

Messen Einstellungsinterviews das, was sie messen sollen? Zur Bedeutung der Bewerberkognitionen über bewertetes Verhalten

Klaus G. Melchers, Martin Kleinmann, Gerald M. Richter,
Cornelius J. König und Ute-Christine Klehe

Zusammenfassung. Mit strukturierten Einstellungsinterviews wird häufig versucht, verschiedene Anforderungsdimensionen zu erfassen, die für eine erfolgreiche berufliche Tätigkeit nötig sind. Allerdings ist noch weitgehend ungeklärt, inwieweit es gelingt, die angestrebten Dimensionen tatsächlich zu erfassen. Zur Untersuchung dieser Frage wurde im Rahmen eines Bewerbungs-Trainings ein strukturiertes Interview durchgeführt, das aus drei Komponenten (Selbstvorstellung, biographischen Fragen und situativen Fragen) bestand. Eine Analyse der Multitrait-Multimethod-Matrix ergab, dass das verwendete Interview wenig konstruktvalid war. Allerdings zeigte sich, dass die Gesamtbeurteilung im Interview mit der Anzahl der Fragen korrelierte, für die die Interviewten die jeweilige Anforderungsdimension korrekt erkannten. Zudem wurden Teilnehmerinnen und Teilnehmer auch intraindividuell besser bei Fragen beurteilt, bei denen sie die jeweils relevante Dimension erkannten, als bei Fragen, bei denen dies nicht der Fall war. Es wird argumentiert, dass die Fähigkeit, relevante Anforderungs- und Beurteilungsdimensionen zu erkennen, einen Teil der prognostischen Validität strukturierter Interviews erklären kann. Schlüsselwörter: Interview, Personalauswahl, Konstruktvalidität, Selbstdarstellung

Do selection interviews measure what they intend to measure? The impact of applicants' cognitions about evaluated behavior

Abstract. Often, the objective of structured selection interviews is to assess different performance dimensions that are relevant for successful job performance. It is, however, largely unknown whether the intended dimensions are really measured. To evaluate this question, a structured interview was used for an application training that consisted of three components (self-presentation, biographic questions, and situational questions). Analysis of the multitrait-multimethod matrix showed that the current interview had low construct validity. We found, however, that performance evaluations for the interview correlated with the number of questions for which interviewees had correctly identified the respective dimension. Furthermore, at the individual level, participants also received better evaluations on questions for which they identified the respective dimension than on questions for which they did not. We argue that the ability to identify relevant performance and evaluation criteria can explain part of the predictive validity of structured interviews.

Key words: interview, personnel selection, construct validity, self-presentation

Eine Vielzahl von Metaanalysen hat wiederholt belegt, dass strukturierte Interviews prognostisch valide Instrumente der Berufseignungsdiagnostik sind (z. B.

Klaus G. Melchers und Martin Kleinmann, Universität Zürich, Gerald M. Richter, Philipps-Universität Marburg, Cornelius J. König und Ute-Christine Klehe, Universität Zürich. Gerald M. Richter arbeitet mittlerweile bei Chiron Vaccines, Marburg. Die vorliegende Arbeit wurde durch eine Sachmittelbeihilfe der DFG (Kl 823/6-1) an Martin Kleinmann gefördert. Wir danken Nicole

Huffcutt, Conway, Roth & Klehe, in press; Huffcutt & Woehr, 1999; McDaniel, Whetzel, Schmidt & Maurer, 1994; Taylor & Small, 2002; Wiesner & Cronshaw, 1988). Aus pragmatischer Perspektive lässt sich die Frage, ob sie das messen, was sie messen sollen, damit schnell und einfach beantworten. Da sie eine

Borchert und Birgitta Natale für ihre Unterstützung bei der Durchführung dieser Untersuchung.

gute prognostische Validität erreichen, müssen sie irgendwelche Fähigkeiten oder Eigenschaften messen, die sich in späterem Berufserfolg niederschlagen. Welche Fähigkeiten oder Eigenschaften dies allerdings genau sind, lässt sich damit noch nicht sagen. Und obwohl zahlreiche Forscherinnen und Forscher seit Jahren die Notwendigkeit betonen, diese Frage zu klären (z. B. Harris, 1999; Hunter & Hirsh, 1987; Motowidlo, 1999; Taylor & Small, 2002), haben wir bis heute nur ein eingeschränktes Wissen darüber, was strukturierte Interviews messen. Das Ziel der vorliegenden Arbeit ist, eine Beantwortung dieser Frage voranzutreiben, d. h. der Frage nach der Konstruktvalidität strukturierter Interviews.

Eine Möglichkeit zur Klärung dieser Frage besteht darin zu untersuchen, mit welchen externen Variablen die Beurteilung in einem Interview korreliert. Der erheblich größere Teil unseres Wissens zur Konstruktvalidität von Einstellungsinterviews stammt aus Studien, die diese Herangehensweise gewählt haben. Zwei kürzlich publizierte Metaanalysen haben die entsprechenden Ergebnisse zusammengetragen (Huffcutt, Conway, Roth & Stone, 2001; Salgado & Moscoso, 2002). In der Metaanalyse von Salgado und Moscoso (2002) fielen die Korrelationen zwischen verschiedenen externen Variablen und den Gesamtbeurteilungen je nach Art des Interviews unterschiedlich aus. Konventionelle Interviews korrelierten vor allem mit Maßen für soziale Fertigkeiten, mit Intelligenz und mit verschiedenen Persönlichkeitsdimensionen der Big Five. Verhaltensorientierte Interviews korrelierten demgegenüber vor allem mit Berufserfahrung, sozialen Fertigkeiten, berufsbezogenem Wissen und mit situativer Urteilsfähigkeit. Ob aber hohe oder niedrige Korrelationen positiv im Sinne der Konstruktvalidität zu bewerten sind, ist erst einmal unklar, denn je nach Zielsetzung des Interviews können prinzipiell eine Vielzahl verschiedener Dimensionen bzw. Variablen erfasst werden (vgl. Schuler, 1992). Ein hoher Zusammenhang zwischen der Beurteilung im Interview und Maßen für soziale Fertigkeiten würde zum Beispiel für die Konstruktvalidität eines Interviews sprechen, das tatsächlich soziale Fertigkeiten erfassen soll. Für ein Interview, das demgegenüber Anforderungsdimensionen erfassen soll, die unabhängig von sozialen Fertigkeiten sein sollten (z. B. organisatorische Fertigkeiten, sicherheitsförderliches Arbeitsverhalten etc.), ist ein solcher Zusammenhang unerwartet und deutet eher auf eine fehlende diskriminante Validität hin. Dieses Problem wurde in der Metaanalyse von Salgado und Moscoso jedoch nicht berücksichtigt.

