figuren analogien (FA) zu diesen schlechten Noten bei. Die IST-2000-R-Testteilnehmer werden 120 Items lang daran gewöhnt, dass es stets ein Item gibt und zu diesem Item jeweils eine Anzahl von Distraktoren, unter denen man auswählt, bevor man zum nächsten Item übergeht. Bei Item 121 wird dieses gelernte Prinzip plötzlich durchbrochen, ohne dass dies den Teilnehmern angekündigt wird. Nun gibt es auf einmal fünf »fixe« Distraktoren für zehn Items – die meisten Personen, die diesen Test einmal angewiesen haben, werden die entsprechenden Bekundungen des Missverständnisses und des Erstaunens der Testteilnehmer kennen. Auch die Aufgabenstellung bei den Subtests Rechenaufgaben und Rechenzeichen sind nicht einfach zu verstehen. Ob die hier genannten Gründe tatsächlich die Ursachen für den relativ gesehen schwächeren Wert der Kontrollierbarkeit (der absolut gesehen immer noch im positiven Bereich liegt) sind, bleibt an dieser Stelle Spekulation: Mit dem AKZEPT!-L liegt aber ein Fragebogen vor, der sehr sensitiv für Akzeptanzeinschränkungen ist und auf das Vorliegen von Problemen aufmerksam macht, die dann identifiziert und im Zuge der Qualitätssicherung und -optimierung angegangen werden können. Gleichzeitig zeigt das mehrdimensionale Instrument auch die Stärken eines Verfahrens auf, im Falle des IST-2000 R beispielsweise dessen gute Werte für die Messqualität und Augenscheinvalidität sowie die gute Gesamtbewertung. Weitere Erkenntnisse gewinnt man, wenn man den AKZEPT!-Fragebogen auf der Ebene der Aufgabengruppen (und nicht wie hier auf der Ebene des Gesamttests) einsetzt, sowie vor allem, wenn man systematische Variationen der Aufgabengestaltung vornimmt.

4.4.3 Vorhersage des Akzeptanzgesamturteils durch die Skalen

Die Teilnehmer wurden schließlich noch gebeten, ein abschließendes Akzeptanzgesamturteil über den Test abzugeben. Dieses Gesamturteil aus der Studie 3 wurde als abhängige Variable in eine multiple schrittweise Regression eingebracht, wobei die vier Skalen sowie die Selbsteinschätzung als Prädiktoren berücksichtigt wurden. Wie in der Ergebnistabelle 3 zu sehen ist, spielt die Belastungsfreiheit keine Rolle für das abschließende

V&R
86
X
259

KLETT
99 x 259

Gesamturteil. Über alle Tests hinweg betrachtet, ist vor allem die Einschätzung der Augenscheinvalidität von zentraler Bedeutung für das Akzeptanzgesamturteil. Der durchaus steigerbare Anteil der Varianz des Akzeptanzgesamturteils, der durch die Prädiktoren aufgeklärt wird, zeigt, dass entweder das Kriterium als Single-Item-Messung keine hohe psychometrische Qualität aufweist oder dass es noch bedeutsame andere Prädiktoren gibt.

4.4.4 Der Effekt der (wahrgenommenen) eigenen Leistung auf das Akzeptanzurteil

Es ist bekannt und konsistenztheoretisch plausibel, dass die Akzeptanz eines diagnostischen Verfahrens systematisch mit den Ergebnissen kovariiert, die eine Person

Tabelle 3: Hierarchische Regressionen – Vorhersage der Akzeptanzgesamtbewertung durch die Skalen sowie durch die selbst eingeschätzte Leistung (Studie 3)

	IST-2000 R	BIS-4	d2	Raven (APM)	Gesamt
Messqualität	–	.51	.40	–	.16
Augenscheinvalidität	.27	.24	.33	.50	.44
Kontrollierbarkeit	.21	–	–	–	–
Belastungsfreiheit	–	–	–	–	–
Selbsteinschätzung	.31	.24	–	.38	.27
R	.55	.73	.61	.72	.62
Korr. R2	.28	.50	.36	.51	.38
F	14,48	20,37	26,12	34,33	66,83
Sig. p <	.01	.01	.01	.01	.01
N	103	59	90	66	318

