

**Berichte aus dem  
Psychologischen Institut der  
Universität Bonn**

Band 23, Heft 1 (1997)

**EINE AUSWERTUNGSOBJEKTIVE  
POSTKORBÜBUNG  
FÜR ASSESSMENT CENTER**

Jochen Musch & Walter Lieberei

**Jochen Musch & Walter Lieberei**

**Eine auswertungsobjektive Postkorbübung  
für Assessment Center<sup>1</sup>**

**Inhalt**

1. Einleitung
2. Forschungsstand
  - 2.1. Auswertungs- und Durchführungsobjektivität
  - 2.2. Konstruktvalidität
  - 2.3. Auswertungsverfahren
3. Konstruktion des Postkorbs
4. Methode
  - 4.1. Durchführung
  - 4.2. Stichprobe
5. Ergebnisse
6. Zusammenfassung und Diskussion
7. Literatur

*Aktualisierte Adressen (Stand: 6.6.2004):*

Dr. Jochen Musch  
Lehrstuhl Psychologie III  
Schloss Ehrenhof Ost 240  
D-68131 Mannheim

Dr. Walter Lieberei  
Giselastr. 35  
D-10317 Berlin

---

<sup>1</sup> Wir danken Joachim Funke, Thorsten Meiser und Niclas Schaper für wertvolle Anmerkungen und Kommentare zu einer früheren Fassung des Manuskripts. Der beschriebene Postkorb wurde inzwischen als Teil der „Bonner Postkorb-Module“ (BPM) vom Hogrefe-Verlag veröffentlicht ([www.testzentrale.de](http://www.testzentrale.de)).

## Zusammenfassung

Unzureichende Standardisierung, die fehlende Explikation eines normativen Bezugssystems und eine auf freier Eindrucksbildung beruhende Auswertung werden als Schwachpunkte von vielen in Assessment Centern eingesetzten Übungen identifiziert. Die Entwicklung einer Postkorbübung zur Überwindung dieser Fehlerquellen und zur Reduzierung der Anfälligkeit gegenüber Halo-Effekten und anderen Urteilsverzerrungen wird vorgestellt. Hauptziele der Entwicklung waren eine standardisierte Stimulusdarbietung, das Anlegen eines eindeutigen, kriterienorientierten Bewertungsmaßstabes und eine hohe Auswertungsobjektivität. Der Postkorb wurde in einem Assessment Center mit 54 Teilnehmern evaluiert. Dabei wies er eine gegenüber anderen Verfahren erhöhte diskriminante Validität auf und vermochte spezifische, von anderen Übungen nicht abgedeckte Varianzanteile aufzuklären.

## Abstract

Insufficient standardization, the missing explication of a normative frame of reference and scoring procedures based on free impression formation are identified as weaknesses of many exercises used in assessment centers. The development of an in-basket exercise is presented designed to overcome these sources of error and to reduce susceptibility to halo effects and other distortions of judgment. Main objectives of the development were a standardized presentation of stimuli, the application of an unequivocal, criteria-oriented frame of reference and a high degree of scoring objectivity. The in-basket exercise was evaluated in an assessment center with 54 participants. In comparison with other exercises, it showed enhanced discriminant validity and was able to explain specific portions of variance which were not covered by other methods.

# 1 Einleitung

Die simulierte Bearbeitung eines Stapels eingegangener Schriftstücke ist eine der am meisten verbreiteten situativen Übungen in Assessment Centern. Nach einer Zählung von Thornton und Byham (1982) wurde eine solche, oft einfach „Postkorb“ genannte Übung in mehr als 95% der Assessment Center (AC) eingesetzt, über die ein publizierter Bericht vorlag. Die Entstehung des Postkorbs als diagnostisches Instrument wird meist mit den Arbeiten von Frederiksen, Saunders und Wand (1957) sowie Bray und Grant (1966) in Verbindung gebracht. Funke (1993) weist allerdings zu Recht darauf hin, daß die Postkorbübung nicht ausschließlich als amerikanischer Exportartikel, sondern auch in der Tradition der Arbeitsprobe und der Psychognostik der 20er Jahre zu sehen ist.

Die Parallelen zwischen den Anforderungen, die der Postkorb an den Bearbeiter und der Berufsalltag an den Manager stellt, sind augenscheinlich (Dukerich, Milliken & Cowan, 1990; Neubauer, 1980). So fand Mintzberg (1973), daß Topmanager in seiner Stichprobe durchschnittlich 36 Schriftstücke am Tag zur Bearbeitung vorgelegt bekamen. Tatsächlich hat die häufig betonte hohe Augenscheinvalidität (Meyer, 1970; Hemphill, 1961; Lopez, 1966) die Verbreitung des Postkorbverfahrens erheblich beschleunigt. Studien zur Kontenvalidierung (z.B. Schmitt & Noe, 1983) belegen zudem, daß die in Postkörben auftretenden Anforderungen Repräsentativität für die tägliche Managementarbeit beanspruchen dürfen. Die augenscheinliche inhaltliche Validität des Verfahrens und die hohe Akzeptanz bei Bewerbern ist jedoch möglicherweise auch dafür verantwortlich, daß in Anbetracht der weiten Verbreitung der Methode vergleichsweise wenig Validierungsstudien durchgeführt wurden (Schippmann, Prien & Katz, 1990). Besonders im deutschsprachigen Raum liegen nur sehr wenige und überwiegend unveröffentlichte Untersuchungen zum Postkorbverfahren vor (z.B. Funke & Rasche, 1988; Didi, 1992; Strunz, 1994), was in auffälligem Gegensatz zu der weiten Verbreitung des Verfahrens bei der Personalauswahl steht. Daß die intendierte Erfassung administrativer, organisatorischer und analytischer Fähigkeiten mit Hilfe des Postkorbverfahrens eine erfolgreiche Prädiktion externer Kriterien erlaubt, ist vor allem durch eine Vielzahl von Studien aus dem englischsprachigen Raum belegt (Ward, 1959, 1960; Frederiksen, 1966; Bray & Grant, 1966; Bentz, 1968, 1971; Wollowick & McNamara, 1969; Meyer, 1970; Ginsbury & Silverman, 1972; Brass & Oldham, 1976; zusammenfassend Schippmann et al., 1990).

## 2 Forschungsstand

### 2.1 Auswertungs- und Durchführungsobjektivität

Viele in Assessment Centern eingesetzte Übungen sind durch unzureichend standardisierte Durchführungsbedingungen gekennzeichnet. Während beispielsweise bei einer Gruppendiskussion oder einem Mitarbeitergespräch die Redebeiträge der Teilnehmer nach Inhalt und Dauer ebenso wenig standardisiert werden können wie die anschließende Urteilsbildung der Beobachter, kommt der Postkorb zumindest in seiner Durchführung dem Ideal eines kontrollierten experimentellen Treatments nahe: Im Gegensatz zu anderen AC-Verfahren hängt die Leistung eines Teilnehmers nicht von der anderer Teilnehmer ab. Darüber hinaus eröffnet die Postkorbübung in einem weit höheren Maße als andere AC-Verfahren die Perspektive einer objektiven, beobachterunabhängigen Auswertung und Interpretation der Leistung von AC-Teilnehmern.