Im Gegensatz dazu konzentrierten Huffcutt et al. (2001) sich in ihrer Metaanalyse auf die Frage, inwiefern Einstellungsinterviews Intelligenz erfassen. Dafür berechneten sie Korrelationen zwischen Beur-

teilungen der Intelligenz in Interviews und der tatsächlichen, mit Hilfe von Intelligenztests bestimmten Intelligenz. Zudem berechneten sie Korrelationen zwischen Beurteilungen anderer Dimensionen und der Test-Intelligenz. Sie fanden, dass die in unstrukturierten Interviews beurteilte Intelligenz bedeutsam mit der Test-Intelligenz korrelierte und dass der gefundene Zusammenhang höher war als die Korrelationen zwischen anderen Interviewdimensionen und der Test-Intelligenz. Für strukturierte Interviews fiel der Zusammenhang zwischen beurteilter und tatsächlicher Intelligenz jedoch insgesamt niedriger aus und war auch nicht höher als die Korrelationen zwischen einigen anderen beurteilten Dimensionen (beruflichem Wissen bzw. Berufserfahrung) und der Test-Intelligenz. Zumindest für strukturierte Interviews scheint es also nicht zu gelingen, Intelligenz tatsächlich valide zu erfassen, selbst wenn diese Dimension konkret beurteilt werden sollte.

Ein weiterer Zugang zur Überprüfung der Konstruktvalidität von Interviews besteht in der Analyse so genannter Multitrait-Multimethod-Matrizen (MTMM-Matrizen). Für solche Matrizen werden verschiedene Dimensionen (die Traits) jeweils auf mehrere Arten (d. h. mit verschiedenen Methoden) erfasst. Damit kann überprüft werden, ob verschiedene Beurteilungen innerhalb eines Interviews hoch miteinander korrelieren, wenn sie tatsächlich die gleiche Anforderungsdimension erfassen sollen, bzw. niedrig miteinander korrelieren, wenn sie unterschiedliche Dimensionen erfassen sollen (einige Autoren nennen diesen Aspekt *interne Konstruktvalidität*, z. B. Kolk, 2001). Ersteres würde auf eine gute konvergente Validität hindeuten, Zweiteres auf eine gute diskriminante Validität.

Obwohl ein solcher Zugang im Rahmen von Studien zur Konstruktvalidität von Assessment Centern (ACs) seit langem gewählt wird (vgl. Kleinmann, 1997a), ist diese Möglichkeit, die Konstruktvalidität von Interviews zu klären, bisher weitgehend übersehen worden. Dies überrascht im ersten Augenblick, da Interviews häufig verwendet werden, um eine ganze Reihe unterschiedlicher Dimensionen zu erfassen (vgl. Huffcutt, Conway et al., 2001). Ein Grund für die geringe Anzahl von Untersuchungen zur internen Konstruktvalidität von Interviews besteht vermutlich darin, dass bei Interviews häufig eine notwendige Voraussetzung für MTMM-Analysen nicht gegeben ist, nämlich dass verschiedene Anforderungsdimensionen tatsächlich jeweils mit mehreren Interview-Komponenten erfasst werden (z. B. mit biographischen und situativen Fragen als zwei unterschiedlichen Komponenten). So gibt es bisher kaum Studien, in denen Interviews verwendet wurden, die aus mehreren Komponenten bestanden (siehe z. B. Campion, Campion & Hudson, 1994; Conway & Penno, 1999;

Huffcutt, Weekley, Wiesner, DeGroot & Jones, 2001; Schuler & Funke, 1989; Schuler & Moser, 1995, für Ausnahmen). Und in lediglich drei dieser Studien wurde die zweite Art der Konstruktvalidierung am Rande mit untersucht (Conway & Peneno, 1999; Huffcutt, Weekley et al., 2001; Schuler & Funke, 1989). Schuler und Funke berichteten (allerdings ohne Angabe der berechneten MTMM-Matrix), dass Beurteilungen von Fragen, die in zwei verschiedenen Interview-Komponenten jeweils die gleiche Dimension messen sollten, miteinander nicht höher korrelierten als Beurteilungen von Fragen, die verschiedene Dimensionen messen sollten. Zu einem vergleichbaren Ergebnis kamen auch Huffcutt, Weekley et al. sowie Conway und Peneno. Dies lässt daran zweifeln, dass in strukturierten Interviews verschiedene Anforderungsdimensionen tatsächlich valide erfasst werden können.

Einschränkend muss allerdings angemerkt werden, dass in den angesprochenen Studien (Conway & Peneno, 1999; Huffcutt, Weekley et al., 2001; Schuler & Funke, 1989) für jede zu erfassende Dimension nur jeweils eine einzige Frage pro Interview-Komponente verwendet wurde. Dadurch erscheint es schwer möglich, die unterschiedlichen Facetten einer Dimension tatsächlich umfassend und reliabel zu messen. So korrelierten bei Huffcutt, Weekley et al. (2001) alle Items relativ niedrig miteinander – in ihrer Studie 2 zum Beispiel betrug die Korrelation im Durchschnitt .05 für Items, die identische Dimensionen erfassen sollten, und .09 für Items, die unterschiedliche Dimensionen erfassen sollten. Bei Conway und Peneno dagegen korrelierten alle Items hoch miteinander (im Durchschnitt zu .50 für Items, die identische Dimensionen erfassen sollten, und zu .48 für Items, die unterschiedliche Dimensionen erfassen sollten), was daran zweifeln lässt, ob tatsächlich unterschiedliche Dimensionen erfasst wurden. Bei Schuler und Funke waren die durchschnittlichen Interkorrelationen für die einzelnen Interview-Komponenten (und damit deren jeweilige interne Konsistenzen) zum Teil relativ niedrig. Im Hinblick darauf, dass die verschiedenen Fragen tatsächlich unterschiedliche Dimensionen erfassen sollten, ist dies erstrebenswert (vgl. auch Conway, Jako & Goodman, 1995). Allerdings lässt dies trotzdem noch keine Aussage darüber zu, inwieweit tatsächlich die intendierten Dimensionen oder irgendwelche anderen Aspekte beurteilt wurden. Hierzu wäre es wünschenswert, innerhalb der Interview-Komponenten jeweils mehrere Fragen pro Dimension zu haben. Wenn diese Fragen die gleiche Dimension erfassen, sollten ihre mittleren Interkorrelationen höher ausfallen als für die ganze Komponente. Damit würde auch die Reliabilität für die Erfassung der Dimensionen gesteigert.

Mit diesen ersten Hinweisen zur Überprüfung der internen Konstruktvalidität von Einstellungsinterviews und den oben berichteten Ergebnissen der Metaanalyse von Huffcutt, Conway et al. (2001) scheint die Interviewforschung insgesamt vor einem vergleichbaren Validitäts-Paradoxon zu stehen wie die AC-Forschung vor einiger Zeit (Kleinmann, 1997a; Klimoski & Brickner, 1987). Einerseits sind strukturierte Einstellungsinterviews zwar prognostisch valide, andererseits scheinen sie jedoch nicht geeignet zu sein, jene Konstrukte oder Dimensionen tatsächlich zu erfassen, die sie eigentlich erfassen sollen.