Erläuterung: In den Zellen: standardisierte Beta Koeffizienten der berücksichtigten Variablen

in diesem Verfahren erzielt (siehe Hausknecht et al., 2004). Ein letzter Analyseschritt galt der Frage, ob die Akzeptanzurteile in Studie 3 von dem eigenen Abschneiden im Test moderiert wurden. Dabei wurde zwischen dem tatsächlichen Testergebnis (welches den Teilnehmern zum Zeitpunkt der Akzeptanzeinschätzung nicht bekannt war) und dem selbst eingeschätzten Abschneiden im Test (im Vergleich zu anderen Personen mit vergleichbarer Bildung) unterschieden. Die Ergebnisse sind in Tabelle 4 dargestellt. Grundsätzlich gaben leistungsstärkere Personen positivere Akzeptanzurteile ab. Für das tatsächliche Abschneiden im Test zeigte sich der deutlichste Zusammenhang mit der Einschätzung der Messqualität. Demgegenüber zeigte sich der Zusammenhang bei den Skalen Augenscheinvalidität

und Belastungsfreiheit vor allem für die selbst eingeschätzte Leistung, hier werden Korrelationen in beachtlicher Höhe erreicht.

5. Fazit und Ausblick

Praktiker gehen davon aus, dass Intelligenz- und Leistungstests eine geringe Akzeptanz bei Bewerbern finden (siehe Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Ihre Entscheidung, aus diesem Grunde auf derartige Tests zu verzichten, vermittelt die Diktion, welche man braucht, um als »Beschützer« der Bewerber zu gelten. De facto geht der Vorwurf, dass Intelligenz- und Leistungstests keine Akzeptanz finden, allerdings schlicht an der Wirklichkeit vorbei. Die Akzeptanzeinschätzungen der Bewerber differieren erheblich von den Akzeptanzeinschätzungen, die den Bewerbern von Organisationsvertretern zugeschrieben werden. Die Akzeptanz von Intelligenz- und Leistungstests ist insgesamt positiv. Dabei sind die Teilnehmer in der Lage, ein differenziertes Urteil über diagnostische Verfahren abzugeben. Setzt man entsprechend einen mehrdimensionalen Akzeptanzfragebogen ein, zeigt sich, dass Verfahren nicht einen, sondern mehrere Akzeptanzwerte haben. Über alle Tests hinweg betrachtet, werden die Messqualität und vor allem die Kontrollierbarkeit von Intelligenz- und Leistungstests besonders positiv, die Augenscheinvalidität wird hingegen kritisch beurteilt. Die Testbearbeitung wird von den Teilnehmern zwar als belastend empfunden, dies wirkt sich aber nicht negativ auf das von den Teilnehmern abgegebene Akzeptanzgesamturteil aus, offensichtlich wird die Belastung als angemessen gewertet. Für das abschließende Gesamtakzeptanzurteil zu einem Verfahren ist – über verschiedene Tests hinweg betrachtet – vor allem die wahrgenommene Augenscheinvalidität ausschlaggebend. Neben diesen Gemeinsamkeiten im Akzeptanzurteil über verschiedene Tests gibt es aber auch deutliche Unterschiede dahin gehend, welche Akzeptanz ein einzelnes Verfahren bei den Teilnehmern findet.

Zu berücksichtigen ist, dass neben den Situationsparametern auch Personmerkmale, insbesondere die (wahrgenommene) Fähigkeit, das Akzeptanzurteil wesentlich mitbestimmen. Erklärungen für diese interindividuell unterschiedliche Wirkung eines diagnostischen Verfahrens finden sich beispielsweise in den kognitiven Theorien der Einstellung. In dem vorliegenden Artikel wurde gezeigt, dass das Akzeptanzurteil (insb. über die Belastungsfreiheit) vom (selbst eingeschätzten) eigenen Abschneiden moderiert, aber nicht grundsätzlich in Frage gestellt wird. Mit dem AKZEPT!-Fragebogen liegt ein reliables und konstruktvalides mehrdimensionales Instrument zur Erfassung der Akzeptanz diagnostischer Verfahren vor. Möglicherweise gibt es weitere relevante Akzeptanzdimensionen, die mit dem AKZEPT!-Fragebogen noch nicht erfasst werden. Zukünftige Studien können prüfen, ob neue Skalen zu weiteren Dimensionen über den AKZEPT!-Fragebogen hinaus inkrementelle Validität an Kriterien aufklären. Der AKZEPT!-Fragebogen und die Beachtung der weiter oben formulierten Postulate zur Messung der Akzeptanz diagnostischer Verfahren können dazu beitragen, dass Arbeiten zur sozialen Akzeptanz nicht länger den Charakter von Trends, sondern von