Um den für die Auswertung erforderlichen Aufwand zu minimieren, findet die Bewertung der Postkorbleistung allerdings häufig analog zur Vorgehensweise bei anderen AC-Verfahren während der Präsentation der Ergebnisse durch den Postkorbbearbeiter im Zuge einer globalen Eindrucksbildung auf vorgegebenen Eigenschaftsdimensionen statt. Durch diese Vorgehensweise wird die Möglichkeit des Einsatzes objektivierter Auswertungsverfahren jedoch verschenkt; die beobachteten Interraterreliabilitäten schwanken mit Werten im Bereich von  $r = -.20$  bis  $.97$  erheblich (Weise, 1975; Schippmann et al., 1990). Nur durch ein für die Praxis häufig zu aufwendiges Beobachtertraining kann der Grad der Übereinstimmung erhöht werden (z.B. von  $r = .52$  auf  $.84$  bei einer Studie von Richards & Jaffee, 1972). Ein gravierender Nachteil der "impressionistischen" Vorgehensweise ist die Gefahr von Urteilsverzerrungen: "Problems with supervisor's rating are legion. Leniency, halo, and restriction-in-range biases may occur" (Thornton & Byham 1982, S. 298). Auch Similar-to-me-, Primacy- und Recency-Effekte gefährden die Objektivität von eindrucksgestützten Beurteilerratings (Robertson, Gratton & Sharpley, 1987). Durch den Einsatz des hier vorgestellten auswertungsobjektiven Postkorbs sollen beobachterspezifische Urteilsverzerrungen ausgeschlossen werden.

## 2.2 Konstruktvalidität

Im Hinblick auf die Konstruktvalidität von AC-Übungen, soweit sie nach der eigenschaftsorientierten Multitrait-Multimethod-Methode von Campbell und Fiske (1959) bestimmt wird, liegen enttäuschende Ergebnisse vor. Sackett und Dreher (1982), Turnage und Muchinsky (1984), Silverman, Dalessio, Woods und Johnson (1986) und andere kamen zu dem Ergebnis, daß die konvergente und diskriminante Validität bei den üblicherweise im AC erhobenen Eigenschaftsdimensionen Anwender und Forscher nicht befriedigen können. Diese Kritik wurde sowohl auf den Postkorb als auch auf andere AC-Verfahren bezogen: Die auf Eindrucksbildung beruhenden Beurteilungen ganz unterschiedlicher Eigenschaftsdimensionen korrelierten untereinander mit  $r = .8$  bis  $.9$  sehr stark; gleichzeitig waren Ratings ein und derselben Eigenschaftsdimension über verschiedene Verfahren hinweg häufig nahezu unabhängig voneinander. Beispielsweise beobachteten Brannick, Michaels und Baker (1989) aufgrund von Haloeffekten in ihrem Postkorbverfahren zwischen den Beurteilungen unterschiedlichster Eigenschaftsdimensionen artifiziell überhöhte Interkorrelationen von durchschnittlich  $r = .84$ . Ähnlich unbefriedigende Werte für die diskriminante Validität klassischer Postkorbübungen berichten Bycio, Alvares und Hahn (1987;  $r = .77$ ) und Gaugler und Thornton (1989;  $r = .83$ ). In der letztgenannten Studie konnte zudem gezeigt werden, daß auch die Reduktion der Anzahl der im Postkorb einzustufenden Eigenschaftsdimensionen von neun auf drei trotz verringerter kognitiver Anforderungen den Halo-Effekt eher verstärkt ( $r = .89$ ). Eine positive Ausnahme bildet lediglich die Studie von Sackett & Dreher (1982), bei der für drei Postkörbe eine mittlere diskriminante Validität von  $r = .51$  beobachtet wurde. Im allgemeinen gelingt es menschlichen Beobachtern in der klassischen, auf Eindrucksbildung beruhenden Postkorbübung offenbar nicht, zwischen unterschiedlichen Eigenschaftsdimensionen wie beispielsweise dem Analysevermögen, dem Organisationsgeschick und dem Entscheidungsverhalten von Bewerbern zu differenzieren.

Aufgrund dieser Befundlage wurde verschiedentlich der Schluß gezogen, daß es um die Konstruktvalidität von AC-Beurteilungen schlecht bestellt sei (Brannick, Michaels & Baker, 1989; Klimoski & Brickner, 1987). Manche Autoren argumentieren, daß sich die Systemprinzipien des AC ohnehin vom eigenschaftstheoretischen Konzept gelöst hätten und verhaltensorientiert seien; AC-Übungen wären demnach eher in der Tradition der Arbeitsprobe als

in der Tradition der Persönlichkeitsdiagnostik zu sehen. Die verschiedenen Übungen werden dabei als „sample“ verschiedener Anforderungen angesehen und nicht als Mittel zur Evozierung von „signs“ für führungsrelevante Eignungsdimensionen (vgl. Wernimont & Campbell, 1968). Das Hauptziel der vorliegenden Untersuchung war die Entwicklung einer durchführungs- und auswertungsobjektiven Postkorbübung, um zu prüfen, ob durch die Reduzierung von Beurteilerfehlern die eigenschaftsbezogene Konstruktvalidität von Assessment Centern erhöht werden kann. Gleichzeitig sollte die Stellung einer solchen Postkorbübung im Verbund klassischer AC-Übungen bestimmt werden.

### 2.3 Auswertungsverfahren

Ein denkbarer Lösungsansatz zur Überwindung der mit der freien Eindrucksbildung im AC verbundenen Probleme und zur Erhöhung der Auswertungsobjektivität bei Postkorbübungen ist der Rückgriff auf schriftliche Aufzeichnungen des Bearbeiters. Das Abschneiden beim Postkorb hängt bei Auswertung der frei formulierten schriftlichen Aufzeichnungen der Teilnehmer allerdings stark von der Zahl produzierter Äußerungen und damit auch von der Wortflüssigkeit und der Schreibgeschwindigkeit ab (Frederiksen, 1966). Sinnvoller erschien uns deshalb eine standardisierte Vorgehensweise bei der detaillierten Auswertung der Güte aller aufgrund der jeweiligen Schriftstücke durchgeführten Handlungen, die der Bearbeiter im Anschluß an die Bearbeitung des Postkorbs vorstellt. Anhand eines vorher festgelegten Kriterienkataloges wird dabei über die Angemessenheit der gewählten Handlungsalternativen entschieden. Durch diese Vorgehensweise können Urteilsverzerrungen weitestgehend ausgeschlossen und die Auswertungsobjektivität entscheidend verbessert werden (Funke, 1993). Die statistische Datenkombination hat sich zudem der freien, „klinischen“ Urteilsbildung auch in der Vorhersage von Außenkriterien in der Regel als überlegen erwiesen (Meehl, 1954; Lechner, 1979). In der AC-Forschung indizieren die Ergebnisse z.B. von Wollowick und McNamara (1969) und Mitchell (1975) ebenfalls diese Überlegenheit. Der für eine Standardisierung des Instruments erforderliche Aufwand ließ sich aufgrund seines Einsatzes bei der Auswahl aus einer Vielzahl von Bewerbern für gleichgeartete Stellen auch ökonomisch rechtfertigen.