In der vorliegenden Studie war unser erstes Ziel, die Konstruktvalidität eines strukturierten Interviews abzuschätzen, bei dem die verschiedenen Dimensionen in jeder Interview-Komponente mit mehreren Fragen gemessen werden. Zum einen wird es dadurch möglich, die mittleren Interkorrelationen für eine Interview-Komponente als Ganzes mit denen für die einzelnen Dimensionen zu vergleichen. Für die einzelnen Dimensionen sollten diese Interkorrelationen höher ausfallen. Zudem sollten die verschiedenen Dimensionen innerhalb jeder Interview-Komponente dadurch reliabler gemessen werden, was zu einer Verbesserung der konvergenten und der diskriminanten Validität beitragen könnte.

Die zweite Frage unserer Studie war, ob neben (oder anstelle) der intendierten Dimensionen die Fähigkeit erfasst wird, relevante Anforderungen oder Kriterien zu erkennen. Auch diese Idee greift Ergebnisse der AC-Forschung auf (allerdings gibt es ähnliche Überlegungen auch bei anderen Personalauswahlverfahren; vgl. Marcus, 2003a, 2003b). Der Grundgedanke hierbei ist, dass Bewerberinnen und Bewerber versuchen, sich in einer Auswahl-situation positiv darzustellen, um damit ihre Chancen auf ein Stellenangebot zu erhöhen. Marcus (2003b) unterscheidet in diesem Zusammenhang vier Aspekte, die für eine positive Selbstdarstellung relevant sind: Erstens die rezeptive Selbstdarstellungskompetenz, bei der es darum geht, die Anforderungen der Situation möglichst genau zu erkennen, zweitens die Selbstdarstellungsperformanz, d. h. die Fähigkeit, sich tatsächlich den Anforderungen entsprechend zu verhalten, drittens die Selbstdarstellungsmotivation, also die Bereitschaft, sein Verhalten den vermuteten Erwartungen entsprechend anzupassen, und viertens die wahre Ausprägung des jeweils erfassten Konstrukts. Wenn man wie Marcus davon ausgeht, dass die wahre Ausprägung des erfassten Konstrukts unabhängig von der Selbstdarstellungstendenz ist, ist vor allem der erste Aspekt für die Beurteilung im Interview von grundlegender Bedeutung. Bewerberinnen und Bewerber, die nicht in der Lage sind, die Anforderungen in eben dieser Bewerbungssituation richtig zu erkennen, kön-

nen ihr Verhalten trotz hoher Motivation zur positiven Selbstdarstellung gar nicht erst entsprechend ausrichten.

Für ACs konnte Kleinmann (1993, 1997a) zeigen, dass die Anzahl richtig erkannter Dimensionen bzw. Kriterien mit der Gesamtleistung von Teilnehmerinnen und Teilnehmern korrelierte. Darüber hinaus wurden diese auch intraindividuell auf den AC-Dimensionen besser beurteilt, die sie korrekt erkannt hatten, als auf den Dimensionen, die sie nicht erkannt hatten. Die Ausprägung der rezeptiven Selbstdarstellungskompetenz hatte also Einfluss auf das Abschneiden im AC.

Da ACs und Interviews für Bewerberinnen und Bewerber insofern vergleichbar sind, dass bei beiden Verfahren in der Regel intransparent ist, welche Kriterien jeweils mit einer AC-Übung oder mit einer Interview-Frage erfasst werden sollen, erscheint es plausibel, dass sich auch für Interviews vergleichbare Zusammenhänge finden lassen zwischen dem Abschneiden und der Fähigkeit, die jeweils relevanten Anforderungsdimensionen und Kriterien zu erkennen (als englische Bezeichnung schlagen wir hierfür die Formulierung *ability to identify criteria* vor). In unserer Studie wollten wir erstmals überprüfen, wie bedeutsam der Einfluss dieser Erkennensfähigkeit auf die Beurteilung in strukturierten Einstellungsinterviews ist. In Anlehnung an Kleinmann (1993, 1997a) sollte dies auf zwei Arten geschehen: zum einen über die Korrelation zwischen der Anzahl korrekt erkannter Anforderungsdimensionen und der Gesamtbeurteilung und zum anderen über den intraindividuellen Vergleich von Antwort-Beurteilungen auf Fragen, für die die jeweilige Anforderungsdimension korrekt erkannt wurde, versus solchen, für die dies nicht der Fall war.

Methode

Stichprobe

Die vorliegende Studie wurde im Rahmen eines Bewerbungs-Trainings durchgeführt. Dieses Training richtete sich in erster Linie an Personen, die kurz vor oder nach Abschluss ihres Studiums standen und die sich mit Hilfe des Trainings auf Einstellungsinterviews vorbereiten wollten. Die Untersuchungstichprobe bestand aus 64 Personen (35 Frauen und 29 Männer), die zwischen 18 und 47 Jahren alt waren ($Md = 26$). Die größte Gruppe kam aus dem Bereich der Rechts- und Wirtschaftswissenschaften (39 Prozent) und die zweitgrößte aus den Naturwissenschaften (14 Prozent). Mehr als 45 Prozent verfügten bereits über einen Studienabschluss. Für die Teilnahme

am Training wurde ein kleiner Unkostenbeitrag erhoben, der auch dazu dienen sollte, die Motivation der Teilnehmerinnen und Teilnehmer zu erhöhen.

Die 12 Interviewer und 13 Interviewerinnen waren überwiegend Psychologie-Studierende mit dem Schwerpunkt Arbeits- und Organisationspsychologie, die mit der Teilnahme an der Studie ihre fachliche Qualifikation erhöhen wollten und speziell geschult wurden. Sie erhielten während der gesamten Datenerhebung keine Informationen über die der Studie zugrunde liegenden Untersuchungsfragen.

Verwendetes Interview

Das von uns verwendete Interview wurde in Zusammenarbeit mit Praktikern aus dem Bereich der Personalauswahl entwickelt. Hierbei wurde die Zielposition „Management-Trainee“ gewählt, um Teilnehmerinnen und Teilnehmern aus unterschiedlichen Studienfächern bzw. Bereichen eine möglichst realistische Bewerbungssituation zu ermöglichen. Entsprechend wurden in einem ersten Schritt typische Alltagssituationen aus dem Arbeitsleben eines Management-Trainees gesammelt. Aus diesen Situationen wurden anschließend von Experten Verhaltensdimensionen abgeleitet, die zu ihrer erfolgreichen Bewältigung jeweils notwendig sind. In einem dritten Schritt wurde in einem Workshop ermittelt, wie gut die ermittelten Dimensionen im Rahmen eines Interviews erfassbar sind. Die Teilnehmer dieses Workshops waren Studierende mit Schwerpunkt Arbeits- und Organisationspsychologie und mit Erfahrung als AC-Beobachter bzw. Interviewer. Vier der gefundenen Dimensionen wurden als gut erfassbar eingeschätzt. Bei ihnen handelte es sich um Arbeitsorganisation/Planung, Führungsverhalten, Informationsverhalten und Kooperation. Arbeitsorganisation/Planung war dabei zum Beispiel definiert als Organisation des eigenen Arbeitsalltags nach bestimmten Kriterien, Setzung von Prioritäten, Einhaltung von Terminen sowie sorgfältiges, zuverlässiges und zielgerichtetes Arbeiten. Dimensionen mit ähnlichem Bedeutungsgehalt werden in der Praxis häufig verwendet (vgl. Huffcutt, Conway et al., 2001; Jeserich, 1990).