Tabelle 4: (Selbst eingeschätztes) Ergebnis im Test und das Akzeptanzurteil (Studie 3)

	IST-2000 R			BIS-4		d2		Raven (APM)		Gesamt
	SD N=102	SE N=103	K N=93	SE N=94	KL N=90	SE N=86	RW N=66	SE N=66	Z ¹ N=351	SE N=349
MQ	-.19	.08	.21**	.04	.23*	.11	.36**	.20	.20**	.03
AV	.18	.21*	.14	.16	.06	.10	.02	.34**	.12*	.21**
KB	.06	.18	.17	.14	-.05	.03	.33*	.12	.17**	.11*
BF	.23*	.37**	.29**	.30**	.18	.45**	.30*	.51**	.24**	.45**

* = .05; ** = .01 Erläuterungen: MQ=Messqualität; AV=Augenscheinvalidität; KB=Kontrollierbarkeit; BF=Belastungsfreiheit; SD: schlussfolgerndes Denken; SE=Selbsteinschätzung der Leistung; K: Verarbeitungskapazität; KL=Konzentrationsleistungswert; RW: Rohwert (Anzahl richtig gelöster Aufgaben); Z: tatsächliche Leistung in allen Leistungstests

¹ Die Leistungen wurden vor der Aggregation pro Test z-transformiert.

DPV

73

x

259

KLETT

PSYCHE

99 x 259

Fortschritten erhalten. Dabei steht die Forschung weitgehend am Anfang. Die einseitige Verortung der entsprechenden Forschungsarbeiten im Bereich der Personalpsychologie sollte überwunden werden, denn die Fragen danach, ob die Teilnehmer die Instruktionen verstehen, wie belastet sie sich fühlen und ob sie einen Bezug zwischen dem diagnostischen Verfahren und der diagnostischen Zielsetzung herstellen können, sind für alle Gebiete relevant, in denen diagnostische Verfahren angewandt werden.

Auf der Ebene der Information ist anzustreben, dass die Akzeptanz von diagnostischen Verfahren in Zukunft standardisiert unter Nutzung eines Referenzinstruments wie z.B. dem AKZEPT!-Fragebogen bestimmt und in den Testmanualen veröffentlicht wird (Ryan & Ployhart, 2000; Trost, 1993). Dies würde dazu führen, dass es eine Art »Normwert« für die Akzeptanz von Tests gibt, sodass geprüft werden kann, ob ein Verfahren eine hohe, durchschnittliche oder niedrige Akzeptanz erzielt. Ein zusätzlicher Bericht über die Ausprägung solcher situativer und personbedingter Variablen, die das Akzeptanzurteil beeinflussen, kann die angemessene Interpretation der Daten unterstützen. Auf der Ebene der Erklärung und Vorhersage sind theoriegeleitete Studien mit systematischen Variationen von Merkmalen wie Vorinformation, Anweisung/Instruktion, Verfahrensqualität (gemessen an den Hauptgütekriterien), Antwortformat und Testdauer wünschenswert. So könnte