Neben den offensichtlichen Vorteilen bringt die psychometrisch begründete Postkorbauswertung auch eine Reihe spezifischer Probleme mit sich, auf die wir im folgenden eingehen und denen ein Lösungsversuch entgegengestellt werden soll.

## 3 Konstruktion des Postkorbs

Ein grundsätzliches Problem objektiver, testorientierter Auswertungsschemata für Postkorbübungen ist die Festlegung normativer Musterlösungen, die oft willkürlich und angreifbar erscheinen. Ursache hierfür ist vor allem der Umstand, daß sich noch immer viele in der Praxis eingesetzte Postkörbe an einer einflußreichen Empfehlung von Lopez (1966) orientieren, derzufolge sich Postkorbitems durch ein hohes Maß von Uneindeutigkeit und Unklarheit auszeichnen sollten, um genügend Spielraum für differentielle Bearbeitungsmuster zu lassen. Diese Empfehlung ist mit der Konstruktion auswertungsobjektiver Postkorbttests unvereinbar: Die für ein valides diagnostisches Instrument unabdingbare Auswertungsobjektivität ist notwendig an eine eindeutige Festlegung der Angemessenheit möglicher Antwortalternativen geknüpft.

Dabei ist in zweideutigen, unstrukturierten Situationen bereits die erschöpfende Festlegung aller denkbaren Handlungsoptionen nur schwer zu erreichen. Kesselman, Lopez und Lopez (1982) erstellten beispielsweise eine Liste von nicht weniger als 684 Handlungsalternativen für 26 Items. Dennoch erwies sich die Liste als nicht vollständig. Zudem ist die zeitraubende Auswertung mit einem umfangreichen Manual für die durch Zeitdruck gekennzeichnete Praxis zu unökonomisch.

Die Festlegung der Angemessenheit von Antwortalternativen gestaltet sich in unstrukturierter Situationen ebenfalls schwierig. Der Rückgriff auf die Sachkompetenz von Experten löst das Problem nämlich nicht: Die beteiligten Fachleute bei Frederiksen (1962) erzielten trotz eines erheblichen Aufwandes keinen Konsens über die Rangreihe der Angemessenheit von Reaktionen auf die Vorgänge des Postkorbs. Die von Frederiksen (1962) gewählte Lösung, auf eine Bewertung der Korrektheit der Entscheidungen gänzlich zu verzichten und stattdessen Skalen für deren "Stil", "Kreativität" und das "Ausmaß organisationaler Veränderungen" zu etablieren, erscheint - obwohl sie viele Nachfolger gefunden hat - ebenfalls unbefriedigend. Zudem besteht die Gefahr, daß derartige Bewertungen aus impliziten Führungstheorien und anderen schwer objektivierbaren Kriterien abgeleitet werden. Auch die empiristische Alternative, stets die von der Mehrheit der AC-Teilnehmer präferierten Vorgehensweisen als korrekt zu bewerten, ist inhaltlich kaum angemessen.

Wir haben uns deshalb entschlossen, ausschließlich solche Handlungsalternativen zu bewerten, die sachlogisch zwingend aus den zur Verfügung gestellten Informationen abzuleiten sind, und auf eine normative Bewertung von Handlungen in mehrdeutigen Situationen gänzlich zu verzichten. Damit einher ging eine Minimierung der Häufigkeit unstrukturierter Situationen zugunsten von Vorgängen, die eine eindeutig als korrekt zu wertende Verhaltensweise erfordern. Bei diesen wurde dann positiv bewertet, wenn beispielsweise eine Information an die zuständige Stelle weitergeleitet wurde, eine Terminkollision erkannt und eine logisch eindeutig aus den Schriftstücken ableitbare Folgerung gezogen wurde. Keine Punkte wurden hingegen für Entscheidungen und Vorgehensweisen vergeben, die vom individuellen Stil, idiosynkratischen Präferenzen und nicht unmittelbar aus dem Stimulusmaterial ablesbaren Randbedingungen determiniert wurden.

Indem die Bewertungskriterien für das Vorliegen einer korrekten Vorgehensweise vorab eindeutig formuliert werden, wird also eine weitestgehende Auswertungsobjektivität sichergestellt. Erkauft wird dieser Vorteil allerdings durch den Verzicht auf mehrdeutige, offene Postkorbitems, die breiteren Raum für eine Projektion individueller Stile lassen, was zur Folge hat, daß kreative und originelle Lösungsvorschläge nicht gewürdigt werden können.

Prinzipiell erscheint der Einsatz von Computern bei der Postkorbdurchführung als ein zur Erhöhung der Auswertungsobjektivität besonders geeignetes Verfahren. Funke und Rasche (1988), die sich ebenfalls die Steigerung der Auswertungsobjektivität und die Verringerung des Zeitaufwands für die Auswertung zum Ziel gesetzt hatten, entwickelten deshalb folgerichtig einen computergestützten "PC-Postkorb". Die Frage nach der Vergleichbarkeit von computergestützten Postkörben mit ihren Papier-und-Bleistift-Äquivalenten ist allerdings ungelöst (Funke, 1993; vgl. Klieme & Stumpf, 1990). Ein erheblicher Nachteil einiger existierender computergestützter Postkörbe (z.B. Horn, 1991; Roest & Horn, 1990) ist zudem die Fragwürdigkeit der aus der Menge des aufgezeichneten Datenstroms gebildeten, häufig beliebig wirkenden Indikatoren (wie etwa die Reihenfolge und Anzahl der Lesevorgänge und die Lesedauer einzelner Dokumente; Funke, 1993) und die Gefahr, PC-spezifische Fertigkeiten wie die Mausbedienung mitzuerfassen und damit Computernovizen unangemessen

zu benachteiligen. Dem steht allerdings die stark wachsende Verbreitung computergestützter Informationssysteme im Büroalltag gegenüber, die die Verwendung entsprechender Postkorbverfahren zunehmend realitätsnäher werden läßt.

Im Bemühen um auswertungsobjektive und sinnvolle Performanzindikatoren setzten Funke und Rasche (1988) bei ihrem computergestützten PC-Postkorb Multiple-Choice-Items ein, die Fachleute unter rationalen Gesichtspunkten mit Plus- und Minusbewertungen insgesamt vier Skalen zuordneten, welche logisch aus den Anforderungen der Situation abgeleitet wurden. Kritisch ist bei dieser die Auswertungsobjektivität und -ökonomie maximierenden Vorgehensweise jedoch - wie die Autoren selbst anmerken -, daß nicht mehr die Generierung von Lösungsmöglichkeiten, sondern nur noch die Auswahl zwischen verschiedenen Vorgaben gefordert wird. Dies wiegt umso schwerer, als alle Bestandteile einer entsprechend der jeweiligen Beurteilungsdimension "guten" Reaktion in den Antwortalternativen schon enthalten sein müssen. Didi (1992) bemerkt hierzu, daß wohl kaum ein Teilnehmer eine Antwortoption, die ein Lob für einen guten, unaufgeforderten Vorschlag eines Mitarbeiters vorsieht, nicht ankreuzen würde; bei der Durchsicht von frei generierten Reaktionen stellte er jedoch fest, daß ein solches Lob tatsächlich nur von wenigen Teilnehmern ausgesprochen wurde. Auch ob ein Teilnehmer das wichtige Konzept der Delegation verfügbar hat oder nicht, kann beim Multiple-Choice-Antwortformat nicht festgestellt werden, da ihm dieses Prinzip in den Wahlmöglichkeiten vorgegeben werden muß. Hinzu kommt das Problem, daß durch die Vorgabe von Handlungsalternativen die Vernetzung der Vorgänge allzuleicht aufgedeckt und dadurch die Schwierigkeit der Bearbeitung reduziert wird. Insgesamt erscheint die Inhaltsvalidität von Arbeitsproben durch das Multiple-Choice-Antwortverfahren gefährdet (in diesem Sinne auch Sackett, 1987). Wir haben aus diesem Grund auf seinen Einsatz verzichtet. Ebenfalls verzichtet wurde auf die verbreitete Maßnahme, neben Bonuspunkten für korrekte Antworten auch Maluspunkte für falsche Entscheidungen in das Bewertungssystem mitaufzunehmen (vgl. Funke & Rasche, 1988). Ein Postkorb ist nämlich sicher zumindest partiell auch als Speedtest aufzufassen: Es kommt durchaus vor, daß ein Teilnehmer nicht alle Vorgänge in der vorgegebenen Zeit zu verarbeiten vermag. Deshalb kann es bei Vergabe von Maluspunkten leicht zu der paradoxen Situation kommen, daß ein gar nicht bearbeitetes Schriftstück günstiger bewertet wird als ein fehlerhaft bearbeitetes.