Als nächstes wurden Interviewfragen formuliert, mit denen die vier Dimensionen gemessen werden sollten. Diese Interviewfragen wurden zum Teil aus den ermittelten Situationen abgeleitet und zum Teil aus früheren Interview-Studien adaptiert, in denen vergleichbare Dimensionen erfasst werden sollten. Acht weitere Studierende, die mit jenen des ersten Workshops vergleichbar waren, ordneten die verschiedenen Interviewfragen anschließend den Dimensionen zu. Nur Items, die nach diesem Prozedere ein-

deutig der jeweiligen Dimension zugeordnet werden konnten, wurden für das Interview verwendet.

Das eigentliche Interview bestand aus drei Komponenten: Erstens einer Selbstvorstellung der Bewerberinnen und Bewerber, zweitens einem biographischen Interview und drittens einem situativen Interview. Diese Komponenten sind in ähnlicher Form auch im Multimodalen Interview von Schuler enthalten, wo sie sich in der Praxis bereits gut bewährt haben (vgl. Schuler, 1992, und Schuler & Moser, 1995).

Die Selbstvorstellung ist ein typischer Bestandteil vieler Einstellungsinterviews und wird auch in ACs häufig verwendet. In unserem Interview wurden die Teilnehmerinnen und Teilnehmer gebeten zu berichten, „was sie bislang gemacht haben, was sie gerne tun, wo ihre Stärken und Schwächen liegen und was sie erreichen möchten“. Sie hatten dafür 5 Minuten Zeit. Bei der Selbstvorstellung wurden alle vier Dimensionen beurteilt.

Biographische und situative Interviews sind ebenfalls häufig verwendete Bestandteile von Einstellungsinterviews, auch wenn sie eher selten gemeinsam in einem Interview verwendet werden. Bei beiden handelt es sich um hoch strukturierte Interviews, für die in Metaanalysen jeweils gute prognostische Validitäten gefunden wurden (Huffcutt et al., in press; Taylor & Small, 2002).

Beim biographischen Interview werden verschiedene Fragen gestellt, die sich auf Situationen und Verhalten der Bewerber in der Vergangenheit beziehen. Der zentrale Gedanke bei biographischen Interviews ist, dass Informationen über vergangenes Verhalten erfragt werden, da dieses vergangene Verhalten ein guter Prädiktor für künftiges Verhalten sein sollte (Janz, 1989). Im Rahmen des situativen Interviews werden dagegen hypothetische, auf die Zielposition bezogene Situationen dargestellt, und Bewerber sollen angeben, wie sie sich in diesen Situationen verhalten würden (Latham, Saari, Pursell & Champion, 1980). Sowohl das biographische als auch das situative Interview nahmen etwa 15 bis 20 Minuten in Anspruch. Jede Interview-Frage dieser beiden Komponenten erfasste eine einzige Dimension, und jede der vier Dimensionen sollte mit jeweils drei Fragen erfasst werden.

Für alle drei verwendeten Komponenten des Interviews wurden die Beurteilungen auf einer Skala von 1 bis 5 erfasst, wobei höhere Werte bessere Beurteilungen bedeuteten. In dem erwähnten Workshop wurden für die Selbstvorstellung für alle Skalenpunkte Beurteilungsanker entwickelt und für die beiden anderen Komponenten für die Skalenpunkte 1, 3 und 5.

Untersuchungsablauf

Interviewer-Schulung. Die Interviewerinnen und Interviewer wurden in drei Gruppen jeweils einen halben Tag lang geschult. Dabei wurden sie mit den verwendeten Komponenten des Interviews und mit den erfassten Dimensionen vertraut gemacht, und sie bekamen Informationen über typische Beurteilungsfehler. Besonderer Wert wurde auf ein einheitliches Verständnis der verwendeten Verhaltensanker und der mit ihnen verbundenen Beurteilungen gelegt. Da sich dieser Aspekt in früheren Beurteiler-Trainings als besonders bedeutsam für die Beurteilungs-Akkuratheit erwiesen hat (Woehr & Huffcutt, 1994), wurde die Beurteilung vor diesem Hintergrund auch praktisch geübt. Zudem wurden die Interviewer darin geschult, Feedbackgespräche mit den Teilnehmern zu führen.

Bewerbungs-Training. Die Teilnehmerinnen und Teilnehmer bekamen zunächst in einer etwa einstündigen Einführungsveranstaltung das Konzept des Trainings erläutert. Zudem bekamen sie eine fiktive Stellenanzeige für die Zielposition, auf die sie ihre schriftliche Bewerbung ausrichten sollten und an der sich auch das Interview orientierte. Diese Stellenanzeige enthielt Hinweise auf die vier Dimensionen, die im Interview erfasst werden sollten (z. B. für die Dimension „Arbeitsorganisation/Planung“: „Sie ... arbeiten generell zuverlässig und zielgerichtet“). Das eigentliche Training fand in der Regel eine Woche später statt und dauerte insgesamt 4 Stunden. Das Interview wurde direkt nach einer kurzen Begrüßung einzeln mit jedem Teilnehmer bzw. jeder Teilnehmerin durchgeführt. Jedes Interview wurde von jeweils zwei Interviewern bzw. Interviewerinnen gemeinsam geführt.

In Anlehnung an Kleinmann (1993) bekamen die Teilnehmerinnen und Teilnehmer im Anschluss an jede der drei Interview-Komponenten einen Nachbearbeitungsbogen, in dem die einzelnen ihnen zuvor gestellten Fragen nochmals aufgeführt waren. Ihnen wurde gesagt, dass ihre Antworten in diesem Bogen dazu dienen sollten, das Interview zu verbessern. Die Teilnehmerinnen und Teilnehmer sollten für jede dieser Fragen mögliche Hypothesen angeben, worauf es bei der jeweiligen Frage ihrer Meinung nach ankam. Pro Frage konnten mehrere Hypothesen notiert werden. Zudem sollten die Teilnehmerinnen und Teilnehmer jene Hypothesen ankreuzen, die für ihr Antwortverhalten tatsächlich relevant waren. Diese verhaltensrelevanten Hypothesen gingen in die späteren Auswertungen ein.

Nach dem Ende des gesamten Interviews besprachen die Interviewerinnen bzw. Interviewer ihre Beurteilungen und bereiteten das Feedback vor. Bei dieser Besprechung sollten sie insbesondere solche Bewer-

tungen diskutieren, bei denen Unterschiede von zwei oder mehr Punkten (auf der 5-stufigen Beurteilungsskala) vorlagen. Unterschiedliche Beurteilungen konnten meist nach kurzer Diskussion geklärt werden, so dass die Urteile in der Regel nur noch maximal einen Punkt auseinander lagen. Eine Einigung musste jedoch nicht erfolgen. Die endgültige Leistungsbewertung der Teilnehmerinnen und Teilnehmer erfolgte durch Mittelwertsbildung der beiden Interviewer-Urteile.