beispielsweise geprüft werden, ob die Augenscheinvalidität durch Informationen und/oder durch eine bestimmte Gestaltung der Aufgaben (z.B. nach Art einer Arbeitsprobe) beeinflusst werden kann, sodass Erkenntnisse zu einer akzeptanzfördernden Gestaltung von Tests gewonnen werden, die nicht zulasten der eigentlichen Testgüte geht. Auch potenzielle Korrelate des Akzeptanzurteils (wie z.B. Testängstlichkeit/Testmotivation) sowie die differenziellpsychologische Beeinflussung durch Personmerkmale wie beispielsweise Erfahrungshintergrund, Leistungs- und Persönlichkeitsmerkmale sowie Alter und Geschlecht bedürfen der weiteren Aufklärung. Dies gilt umso mehr, da einige dieser Merkmale wiederum mit Außenkriterien, wie z.B. Berufserfolg, kovariieren (siehe z.B. Hülshager & Maier, 2008). Letztendlich geht es auch darum, den Zusammenhang zwischen dem Akzeptanz erleben und dem Verhalten von Personen während und nach der Bearbeitung des diagnostischen Verfahrens in Erfahrung zu bringen und beherrschbar werden zu lassen.

Die Forschung und Praxis zur sozialen Akzeptanz von diagnostischen Verfahren wird in dem Maße produktiv, in dem sie zusätzlich zum Nebenkriterium Akzeptanz die Hauptgütekriterien der diagnostischen Verfahren berücksichtigt. Es kommt nicht allein darauf an, dass die Verfahren bei den Teilnehmern gut ankommen, sondern es kommt darauf an, dass diagnostische Verfahren valide sind und zusätzlich gut ankommen.

ZUSAMMENFASSUNG

Das Erleben und Bewerten diagnostischer Verfahren aus der Sicht der Verfahrensteilnehmer wird immer häufiger thematisiert. Dies gilt – angesichts des Bewerbermangels – insbesondere im Bereich der Personalpsychologie. Aber auch andere Anwendungsfelder wie die Pädagogische oder Klinische Psychologie sind auf die Motivation und Compliance der Teilnehmer und somit auf die Akzeptanz der Verfahren angewiesen. Entsprechend berücksichtigen zahlreiche Anwender bei der Auswahl diagnostischer Verfahren Akzeptanzgesichtspunkte, wobei sie aber weitgehend auf Spekulationen angewiesen sind. Der Artikel skizziert zunächst kurz potenzielle Auswirkungen der Akzeptanzeinschätzungen sowie unterschiedliche Konzepte der sozialen Akzeptanz und formuliert fünf Postulate, die bei der Messung der Akzeptanz diagnostischer Instrumente berücksichtigt werden sollten. Im Mittelpunkt der Arbeit steht der AKZEPT!-Fragebogen, ein technisch hochwertiger mehrdimensionaler Fragebogen zur Erfassung der Akzeptanz diagnostischer Verfahren. Mit Hilfe dieses Fragebogens wurden in drei Studien Akzeptanzurteile über fünf Intelligenz- und Leistungstests (BIS-4, d2, IST-2000 R, Ravens APM und WIT-2) gesammelt, wobei auch Personvariablen (das selbst eingeschätzte eigene Abschneiden im Test) und Situationsvariablen (Testung im »Ernstfall« oder im freiwilligen Kontext) berücksichtigt wurden. Der Artikel stellt ausgewählte Ergebnisse vor und zeigt Entwicklungslinien für die Akzeptanzforschung sowie für die Testpraxis auf.

Schlagwörter: Akzeptanz, soziale Validität, Eignungsdiagnostik, Personalpsychologie, Personalmarketing

ABSTRACT

Recent discussions have brought up the issue of evaluating diagnostic procedures from the perspective of how participants experience and evaluate these procedures. This applies in particular to the field of personnel psychology (in the face of the »war for talents«), but also other applied fields like the educational or clinical psychology are dependent on the motivation and compliance of the participants and on the acceptance of the diagnostic procedures. Consequently, many practitioners are considering acceptance as a criterion when selecting diagnostic procedures, without being sure what exactly acceptance entails. Against this background, the present article first briefly outlines potential effects of acceptance and different acceptance concepts. After formulating five postulates that should be considered when assessing the acceptance of diagnostic procedures, the article introduces the AKZEPT! (AKZEPT!) questionnaire as a methodologically sound and sophisticated multidimensional measure instrument for acceptance of diagnostic procedures. Three studies are described using the AKZEPT! questionnaire to gather acceptance data on five widely-used intelligence and achievement tests (BIS-4, d2, IST-2000 R, Raven's APM, and WIT-2) taking into consideration how individual differences (e.g., achievement) and situational differences (e.g., voluntary vs. obligatory participation) influence the experience, evaluation, and acceptance of the five tests. Selected results are presented, and future directions for possible improvements for acceptance research and testing practice are discussed.