Die Konstruktion fast aller in der Praxis verwendeten Postkörbe orientiert sich an einigen wenigen veröffentlichten Beispielen (z.B. Jeserich, 1981, S. 172-193; Kliem, 1983, S. 194-196; Frederiksen et al., 1957, S. 25-28). Ungeachtet häufiger Hinweise auf die Notwendigkeit einer anforderungsanalytisch geleiteten Postkorbkonstruktion fehlen nämlich bis heute differenzierte Vorschläge und Erzeugungsvorschriften für eine standardisierte Vorgehensweise bei der Erstellung von Postkorbübungen: "... in all the research reviewed, there was no evidence of any effort to develop a test plan or item budget for construction of in-basket tests following conventional procedures. [...] There simply are no published or widely distributed reports which describe how to develop an in-basket that is well grounded with regard to content validity" (Schippmann et al., 1990, S. 851). Bei der Konstruktion der Schriftstücke orientierten wir uns zunächst ebenfalls an bestehenden Vorbildern und Konventionen. Inhaltlich orientierten wir uns an zur Zielposition funktionsanalogen organisatorischen und administrativen Anforderungen an einen Gruppenleiter der untersten Führungsebene, ohne durch eine direkte Übernahme des Settings und der Problemstellungen den Einfluß von spezifischen Vorkenntnissen zu groß werden zu lassen. In der Rolle des Abteilungsleiters einer Zeitarbeitsfirma trägt der Bearbeiter deshalb Personalverantwortung für die Abteilung „Kraftfahrer“ und übernimmt aufgrund einer Erkrankung des Hauptabteilungsleiters „Personal“ auch dessen Vertretung. Ein Organigramm und eine Charakterisie-



rung der für die Postkorbbearbeitung relevanten Mitarbeiter geben die erforderlichen Hintergrundinformationen; die ausschließliche Darbietung dieser und aller anderen für die Durchführung relevanten Informationen in Schriftform gewährleistet eine vollständig objektivierte Instruktionserteilung. Die 17 zu bearbeitenden Schriftstücke beziehen sich auf Themenkreise und Vorgänge wie die Weiterbildung der Mitarbeiter, die Projektmitarbeit für den Bau eines neuen Verwaltungsgebäudes, krankheitsbedingte Fehlzeiten, eine Schadensverursachung durch ein Dienstfahrzeug, EDV-Probleme usw. Durch unterschiedliche Schreibstile und Briefköpfe sowie typische Formulare für hausinterne Mitteilungen wurde eine größtmögliche Authentizität der Schriftstücke angestrebt. Punkte werden bei der Auswertung beispielsweise für das Erkennen von Terminkollisionen, die Koordination von Terminen, die Weiterleitung von Informationen an betroffene Stellen, das Erkennen von Zusammenhängen, sachlich erforderliche Entscheidungen und die Identifikation von Irrläufern vergeben. Handlungsalternativen in Situationen, die interpretativen Spielraum lassen, sowie kreative und originelle Lösungsvorschläge werden nicht bewertet. Unter inhaltlichen Gesichtspunkten sind die zu vergebenden Punkte drei Subskalen zugeordnet: Logische Schlußfolgerungen werden mit der Skala „Analyseverhalten“, administrative Maßnahmen mit der Skala „Organisation & Planung“ und Entscheidungen mit der Skala „Entscheidungsverhalten“ erfaßt. Abbildung 1 (folgende Seite) zeigt exemplarisch zwei im Postkorb enthaltene Schriftstücke.

Abbildung 1: Ausschnitte aus zwei Beispielschriftstücken

[...]

Sehr geehrter Herr Lenk,

Ihr Mitarbeiter, Herr Schneider, ist am gestrigen Donnerstag mit einem Lastwagen der Spedition Speedy (amtliches Kennzeichen F-AB 356) von der Straße abgekommen und über mein Grundstück im Philosophenweg 45 (Ecke Schillerstraße) gefahren; dabei hat er einen Teil der Gartenanlage (eine Hecke, mehrere Blumenbeete) zerstört. Zusätzlich wurde ein Telefonmast auf meinem Grundstück in Mitleidenschaft gezogen, der inzwischen von den Stadtwerken ausgetauscht werden mußte.

Ich fordere hiermit zum einen Schadensersatz für die mir von den Stadtwerken in Rechnung gestellten Reparaturkosten in Höhe von DM 4.523.- Zusätzlich erwarte ich, daß Sie die anstehenden Gärtnereikosten für die Wiederherstellung meines Gartens in den vorherigen Zustand in voller Höhe übernehmen.

Von der Spedition Speedy wurden meine diesbezüglichen Forderungen mit dem Hinweis abgelehnt, für Zeitmitarbeiter hafte die ihn beschäftigende Zeitarbeitsfirma; nach dieser Rechtsauffassung haben Sie für den von Ihrem Mitarbeiter verursachten Schaden aufzukommen.

Ihrer Antwort sehe ich bis zum 13. Mai entgegen.

Mit freundlichen Grüßen

[...]

### **EILIGES FAX!**

Sehr geehrter Herr Lenk,

nachdem Herr Schneider bereits am vergangenen Donnerstag (bei der Heimfahrt von einer Lieferung) einen Unfall verschuldet hat, ist er am heutigen Montag offenbar nicht zum Dienst angetreten; soeben hat uns die Firma Merklinger angerufen, sie würden immer noch auf jemanden warten, der die Bananenkisten abholt! Wo bleibt Herr Schneider nur?

Wir können es uns nicht leisten, Aufträge unserer Kunden einfach unerledigt zu lassen. Wenn Sie Wert darauf legen, weiter mit uns zusammenzuarbeiten, sollten Sie sich stärker um Zuverlässigkeit bemühen, als das bisher der Fall war.