Während der Interviewer-Besprechung wurden die Teilnehmerinnen und Teilnehmer in einem anderen Raum mit sechs verschiedenen Dimensionen vertraut gemacht, die im Interview potenziell hätten beurteilt werden können. Unter diesen Dimensionen waren die vier, die tatsächlich erfasst werden sollten, sowie zwei weitere (Ausdrucksvermögen und Selbstvertrauen), die als Distraktoren dienten. Diese Distraktoren stammten ebenfalls aus dem Workshop. Die Teilnehmerinnen und Teilnehmer bekamen eine Liste mit Verhaltensbeispielen für die verschiedenen Dimensionen ausgeteilt und sie erhielten die ausgefüllten Nachbefragungsbögen mit ihren notierten Hypothesen zurück. Sie sollten nun jede ihrer Hypothesen einer der Dimensionen zuordnen bzw. angeben, wenn die Hypothese keiner der Dimensionen entsprach. Anschließend bekamen die Teilnehmerinnen und Teilnehmer Informationen zu weiteren Personalauswahlverfahren und allgemeine Tipps für Bewerbungen. Danach bekamen sie ein ausführliches Feedback über ihre Leistungen und zu ihren Bewerbungsunterlagen.

Ergebnisse

Reliabilität des Interviews

Wir ermittelten drei verschiedene Schätzer für die Reliabilität des verwendeten Interviews: Zum einen die mittlere Interviewer-Übereinstimmung, also die Korrelation der Gesamtwerte beider Interviewer über alle 28 Fragen. Sie betrug .74 vor der Beurteiler-Konferenz und .85 danach. Für die Einzelkomponenten betragen die Werte vor der Konferenz zwischen .71 und .81 und danach zwischen .81 und .89. Diese Ergebnisse sind vergleichbar mit metaanalytisch bestimmten Werten (Conway et al., 1995; Taylor & Small, 2002).

Zum anderen berechneten wir die internen Konsistenzen (Cronbach's α) und die mittleren Interkorrelationen der Fragen für die einzelnen Interview-Komponenten und für das Gesamtinterview. Grundlage hierfür und für alle nachfolgenden Analysen waren die Beurteilungen nach der Interviewer-Besprechung. Die entsprechenden Ergebnisse sind in Tabelle 1 an-

gegeben. Diese sind einerseits für die verschiedenen Komponenten als Ganzes angegeben, andererseits nochmals aufgeschlüsselt für die einzelnen Dimensionen des biographischen und des situativen Interviews. Da Cronbach's α von der Anzahl der Fragen abhängt, sind zusätzlich die durchschnittlichen Interkorrelationen der Beurteilungen angegeben. Für die Interpretation ist zu beachten, dass hohe Konsistenzen bzw. Interkorrelationen innerhalb einer Komponente als Ganzes kritisch sind. Sie würden eher darauf hindeuten, dass die Fragen vergleichbar beurteilt werden, obwohl sie unabhängige Dimensionen erfassen sollten (Conway et al., 1995). Für die einzelnen Dimensionen hingegen sollten die Konsistenzen und die Interkorrelationen möglichst hoch ausfallen.

Die interne Konsistenz und die mittlere Korrelation sind für die Selbstvorstellung mit .78 bzw. mit .47 deutlich höher als für die anderen Komponenten, was auf Probleme hindeutet, ob mit dieser Komponente tatsächlich unabhängige Dimensionen erfasst wurden. Für das biographische und das situative Interview fällt auf, dass die mittleren Korrelationen für die einzelnen Dimensionen jeweils höher ausfallen als für die jeweilige Gesamt-Komponente (in beiden Fällen mit Ausnahme von Informationsverhalten), was durchaus im Sinne einer besseren Konstruktvalidität ist. Die Fragen zum Informationsverhalten wiesen in beiden Komponenten keine interne Konsistenz auf.

Die ermittelten internen Konsistenzen und Interkorrelationen sind für die Selbstvorstellung und für

Tabelle 1. Interne Konsistenzen und mittlere Interkorrelationen der Fragen für die verschiedenen Interview-Komponenten und für die einzelnen Dimensionen im Biographischen und im Situativen Interview ($N = 64$).

Komponente	Anzahl Fragen	Cronbach's α	mittlere Korrelation
Selbstvorstellung	4	.78	.47
Biographisches Interview	12	.34	.05
Situatives Interview	12	.69	.15
Gesamt	28	.70	.08
Einzeldimensionen Biographisches Interview			
Arbeitsorganisation/ Planung	3	.31	.13
Führungsverhalten	3	.33	.17
Informationsverhalten	3	-.04	-.01
Kooperation	3	.29	.11
Einzeldimensionen Situatives Interview			
Arbeitsorganisation/ Planung	3	.41	.18
Führungsverhalten	3	.53	.28
Informationsverhalten	3	-.09	-.03
Kooperation	3	.50	.26

das situative Interview als Ganzes in etwa vergleichbar mit den Werten aus der Metaanalyse von Conway et al. (1995), die Werte für das biographische Interview als Ganzes sind etwas niedriger. Letztere sind allerdings vergleichbar mit den Ergebnissen der Untersuchung von Schuler und Funke (1989).

Konstruktvalidität

Für die Analyse der MTMM-Matrix wurden zuerst pro Interview-Komponente Durchschnittswerte für die einzelnen Dimensionen berechnet. Für die Selbstvorstellung war dies der Mittelwert der beiden Beurteiler und für das biographische und das situative Interview der Durchschnitt der mittleren Urteile für die drei Fragen pro Dimension. Die Durchschnittswerte und die zu ihnen gehörenden Standardabweichungen sind in Tabelle 2 zu finden. Dort ist zudem die ermittelte MTMM-Matrix angegeben.

Hohe Korrelationen der Werte, die die gleichen Dimensionen messen sollten, aber mit verschiedenen Methoden (also den drei Interview-Komponenten) ermittelt wurden, deuten auf eine gute konvergente Validität hin. Idealerweise sollten diese monotrait-heteromethod-Korrelationen deswegen möglichst groß sein. Demgegenüber sollten Korrelationen der Werte, die unterschiedliche Dimensionen mit ein und derselben Methode erfassen, möglichst gering ausfallen, da sie Indikatoren für die diskriminante Validität sind. Zudem sollten diese heterotrait-monomethod-Korrelationen kleiner sein als die monotrait-heteromethod-

Korrelationen. Und schließlich sollten auch die heterotrait-heteromethod-Korrelationen, die ebenfalls Indikatoren der diskriminanten Validität sind, möglichst gering ausfallen und ebenfalls kleiner sein als die monotrait-heteromethod-Korrelationen.