L I T E R A T U R

- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). Intelligenz-Struktur-Test 2000 R. Göttingen: Hogrefe.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test-taking. *Personnel Psychology*, 43, 695-716.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P. & Campion, M. A. (2001). Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, 54, 387-419.
- Brickenkamp, R. (2002). Test d2 – Aufmerksamkeits- und Belastungstest. (9., überarbeitete und neu normierte Auflage). Göttingen: Hogrefe.
- Bulheller, S. & Häcker, H. (1998). Raven's Progressive Matrices und Vocabulary Scales, von J. C. Raven, J. Raven und J. H. Court. Frankfurt: Swets & Zeitlinger.
- Cattell, R. B., Eber, H. W. & Tatsuoka, M. M. (1970). Handbook for the sixteen Personality Factor Questionnaire (16 PF). Champaign, IL: Institute for Personality and Ability Testing.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H. & Ng, K. Y. (2001). Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology*, 86, 425-445.
- Gilliland, S. W. (1993). The Perceived Fairness Of Selection Systems: An Organizational Justice Perspective. *Academy of Management Review*, 18, 694-734.
- Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683.
- Hülshöger, U. R. & Maier, G. W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf. Eine Bestandsaufnahme internationaler und nationaler Forschung. *Psychologische Rundschau*, 58, 108-122.
- Jäger, A. O. (1986). Validität von Intelligenztests. *Diagnostica*, 32, 272-289.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). Berliner Intelligenzstruktur-Test (Form 4) Handanweisung. Göttingen: Hogrefe.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseparadigmen als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61-75.
- Kersting, M. (2003). Augenscheinvalidität. In K. D. Kubinger & R. S. Jäger (Hrsg.), Schlüsselbegriffe der Psychologischen Diagnostik (S. 54-55). Weinheim: Beltz.
- Kersting, M., Althoff, K. & Jäger, A. O. (2008). WIT-2. Der Wilde-Intelligenztest. Verfahrenshinweise. Göttingen: Hogrefe.
- Lounsbury, J. W., Bobrow, W. & Jensen, J. B. (1989). Attitudes Toward Employment Testing: Scale Development, Correlates, and »Known-Group« Validation. *Professional Psychology: Research and Practice*, 20, 340-349.
- Marcus, B. (2003). Attitudes toward Personnel Selection Methods: A Partial Replication and Extension in a German Sample. *Applied Psychology: An International Review*, 52, 515-532.
- Nevo, B. & Sfez, J. (1986). Examinees' Feedback Questionnaire (EFeQ). In B. Nevo & R. S. Jäger (Eds.), *Psychological Testing: The examinee perspective* (pp. 21-47). Göttingen: Hogrefe.
- Ryan, A. M. & Ployhart, R. E. (2000). Applicant perceptions of selection pro-

- cedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26, 565-606.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184-191.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6, 60-70.
- Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Psychologie und Praxis*, 27, 33-44.
- Steck, P. (1997). Psychologische Testverfahren in der Praxis. *Diagnostica*, 267-284.
- Trost, G. (1993). Attitudes and reactions of West German students with respect to scholastic aptitude tests in selection and counseling programs. In B. Nevo & R. S. Jäger (Hrsg.), *Educational and Psychological Testing: The test taker's outlook* (S. 177-201). Göttingen: Hogrefe.

Berichtigung

In Heft 6-2008, Seite 301, veröffentlichten wir die TBS-TK Rezension: »Beck Depressions-Inventar (BDI-II). Revision«. Bei der Aufzählung der Autoren fehlte Dr. Susanne Goldschmidt von der Universität Leipzig. Die Redaktion entschuldigt sich für das Versehen.

Anzeige KORTE

86x50

Anzeige PITS

177 x 127