Mit freundlichen Grüßen

Spedition Speedy

Einen Punkt auf der Skala „Organisation & Planung“ enthält der Bearbeiter beispielsweise, wenn er die Schadensersatzforderung an die Rechtsabteilung weiterleitet. Für das „Entscheidungsverhalten“ wird positiv gewertet, wenn ein Ersatzfahrer zur Firma Merklinger geschickt wird. Einen weiteren Punkt auf der Skala „Analyseverhalten“ erhält der Postkorbbearbeiter schließlich, wenn er den möglichen Zusammenhang zwischen dem LKW-Unfall und dem aus einem weiteren Schriftstück herauszulesenden hohen Alkoholkonsum von Herrn Schneider erkennt.

Geprüft wird im folgenden die Hypothese, daß ein im geschilderten Sinne auswertungsobjektiver Postkorb einen spezifischen Beitrag zur Potentialdiagnose zu erbringen vermag, der über den Beitrag von klassischen AC-Übungen, die auf freier Eindrucksbildung der Beobachter basieren, hinausgeht. Zusätzlich soll die Frage geprüft werden, ob durch die erhöhte Auswertungsobjektivität das Ausmaß von Beurteilerfehlern reduziert und dadurch die eigenschaftsbezogene Konstruktvalidität von Assessment Centern erhöht werden kann.

## 4 Methode

Die Daten der Studie wurden in einem Dienstleistungsunternehmen der Luftfahrtbranche erhoben. Das Unternehmen, das rund 5.000 Mitarbeiter in vier Geschäftsbereichen beschäftigt, wurde Anfang 1993 privatisiert und hatte in der Folgezeit eine erhöhte Nachfrage nach Personalauswahlverfahren, da viele Stellen intern neu ausgeschrieben oder erst geschaffen wurden. Die Aufbauorganisation des Unternehmens ist durch eine flache Hierarchie mit einer maximalen Führungsspanne von 12 Mitarbeitern gekennzeichnet. Auswahlrichtlinie für Führungskräfte ist nach den Unternehmensgrundsätzen ein „kooperativer“ Führungsstil. Vor der Privatisierung fanden keine nennenswerten überfachlichen Personalentwicklungsmaßnahmen und Führungskräftebildungen statt.

### 4.1 Durchführung

Ausgewertet wurden die Ergebnisse von Assessment Centern, die 1995 von der Personalentwicklungsabteilung (PE) für den größten Geschäftsbereich des Unternehmens durchgeführt wurden. Ziel der Personalauswahl waren drei ähnlich geartete Positionen der untersten Führungsebene. Diese waren in größerem Umfang neu zu besetzen, was den Aufwand für die Entwicklung standardisierter Diagnoseinstrumente rechtfertigte. Die AC-Teilnehmer wurden mit einem strukturierten Interview ausgewählt, an dem der Linienvorgesetzte sowie ein PE-Mitarbeiter beteiligt waren. Bei dieser Vorselektion wurde großzügig verfahren, um unabhängig von der Selektionsentscheidung für möglichst viele Mitarbeiter des Unternehmens eine gesicherte Basis für Weiterbildungsempfehlungen zu erhalten. Die Selektionsquote betrug ca. 50 %.

Maßgeblich für die AC-Konzeption waren zwei von der Abteilung PE organisierte Workshops, zu denen Vertreter der betroffenen Fachbereiche eingeladen wurden. Die Fachvertreter gehörten einer Führungsebene an, die ein bis zwei Ebenen über der AC-Zielebene angesiedelt ist. Bei den moderierten Workshops wurden die leistungsbestimmenden Aufgabenbereiche und kritischen Situationen der drei Führungsfunktionen festgelegt und daraus ein jeweils positionsspezifischer Pool an anforderungsnahen AC-Übungen entwickelt. Außerdem wurden die erforderlichen Anforderungs- bzw. Eignungsmerkmale und deren Ausprägungsgrade bestimmt, so daß sich ein „Soll-Profil“ für jede der drei Stellen ergab. Alle AC setzten sich aus fünf Übungen zusammen: einer Gruppendiskussion (GD), einem Mitarbeiterge-

sprach (MAG), einer Präsentation (PRÄ), einer schriftlichen Fallstudie (SFS) und einem Postkorb (PK). In je nach Übung unterschiedlicher Zusammenstellung wurden dabei von den Beobachtern folgende Anforderungsmerkmale eingestuft: Analyseverhalten, Organisation & Planung, Entscheidungsverhalten, Führungsverhalten, Überzeugungskraft, Einfühlungsvermögen, Kooperationsverhalten, Konfliktlöseverhalten, Initiative & Engagement, Flexibilität, Ausdrucksvermögen und Streßbewältigung. Beim Postkorb wurden die drei Subskalen „Analyseverhalten“, „Organisation & Planung“ und „Entscheidungsverhalten“ durch eine eindrucksgeladete Einstufung der Streßbewältigung bei der Vorstellung der Ergebnisse ergänzt. Gewertet wurden ausschließlich überfachliche Kompetenzaspekte, die auf aggregierter Ebene nach logisch-rationalen Gesichtspunkten zu den drei Bereichen „Methodenkompetenz“, „Sozialkompetenz“ und „Selbstkompetenz“ zusammengefaßt wurden (siehe Abbildung 2).

Abb. 2: Matrix der Anforderungsmerkmale und AC-Übungen mit den Mittelwerten und Standardabweichungen der 54 Teilnehmer

### Methodenkompetenz

	GD	MAG	PRÄ	PK	SFS	Ist	Soll
Analyseverhalten				3,2 (1,2)	4,1 (1,2)	3,6 (0,9)	5,0
Organisation und Planung			4,1 (1,1)	4,4 (0,8)	3,8 (1,2)	4,1 (0,7)	5,0
Entscheidungsverhalten				4,2 (1,3)	3,6 (1,6)	3,9 (1,1)	5,5

### Sozialkompetenz

	GD	MAG	PRÄ	PK	SFS	Ist	Soll
Führungsverhalten	3,4 (1,2)	4,0 (1,1)				3,7 (1,0)	5,0
Überzeugungskraft	3,5 (1,0)	4,0 (1,0)	3,9 (1,2)			3,8 (0,8)	5,5
Einfühlungsvermögen	4,0 (0,9)	4,2 (1,0)				4,1 (0,8)	4,5
Kooperationsverhalten	3,7 (1,0)					3,7 (1,0)	5,0
Konfliktlöseverhalten	3,6 (0,9)	4,1 (0,9)				3,8 (0,7)	5,0

### Selbstkompetenz

	GD	MAG	PRÄ	PK	SFS	Ist	Soll
Initiative und Engagement	3,7 (1,2)	4,2 (1,0)	4,0 (1,3)			4,0 (0,8)	5,5
Flexibilität		4,1 (0,9)	4,1 (1,3)			4,1 (0,9)	5,0
Ausdrucksvermögen			4,1 (1,0)		4,3 (1,2)	4,2 (0,9)	4,5
Streßbewältigung			4,2 (1,1)	4,5 (0,9)		4,4 (0,8)	4,5

Anmerkungen: GD = Gruppendiskussion, MAG = Mitarbeitergespräch, PRÄ = Präsentation, PK = Postkorb, SFS = Schriftliche Fallstudie; SOLL = Sollprofil einer der Zielstellen.