In Tabelle 2 ist zu sehen, dass nur 5 der 12 ermittelten monotrait-heteromethod-Koeffizienten signifikant waren. Der Mittelwert aller 12 Koeffizienten (berechnet mit Hilfe der Fisher Z-Transformation) betrug .20. Dies deutet darauf hin, dass die konvergente Validität unseres Interviews gering ist. Im Vergleich dazu fielen die heterotrait-monomethod-Koeffizienten im Durchschnitt höher aus (Mittelwert $r = .30$), und 12 der 18 Werte waren signifikant. Dies deutet auf eine fehlende diskriminante Validität hin. Lediglich für die heterotrait-heteromethod-Koeffizienten konnten wir einen niedrigen Mittelwert von $r = .10$ ermitteln, der widerspiegelt, dass lediglich 4 der insgesamt 36 Koeffizienten signifikant wurden.

Aufgrund der nicht vorhandenen internen Konsistenz der Dimension Informationsverhalten im biographischen und im situativen Interview wurde die Auswertung der MTMM-Matrix nochmals unter Ausschluss dieser Dimension wiederholt. Dabei ergab sich eine mittlere monotrait-heteromethod-Korrelation von $r = .22$, eine mittlere heterotrait-monomethod-Korrelation von $r = .28$ und eine mittlere heterotrait-heteromethod-Korrelation von $r = .06$. Durch Weglassen der nicht reliabel erfassten Dimension Informationsverhalten verändert sich das qualitative Ergebnismuster also nicht.

Tabelle 2. Multitrait-Multimethod-Matrix für das Interview ($N = 64$).

	<i>M</i>	<i>SD</i>	Selbstvorstellung				Biographisches Interview				Situatives Interview			
			AP	FV	IV	KO	AP	FV	IV	KO	AP	FV	IV	KO
Selbstvorstellung														
AP	3.46	.94	–											
FV	2.74	1.07	.49**	–										
IV	3.11	.93	.51**	.44**	–									
KO	3.34	.88	.45**	.49**	.45**	–								
Biographisches Interview														
AP	3.73	.63	.40**	.11	.27*	.24	–							
FV	3.29	.80	.22	.40**	.09	.29*	.08	–						
IV	3.81	.50	.18	.10	.28*	.15	.20	.25*	–					
KO	3.70	.65	–.01	–.03	.17	.14	.04	–.16	.12	–				
Situatives Interview														
AP	3.34	.65	.38**	–.02	.25*	.11	.37**	–.01	.23	.16	–			
FV	3.26	.82	.25*	–.05	.22	.15	–.11	.10	.16	.23	.32**	–		
IV	3.36	.51	.19	.07	.06	.09	–.01	.07	.11	.12	.29*	.24	–	
KO	3.53	.77	–.06	–.21	–.02	–.05	.01	–.15	.07	.24	.35**	.39**	.40**	–

Anmerkungen: AP = Arbeitsorganisation/Planung, FV = Führungsverhalten, IV = Informationsverhalten, KO = Kooperation.
* $p < .05$, ** $p < .01$ (2-seitig).

Auf Anregung eines Gutachters berechneten wir schließlich noch eine MTMM-Matrix mit Korrelationen zwischen einzelnen Fragen. Dabei ergab sich eine mittlere monotrait-heteromethod-Korrelation von $r = .11$, eine mittlere heterotrait-monomethod-Korrelation von $r = .33$ und eine mittlere heterotrait-heteromethod-Korrelation von $r = .06$. Im Vergleich zur oben beschriebenen Analyse zeigen diese Ergebnisse, dass die Verwendung mehrerer Fragen pro Dimension sich zumindest auf die konvergente Validität positiv auswirkt.

Erfassung der Fähigkeit, relevante Beurteilungsdimensionen zu erkennen

Für das Gesamtinterview erkannten die Teilnehmerinnen und Teilnehmer im Durchschnitt in 15.9 von 28 möglichen Fällen die erfasste Dimension richtig und gaben sie als handlungsrelevant an (Spannbreite: 5 bis 23). Betrachtet man die einzelnen Komponenten, erkannten sie im Durchschnitt in der Selbstvorstellung 1.6 von 4 Dimensionen (Spannbreite: 0 bis 4), im biographischen Interview 7.6 von 12 (Spannbreite: 0 bis 11) und im situativen Interview 6.7 von 12 (Spannbreite: 1 bis 10).

Im nächsten Schritt überprüften wir, wie reliabel die Erfassung der Erkennensfähigkeit war. Dazu wurde für die 4 Beurteilungen der Selbstvorstellung und für die jeweils 12 Fragen des biographischen und des situativen Interviews jeweils kodiert, ob die entsprechende Beurteilungsdimension richtig erkannt und als verhaltensrelevant angegeben wurde (kodiert mit 1) oder nicht (kodiert mit 0). Für die so ermittelten Werte wurden interne Konsistenzen (Kuder-Richardson-20) berechnet. Für die Selbstvorstellung fanden wir eine Konsistenz von .25, für das biographische Interview von .60 und für das situative Interview von .49. Für das Gesamtinterview mit allen 28 Beurteilungen ergab sich ein Wert von .59. Der vergleichsweise niedrige Wert für die Selbstvorstellung könnte durch die deutlich geringere Item-Zahl her-

rühren. Die übrigen Werte dagegen fielen höher aus, so dass aussagekräftige statistische Analysen möglich waren.

Bedeutung der Fähigkeit, relevante Beurteilungsdimensionen zu erkennen

Wie oben erwähnt, lässt sich die Bedeutung der Fähigkeit, relevante Beurteilungsdimensionen zu erkennen, auf zwei verschiedene Arten untersuchen. Zum einen kann man interindividuell überprüfen, wie hoch die Korrelation der Anzahl der richtig erkannten Dimensionen mit der Gesamtbeurteilung im Interview bzw. in den verschiedenen Komponenten ausfällt. Zum anderen kann man prüfen, ob die Antworten der Interview-Teilnehmerinnen und -Teilnehmer besser bewertet werden, wenn sie für eine Frage richtig erkannt hatten, welche Dimension erfasst werden sollte, als wenn sie dies nicht erkannt hatten.

Da die Anzahl richtig erkannter Beurteilungsdimensionen für die einzelnen Interview-Komponenten nicht normalverteilt war, berechneten wir Rangkorrelationen. Wir fanden signifikante Korrelationen zwischen der Anzahl der richtig erkannten Dimensionen und der Beurteilung der Teilnehmerinnen und Teilnehmer für das biographische Interview, $r = .29$, $p < .05$, und für das situative Interview, $r = .30$, $p < .05$. Lediglich für die Selbstvorstellung war kein systematischer Zusammenhang nachweisbar, $r = .05$, $p > .70$. Im Hinblick auf das Gesamtinterview (Durchschnittswert aller Beurteilungen) fanden wir ebenfalls eine signifikante Korrelation, $r = .27$, $p < .05$ ($N = 64$ für alle Korrelationen).

Im Hinblick auf die zweite mögliche Auswertung berechneten wir für alle Teilnehmerinnen und Teilnehmer ihre durchschnittliche Beurteilung bei Fragen, bei denen sie die Dimension erkannt hatten, und ihre durchschnittliche Beurteilung bei Fragen, bei denen dies nicht der Fall war. Die durchschnittlichen Beurteilungen für die einzelnen Interview-Kompo-

Tabelle 3. Mittlere Beurteilungen der Interview-Teilnehmerinnen und -Teilnehmer bei Fragen, bei denen sie die erfasste Dimension richtig erkannt hatten, und bei Fragen, bei denen dies nicht der Fall war.