An jedem AC nahmen nach einem festgelegten Zeitplan bis zu 12 Bewerber teil. Um eine unmittelbare Konkurrenzsituation zu vermeiden, setzte sich das Teilnehmerfeld in den meisten Fällen aus Bewerbern unterschiedlicher Niederlassungen zusammen. Bei den PE-Beobachtern handelte es sich um interne oder externe Diplom-Psychologen. Die Fachbeobachter rekrutierten sich überwiegend aus den unmittelbaren Linienvorgesetzten der Teilnehmer. Nach einem rotierenden Beobachtersystem wurde jeder Teilnehmer pro Übung von einem Fach- und einem PE-Beobachter eingestuft. Insgesamt waren für 12 Teilnehmer drei Fach- und drei PE-Beobachter vorgesehen. Das Beobachtungssystem war dahingehend optimiert, daß jeder Beobachter jeden Teilnehmer in einer der Übungen beurteilen konnte. Gleichzeitig wechselte systematisch das Tripel „Teilnehmer - Fachbeobachter - PE-Beobachter“, um den Einfluß der Beobachterkonstellation auf das AC-Ergebnis zu minimieren.

Die Dauer des Assessment Centers betrug in der Regel drei Tage; lediglich bei einer Teilnehmerzahl von weniger als sieben reduzierte sich die Dauer des AC auf zwei Tage. Der erste Tag begann vormittags mit einer mehrstündigen Beobachterschulung. Sofern es sich um erfahrene Beobachter handelte, beschränkte sich die Schulung auf eine Auffrischung früherer Schulungsinhalte und eine Einweisung in die jeweils relevanten Übungen. Daran schloß sich die Teilnehmereinweisung mit einer Erläuterung des AC-Ablaufs an. Am Ende des ersten Tages fand die erste AC-Übung, die Gruppendiskussion, statt. In einer Gruppe von bis zu sechs Teilnehmern waren nach einer Vorbereitungszeit von 15 Minuten unternehmensrelevante Themen zu diskutieren. Die Diskussionsdauer war auf 45 Minuten beschränkt. Die übrigen vier Übungen fanden am zweiten Tag, verteilt auf den Vormittag und den Nachmittag, statt. Beim Mitarbeitergespräch war nach einer zwanzigminütigen Vorbereitungszeit eine vorgegebene Problemstellung mit einem Mitarbeiter, der vom PE-Beobachter verkörpert wurde, zu klären. Die Gesprächsdauer betrug 20 Minuten. Die Vorbereitungszeit bei der Präsentation war mit 40 Minuten länger bemessen, da in den Vortrag Visualisierungshilfen miteinbezogen werden konnten. Der PE-Beobachter und der Fachbeobachter simulierten die fiktiven Zuhörer und stellten während der zweiten Hälfte der zwanzigminütigen Präsentation Fragen an den Vortragenden. Bei der schriftlichen Fallstudie fielen die Vorbereitungs- und Durchführungsphase zusammen. Innerhalb von einer Stunde sollte zu einem führungsrelevanten Thema schriftlich Stellung genommen werden. Ausgangspunkt bildeten dabei meist einschlägige Zeitschriftenartikel, deren Aussagen auf die Unternehmenssituation übertragen werden mußten.

Für die Bearbeitung des Postkorbs standen 60 Minuten zur Verfügung. Im Anschluß an die Bearbeitung wurden die Ergebnisse dem PE- und dem Fachbeobachter präsentiert. Dafür waren 30 Minuten vorgesehen. Während der Präsentation erfolgte die Vergabe der Punkte anhand eines objektiven Auswerteschemas, das für jeden Vorgang des Postkorbs die zu erkennenden Erfordernisse auflistet und für jede korrekte Handlung, Analyse oder Entscheidung die Vergabe eines Rohwertpunktes in einer der Skalen „Analyseverhalten“, „Organisation & Planung“ und „Entscheidungsverhalten“ vorsieht.

Die Auswertung der anderen Übungen erfolgte unmittelbar im Anschluß an ihre Durchführung, mit Ausnahme der schriftlichen Fallstudie, die am Abend des zweiten Tages von den Beobachtern anhand geeigneter Bewertungsanker bewertet wurde; die Dauer der Auswertungsphase betrug dabei je nach Übung zwischen 20 und 30 Minuten. Die Leistung eines Teilnehmers bei einer Übung wurde jeweils anhand von bis zu sechs Anforderungsmerkmalen bewertet (s. Abb. 2). Der numerische Wert eines Anforderungsmerkmals der Ergebnismatrix errechnete sich dabei aus der Mittelung der zugeordneten Verhaltensanker. Im Konsens zwischen den beiden Beobachtern wurden pro Anforderungsmerkmal vier bis fünf Ver-

haltensanker auf einer Skala von „1“ bis „7“ eingestuft; „1“ stellte dabei die Minimal-, „7“ die Maximalausprägung dar.

Das AC wurde am dritten Tag vormittags mit einer Beobachterkonferenz fortgesetzt, in der die Ergebnisse der beiden vorherigen Tage zusammengetragen, Empfehlungen hinsichtlich einer Stellenbesetzung ausgesprochen und Entwicklungsmaßnahmen festgelegt wurden. Abgeschlossen wurde das AC mit Feedback-Gesprächen, bei denen den Teilnehmern von jeweils einem PE- und einem Fachbeobachter die AC-Ergebnisse und die daraus resultierenden Konsequenzen (Empfehlung hinsichtlich der Stellenbesetzung, Weiterbildungsmaßnahmen usw.) mitgeteilt wurden. Nach Abschluß des AC wurden jedem Teilnehmer innerhalb von 14 Tagen die Inhalte des Rückmeldegesprächs in Form eines Kurzberichts an die Privatadresse zugesandt. In verkürzter Form wurde das Resultat auch dem Linienvorgesetzten mitgeteilt, der das AC-Ergebnis bei seiner Entscheidung über die Stellenbesetzung miteinfließen lassen konnte.

#### 4.2 Stichprobe

Am AC nahmen 54 Personen teil, 46 Männer und acht Frauen. Diese Quote spiegelt den hohen Männeranteil in diesem Geschäftsfeld wider. Das Alter der Teilnehmer lag zwischen 26 und 53 Jahren ( $M = 41,4$ ,  $SD = 5,5$ ) und war annähernd normalverteilt (Shapiro-Wilks'  $W = .99$ ,  $p < .90$ ). Bei den AC-Teilnehmern handelte es sich ausnahmslos um interne Bewerber mit mehrjähriger Berufserfahrung, die sich aus zehn über die gesamte Bundesrepublik verteilten Niederlassungen zusammensetzten. Das entspricht ca. 2/3 aller Niederlassungen des Unternehmens. In der Regel bewarben sich die Teilnehmer auf Stellen in der eigenen Niederlassung. Überregionale Bewerbungen stellten die Ausnahme dar.

### 5 Ergebnisse

Zwischen dem Postkorb- bzw. AC-Ergebnis und den Stichprobenvariablen Geschlecht, Alter und Niederlassung zeigten sich keine signifikanten Zusammenhänge. Die beobachtete geringe Korrelation zwischen dem Postkorb-Gesamtwert und dem Alter der Teilnehmer ( $r = .05$ , n.s.) ist ein Beleg dafür, daß die mit dem Alter zunehmende Berufserfahrung für sich genommen keine bedeutende Rolle für das Abschneiden im Postkorb spielt, was seine Tauglichkeit als Prognoseinstrument unterstreicht.<sup>2</sup>

Die mittlere Korrelation zwischen Eigenschaften mit der gleichen Bezeichnung über verschiedene Verfahren hinweg (konvergente Validität) war mit  $r = .27$  relativ gering. Demgegenüber waren die Einstufungen verschiedener Merkmale innerhalb des gleichen Verfahrens bei der Gruppendiskussion, dem Mitarbeitergespräch, der schriftlichen Fallstudie und der Präsentation mit  $r = .73$  sehr stark korreliert. Dies weist auf mögliche Halo-Effekte, eine überlappende Merkmalserfassung durch die Beurteiler oder auf Überschneidungen der erfaßten Dimensionen hin. Die diskriminante Validität muß damit als unzureichend bezeichnet werden. Entsprechend zeigt eine Faktorenanalyse über die Anforderungsmerkmale (Tab. 1) nahezu ausschließlich verfahrensspezifische Faktoren und keine Faktoren, die die verwendeten Dimensionen reproduzieren. Die Verfahrensfaktoren klären 79 % der Gesamtvarianz auf. Die drei Subskalen des Postkorbs, bei denen beobachterspezifische Halo-Effekte infolge

<sup>2</sup> Für die Auswahlpraxis von großer Bedeutung sind - denkt man an wiederholte Teilnahmen bei Assessment Centern und die zunehmende Verbreitung sogenannter „Testknacker“ - allerdings empirische Hinweise darauf, daß selbst durch ein einfaches Training zur Bearbeitung von Postkörben deutliche Übungsgewinne erzielt werden können (Brannick, Michaels & Baker, 1989).

der Auswertungs- und Interpretationsobjektivität nicht wirksam werden konnten, korrelierten untereinander mit  $r = .54$  deutlich geringer als die Skalen der übrigen Verfahren. Der Postkorb weist damit eine höhere diskriminante Validität als die anderen zum Einsatz gekommenen Übungen auf.

Tabelle 1: Faktorenanalyse über die Anforderungsmerkmale

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5
Mitarbeitergespräch: Führungsverhalten	.85				
Mitarbeitergespräch: Überzeugungskraft	.91				
Mitarbeitergespräch: Einfühlungsvermögen	.71				
Mitarbeitergespräch: Konfliktlöseverhalten	.89				
Mitarbeitergespräch: Initiative und Engagement	.84				
Mitarbeitergespräch: Flexibilität	.90				
Präsentation: Organisation und Planung		.79			
Präsentation: Überzeugungskraft		.90			
Präsentation: Initiative und Engagement		.95			
Präsentation: Flexibilität		.89			
Präsentation: Ausdrucksvermögen		.90			
Präsentation: Streßbewältigung		.94			
Gruppendiskussion: Führungsverhalten			.86		
Gruppendiskussion: Überzeugungskraft			.83		
Gruppendiskussion: Einfühlungsvermögen				.60	
Gruppendiskussion: Kooperationsverhalten			.84		
Gruppendiskussion: Konfliktlöseverhalten			.81		
Gruppendiskussion: Initiative und Engagement			.84		
Schriftliche Fallstudie: Analyseverhalten				.82	
Schriftliche Fallstudie: Organisation und Planung				.77	
Schriftliche Fallstudie: Entscheidungsverhalten				.72	
Schriftliche Fallstudie: Ausdrucksvermögen				.82	
Postkorb: Analyseverhalten					.80
Postkorb: Organisation und Planung					.72
Postkorb: Entscheidungsverhalten					.84
Postkorb: Streßbewältigung					.84
aufgeklärte Varianz	35,5 %	15,8 %	10,7 %	9,2 %	7,6 %

Anmerkungen: Ladungen auf orthogonalen Faktoren nach Hauptkomponentenanalyse und Varimaxrotation. Nur Ladungen  $>.6$  sind wiedergegeben.

In Tab. 2 ist die Korrelationsmatrix der im AC eingesetzten Verfahren wiedergegeben. Der Postkorb weist keine positive Korrelation zu den übrigen im AC eingesetzten Verfahren auf; mit dem Abschneiden bei der Präsentation besteht sogar eine signifikant negative Beziehung ( $r = -.27, p < .05$ ).

Tabelle 2: Korrelationsmatrix der im AC eingesetzten Verfahren

	Gruppen- diskussion	Mitarbeiter- gespräch	Präsentation	Fallstudie
Postkorb	-.19	-.07	-.27 *	-.22
Gruppendiskussion		.47 **	.31 *	.43 **
Mitarbeitergespräch			.25	.30 *
Präsentation				.37 **

Anmerkungen:  $N = 54$ ; \*  $p < .05$ ; \*\*  $p < .01$

Über diese Korrelationsmatrix wurde eine Hauptkomponentenanalyse durchgeführt. Nach Varimaxrotation ergaben sich zwei Faktoren, die zusammen 64% der Varianz aufklären (Tab. 3). Hohe Ladungen auf dem ersten Faktor weisen die Gruppendiskussion und das Mitarbeitergespräch auf. Dieser Faktor spiegelt demnach überwiegend soziale und kommunikative Kompetenzen wider. Der Postkorb lädt auf diesem Faktor nicht. Auf dem zweiten, orthogonalen Faktor laden die Präsentation und die Postkorbübung, und zwar mit entgegengesetzten Vorzeichen. Während die Präsentation auch auf dem ersten Faktor lädt, werden mit dem Postkorb offenbar spezifische und von anderen Verfahren nicht abgedeckte Varianzanteile erfaßt.

Tabelle 3: Faktorenanalyse über die eingesetzten Verfahren

	F1	F2
Gruppendiskussion	.79	.20
Mitarbeitergespräch	.83	-.06
Präsentation	.39	.61
Fallstudie	.59	.43
Postkorb	.06	-.87
aufgeklärte Varianz	44 %	20 %

Anmerkungen: Ladungen auf orthogonalen Faktoren nach Hauptkomponentenanalyse und Varimaxrotation.

Die eigenständige Bedeutung der Postkorbübung wird auch anhand der Korrelationen der Verfahren mit dem AC-Gesamtwert deutlich. Während die übrigen Verfahren auch mit dem um die Eigenkorrelation bereinigten Gesamtwert statistisch bedeutsam assoziiert sind, ist das Abschneiden im Postkorb signifikant negativ mit dem Abschneiden bei den übrigen Übungen korreliert ( $r = -.27, p < .05$ ). Ein erfolgreiches Abschneiden beim Postkorb geht



demnach mit einer weniger erfolgreichen Absolvierung anderer Übungen einher - und umgekehrt.