Komponente	richtig erkannt	nicht richtig erkannt	<i>t</i> -Werte	<i>d</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Selbstvorstellung	3.32 (.87)	3.17 (.83)	$t(51) = 1.30$.18
Biographisches Interview	3.74 (.39)	3.45 (.62)	$t(62) = 3.30^{**}$.42
Situatives Interview	3.46 (.62)	3.26 (.57)	$t(63) = 2.49^*$.31
Gesamt	3.58 (.36)	3.28 (.44)	$t(63) = 5.53^{**}$.69

Anmerkungen: *d*-Werte wurden so berechnet, wie von Cohen (1988, pp. 48–49) für abhängige Daten vorgeschlagen. * $p < .05$, ** $p < .01$ (2-seitig).

nen und für das Gesamtinterview sind in Tabelle 3 angegeben. Die unterschiedlichen *N*s ergaben sich dabei dadurch, dass für die Selbstvorstellung bzw. für das biographische Interview einige Personen entweder keine der Dimensionen richtig erkannten oder alle, so dass jeweils einer der beiden benötigten Mittelwerte nicht berechnet werden konnte.

Deskriptiv wurden die Teilnehmerinnen und Teilnehmer in allen Komponenten besser bewertet, wenn sie die erfasste Dimension richtig erkannt hatten, als wenn sie diese nicht erkannt hatten. *t*-Tests für abhängige Daten ergaben signifikante Unterschiede für das biographische Interview, für das situative Interview und für das Gesamtinterview. Der Unterschied für das Gesamtinterview wies mit $d = .69$ dabei einen mittleren bis großen Effekt auf. Lediglich für die Selbstvorstellung wurde der Unterschied nicht signifikant.

Diskussion

Im Hinblick auf die Frage, ob Einstellungsinterviews tatsächlich die Dimensionen messen, die sie messen sollen, sprechen die Ergebnisse der MTMM-Analyse eher gegen diese Facette der Konstruktvalidität von Interviews. Die monotrait-heteromethod-Koeffizienten fielen für eine befriedigende konvergente Validität zu niedrig aus. Noch kritischer ist, dass die heterotrait-monomethod-Koeffizienten im Vergleich sogar höher anstatt niedriger waren. Nach Campbell und Fiske (1959) spricht dies für eine fehlende diskriminante Validität. Das Ergebnismuster ändert sich auch dann nicht bedeutsam, wenn die nicht reliabel erfasste Dimension Informationsverhalten für die Analysen ausgeschlossen wird. Im Vergleich zu Hinweisen aus früheren Untersuchungen (Conway & Peneno, 1999; Huffcutt, Weekley et al., 2001; Schuler & Funke, 1989) bedeutet dies, dass selbst bei Verwendung mehrerer Fragen pro zu erfassender Dimension nicht davon ausgegangen werden kann, dass die intendierten Dimensionen im Interview tatsächlich valide erfasst werden. Dieses Ergebnis ist vergleichbar mit der Befundlage, wie wir sie aus der AC-Forschung kennen (Kleinmann, 1997a; Klimoski & Brickner, 1987).

Nach unseren Ergebnissen korreliert die Fähigkeit, relevante Beurteilungsdimensionen zu erkennen, positiv mit der Leistungsbeurteilung durch die Interviewerinnen und Interviewer. Dieses Ergebnis ist ebenfalls vergleichbar mit Befunden aus der AC-Forschung (Kleinmann, 1993).

Die gefundenen Zusammenhänge zwischen der Erkennensfähigkeit und der Beurteilung im biographischen und im situativen Interview eröffnen die Möglichkeit, dass genau diese Erkennensfähigkeit zumindest einen Teil der prognostischen Validität von

strukturierten Interviews erklären kann. Genauso, wie es für Bewerber in einem Vorstellungsgespräch oftmals nicht transparent ist, nach welchen Kriterien sie in diesem Gespräch beurteilt werden, ist es für Stelleninhaber oftmals ebenfalls nicht transparent, von welchen Kriterien ihre Leistungsbeurteilung und damit ihr Berufserfolg genau abhängt. Wenn sich Personen jedoch darin unterscheiden, diese Kriterien zu erkennen, dann könnte diese Fähigkeit gemeinsame Varianz im Prädiktor (dem Einstellungsinterview) und im Kriterium (dem Berufserfolg) erklären (Kleinmann, 1993; Marcus, 2003a, 2003b). Für ACs gibt es Daten, die diesen Zusammenhang bereits stützen (Kleinmann, 1997b). Zudem würde eine solche Erklärung auch gut zum Befund von Salgado und Moscoso (2002) passen, dass die Gesamtbeurteilung in verhaltensorientierten Interviews bedeutsam mit externen Maßen für soziale Fertigkeiten oder situative Urteilsfähigkeit korreliert. Sowohl das biographische als auch das situative Interview gehörten in der Metaanalyse von Salgado und Moscoso zu diesen verhaltensorientierten Interviews.

Lediglich für die Selbstvorstellung war der Zusammenhang zwischen Erkennensfähigkeit und der Beurteilung weniger stark. Dies könnte allerdings damit zusammenhängen, dass die Erkennensfähigkeit für diese Komponente aufgrund der geringen Item-Zahl nicht so reliabel erfasst werden konnte wie für die anderen Komponenten. Zudem deuten auch die relativ hohen Korrelationen der Beurteilungen und die hohe interne Konsistenz der Selbstvorstellung darauf hin, dass dort eher ein gemeinsamer Faktor (z.B. im Sinne eines Halo-Effekts) beurteilt wurde. Wie bereits erwähnt, betrachten auch andere Autoren hohe interne Konsistenzen in solch einem Fall eher als Hinweis für fehlende diskriminante Validität (vgl. Conway et al., 1995). Dies macht es als Konsequenz schwieriger, differentielle Effekte der Erkennensfähigkeit auf die einzelnen Beurteilungen nachzuweisen.

Einschränkungen und Ansätze für die weitere Forschung

Im Hinblick auf die Befunde zur internen Konstruktvalidität könnte vor allem eingewendet werden, dass diese unter anderem deswegen so negativ ausgefallen sind, weil die internen Konsistenzen für die Fragen zu den einzelnen Dimensionen relativ niedrig waren. Zwar fielen die mittleren Korrelationen (mit Ausnahme von „Informationsverhalten“) für die einzelnen Dimensionen jeweils höher aus als für das biographische oder situative Interview als Ganzes. Trotzdem waren die ermittelten Werte zumindest für das biographische Interview noch nicht befriedigend. Allerdings

waren die Konsistenzen für das verwendete Interview als Ganzes bzw. für die einzelnen Komponenten durchaus vergleichbar mit metaanalytischen Befunden (Conway et al., 1995) und damit mit einer Vielzahl anderer Interviews.

Bezüglich der für das biographische und für das situative Interview gefundenen Korrelationen zwischen Erkennensfähigkeit und Leistung könnte eingesetzt werden, dass diese Korrelationen dadurch beeinflusst wurden, dass Teilnehmerinnen und Teilnehmer durch die Nachbefragung zur Selbstvorstellung sensibilisiert wurden und sich erst deswegen vermehrt Gedanken während der späteren Komponenten machten. Für künftige Studien wäre es deswegen wünschenswert, die Abfolge der verschiedenen Komponenten auszubalancieren. So würde eine Prüfung möglich, ob sich die Korrelationen zwischen Erkennensfähigkeit und Leistung für eine Interview-Komponente unterscheiden, je nachdem ob vorher bereits eine mögliche Sensibilisierung erfolgte oder nicht. Wenn man allerdings berücksichtigt, dass gängige Bewerbungs-Ratgeber (z. B. Hesse & Schrader, 2002) zum Teil relativ explizit erläutern, worauf bestimmte Interview-Fragen abzielen, so erscheint es durchaus möglich, dass der zusätzliche Sensibilisierungs-Effekt durch die erste Nachbefragung relativ gering ausfällt. Auch dies könnte untersucht werden, indem zum Beispiel die Erwartung von Teilnehmerinnen und Teilnehmern bezüglich möglicher Beurteilungsdimensionen erhoben wird.

Künftige Studien sollten schließlich noch versuchen zu klären, mit welchen anderen Konstrukten Erkennensfähigkeit in Zusammenhang steht. Maße für soziale Fertigkeiten oder situative Urteilsfähigkeit, wie dies die Studie von Salgado und Moscoso (2002) nahe legt, wären eine Möglichkeit, Maße für allgemeine Intelligenz oder spezifischere kognitive Fähigkeiten eine andere. Einschränkend muss allerdings angefügt werden, dass solche Maße lediglich dazu beitragen könnten, Unterschiede zwischen Personen zu erklären. Wie unsere intraindividuellen Analysen gezeigt haben, ist es aber auch für eine einzelne Person von Bedeutung, ob sie bei einer bestimmten Frage die jeweilige Dimension erkannt bzw. nicht erkannt hat.¹

Praktische Konsequenzen

Für Praktikerinnen und Praktiker sind die vorliegenden Ergebnisse bezüglich der fehlenden Konstruktvalidität eines strukturierten Einstellungsinterviews unproblematisch, solange sie mit solchen Interviews

lediglich Personalauswahl betreiben wollen. Da diese Interviews prognostisch valide Instrumente der Berufseignungsdiagnostik sind (z. B. Huffcutt et al., in press; Huffcutt & Woehr, 1999; McDaniel et al., 1994; Taylor & Small, 2002; Wiesner & Cronshaw, 1988), können sie trotz unklarer Konstruktvalidität gut eingesetzt werden. Vermieden werden sollte jedoch die Interpretation oder Rückmeldung der Beurteilungen einzelner Dimensionen. Zudem ist es nicht möglich, aus Interview-Ergebnissen Personalentwicklungsvorschläge abzuleiten, die dazu dienen sollen, Kompetenzen in jenen Bereichen aufzubauen, für die die entsprechenden Dimensionen kritisch beurteilt wurden.

Die problematischen Befunde zur internen Konstruktvalidität strukturierter Interviews sollten jedoch nicht dazu führen, künftig die Erfassung unterschiedlicher Anforderungsdimensionen im Rahmen der Berufseignungsdiagnostik gar nicht mehr zu versuchen. Nach wie vor stellt eine sorgfältige und umfassende Anforderungsanalyse eine entscheidende Grundlage für die erfolgreiche Auswahl geeigneter Personen dar. Im Rahmen dieser Anforderungsanalyse sollte deswegen auch weiterhin versucht werden, erfolgsrelevante Anforderungsdimensionen und -kriterien zu identifizieren. Wenn diese trotz entsprechender Bemühungen nicht mit Hilfe von Einstellungsinterviews erfasst werden können, ist es nötig, andere Methoden zu verwenden, mit denen sie besser und zuverlässiger erfasst werden können.

Literatur

- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Campion, M. A., Campion, J. E. & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79, 998–1002.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conway, J. M., Jako, R. A. & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579.
- Conway, J. M. & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13, 485–506.
- Harris, M. M. (1999). What is being measured? In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 143–157). Thousand Oaks, CA: Sage.
- Hesse, J. & Schrader, H. C. (2002). *Das perfekte Vorstellungsgespräch*. Frankfurt: Eichborn.
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Klehe, U.-C. (in press). Evaluation and comparison of the situational

¹ Wir danken Herrn Prof. Schuler dafür, dass er uns auf dieses Argument aufmerksam gemacht hat.

- and behavior description interview formats. *International Journal of Selection and Assessment*.
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., DeGroot, T. G. & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, 54, 619–644.
- Huffcutt, A. I. & Woehr, D. J. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior*, 20, 549–560.
- Hunter, J. E. & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 321–357). Oxford: Wiley.
- Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 158–168). Newbury Park, CA: Sage.
- Jeserich, W. (1990). *Mitarbeiter auswählen und fördern. Assessment-Center-Verfahren*. München: Hanser.
- Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, 78, 988–993.
- Kleinmann, M. (1997a). *Assessment-Center: Stand der Forschung, Konsequenzen für die Praxis*. Göttingen: Verlag für Angewandte Psychologie.
- Kleinmann, M. (1997b). Transparenz der Anforderungsdimensionen: Ein Moderator der Konstrukt- und Kriteriumsvalidität des Assessment-Centers. *Zeitschrift für Arbeits- und Organisationspsychologie*, 41, 171–181.
- Klimoski, R. & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243–260.
- Kolk, N. J. (2001). *Assessment centers: Understanding and improving construct-related validity*. Enschede: Ipskamp.
- Latham, G. P., Saari, L. M., Pursell, E. D. & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422–427.
- Marcus, B. (2003a). Das Wunder sozialer Erwünschtheit in der Personalauswahl. *Zeitschrift für Personalpsychologie*, 2, 129–132.
- Marcus, B. (2003b). Persönlichkeitstests in der Personalauswahl: Sind „sozial erwünschte“ Antworten wirklich nicht wünschenswert? *Zeitschrift für Psychologie*, 211, 138–148.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L. & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- Motowidlo, S. J. (1999). Asking about past behavior versus hypothetical behavior. In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 179–190). Thousand Oaks, CA: Sage.
- Salgado, J. F. & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, 299–324.
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica*, 38, 281–300.
- Schuler, H. & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183–192). Newbury Park, CA: Sage.
- Schuler, H. & Moser, K. (1995). Die Validität des Multimodalen Interviews. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 2–12.
- Taylor, P. J. & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277–294.
- Wiesner, W. H. & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275–290.
- Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189–205.

Eingegangen: 12. 01. 2004

Revision eingegangen: 08. 06. 2004

Dr. Klaus G. Melchers

Psychologisches Institut der Universität Zürich
Arbeits- und Organisationspsychologie
Rämistrasse 62
8001 Zürich
Schweiz
E-Mail: k.melchers@psychologie.unizh.ch