Tabelle 4: Korrelationen der Verfahren mit dem AC-Gesamtwert

Verfahren	AC-Gesamtwert	"bereinigter" AC-Gesamtwert
Postkorb	.09	-.27 *
Gruppendiskussion	.70 **	.46 **
Mitarbeitergespräch	.66 **	.42 **
Präsentation	.63 **	.28 *
Fallstudie	.70 **	.39 **

*Anmerkungen: Beim bereinigten Gesamtwert wurde der Einfluß des jeweiligen Verfahrens aus dem Gesamtwert eliminiert.*

*N = 54; \*  $p < .05$ ; \*\*  $p < .01$*

## 6 Zusammenfassung und Diskussion

Die Befunde von Sackett und Dreher (1982) und anderen zur Konstruktvalidität nach der Multitrait-Multimethod-(MTMM)-Methode wurden in der vorliegenden Untersuchung repliziert. Die verfahrensübergreifende Zusammenfassung von Eigenschaftsmessungen erscheint demnach aufgrund der mangelnden konvergenten und diskriminanten Validität fragwürdig. Auch im vorliegenden AC treten bei allen situativen Verfahren höhere Korrelationen zwischen den verschiedenen Eignungsdimensionen innerhalb eines Verfahrens auf als zwischen Werten zu derselben Eignungsdimension aus verschiedenen Verfahren. Nicht die Konstanz bei der Bewertung von Eignungsdimensionen über verschiedene Verfahren spiegelt sich also in den Daten, sondern eher die Ähnlichkeit der Dimensionen innerhalb der Verfahren.

Die Faktorenanalyse zeigt in Übereinstimmung damit deutlich verfahrensbezogene Faktoren anstelle von Faktoren, die die verwendeten Dimensionen reproduzieren. Während also die allen anderen Verfahren ebenbürtige oder überlegene Prognoseleistung von Assessment Centern unbestritten ist (Gaugler, Rosenthal, Thornton & Bentson, 1987), erweist sich die landläufige, traitorientierte Erklärung für diese Prognoseleistung als unzulänglich: Nach den Kriterien der Multitrait-Multimethod-Konstruktvalidierung ist die Assessment-Center-Technologie nicht valide (vgl. Sackett & Dreher, 1982, 1984).

Bei der Diskussion über die unbefriedigende Qualität von traitorientierten Messungen in Assessment Centern sollte allerdings nicht übersehen werden, daß hohe Korrelationen zwischen einzelnen Verhaltensereignissen aus zwei verschiedenen Situationen gar nicht zu erwarten sind, wenn spezifische Besonderheiten der situativen Übungen - die im AC ja die gesamte Breite von Management- und Führungsaufgaben abdecken sollen - die Konsistenz maskieren. Ohnehin wird der Bedeutung des Aggregationsprinzips für die Reliabilität der Messung von Dispositionskonstrukten in traitorientierten Assessment Centern nicht in ausreichendem Maße Rechnung getragen, wenn beispielsweise in einer Gruppendiskussion aus einer kleinen Zahl von Wortbeiträgen Ratings auf fünf oder mehr Eigenschaftsdimensionen

abgeleitet werden. In Anbetracht der „incredible demands“ (Bycio, Alvares & Hahn, 1987), die das Verfahren der freien Eindrucksbildung an die Beobachter stellt, bleibt diesen angesichts fehlender konkreter Vorgaben für die Bewertung kritischer Verhaltensweisen und aufgrund der bekannten Grenzen der menschlichen Informationsverarbeitungskapazität möglicherweise gar nichts anderes übrig, als sich ad hoc auf eine einfache, alltagstheoretische Personwahrnehmung zurückzuziehen (Maukisch, 1989). Die zu hohen Anforderungen an die Beobachter können potentiell durch die Verwendung von auswertungsobjektiven Verfahren wie dem hier beschriebenen Postkorb wirksam reduziert werden. Zwar trug im vorliegenden AC auch eine erhöhte Auswertungsobjektivität nur teilweise zu einer Erhöhung der MTMM-Konstruktvalidität bei; dabei ist jedoch zu berücksichtigen, daß lediglich ein Verfahren - der Postkorb - mit optimierten Objektivitätskennwerten zum Einsatz kam. Für dieses wurde eine deutlich erhöhte diskriminante Validität und damit eine geringere Anfälligkeit gegenüber Urteilsverzerrungen und Haloeffekten festgestellt. Im Verbund mit der faktoranalytisch bestimmten eigenständigen Stellung des auswertungsobjektiven Postkorbs gibt dies Anlaß zu der Hoffnung, daß durch Entwicklung weiterer auswertungs- und durchführungsobjektiver AC-Übungen eigenschaftsorientierte Assessment Center auf eine tragfähigere Grundlage gestellt werden können. Bei der Neuentwicklung von geeigneten AC-Verfahren wird deshalb besonders auf eine befriedigende Standardisierung in der Durchführung und Auswertung zu achten sein.

Vor dem Hintergrund der noch nicht endgültig geklärten Kontroverse um die eigenschaftstheoretische Fundierung von AC-Übungen hat sich in der theoretischen Diskussion weitgehend der Ansatz von Neubauer (1980) durchgesetzt, das AC als primär verhaltensorientiertes Beurteilungsinstrument aufzufassen. Managementerfolg wird dabei als Ergebnis einer Interaktion von Person und Situation verstanden: Die Beurteilung der Eignung erfolgt nicht über Dimensionen der Führungseignung, sondern situations- und anforderungsbezogen.

Für diese Konzeption spricht zunächst, daß sich der isolierte Einsatz eigenschaftstheoretisch fundierter psychologischer Tests zur Personalauswahl bislang als ungeeignet erwiesen hat, weil abgesehen von Intelligenztests (die aus verschiedenen Gründen zumindest in Deutschland in der betrieblichen Auswahldiagnostik eine immer geringere Rolle spielen) nur sehr niedrige Zusammenhänge zwischen den Testwerten und dem Erfolg im beruflichen Alltag gefunden wurden (Hess & Schmitt-Planert, 1985).

Gerade der Übergang vom Mitarbeiter zur Führungskraft ist oft durch tiefgreifend veränderte situative Anforderungen gekennzeichnet, die in dem häufigen Phänomen des Versagens ehemals hervorragender Sachbearbeiter als Führungskraft sichtbar werden. Wünschenswert ist deshalb aus einer verhaltenstheoretischen Sichtweise, die den Berufsalltag eines Managers kennzeichnenden Anforderungen stichprobenartig und möglichst repräsentativ mit verschiedenartigen Übungen abzubilden. Jaffee (1971) betont beispielsweise die Notwendigkeit der Ergänzung von Postkorbergebnissen durch führerlose Gruppendiskussionen, bei denen Führungspotential nicht nur theoretisch, sondern durch effektive Kommunikation mit anderen auch in der Praxis nachgewiesen werden muß. "Different situational exercises are not developed to simply provide multiple input on what Sackett and Dreher refer to as constructs. Multiple exercises are included in an attempt to adequately sample the relevant content domain of incumbent behavior" (Neidig & Neidig, 1984, S. 183).

Bei einer solchen Konzeption stellt man den durch die unbefriedigenden konvergenten und diskriminanten Validitäten genährten Bedenken gegen eine eigenschaftstheoretische Vorgehensweise die eher pragmatische Sichtweise gegenüber, daß es bei der Validierung an einem

Außenkriterium unwesentlich ist, ob Dimensions- oder Verfahrensfaktoren die Varianz in den Bewertungen verursachen (Schuler, 1989). Die verwendeten Verfahren werden also nicht länger dem Multitrait-Multimethod-Gedanken untergeordnet, um eigenschaftsbasierte Eignungsbeurteilungen durch den Einsatz mehrerer Verfahren abzusichern (Campbell & Fiske, 1959); sie werden vielmehr als notwendige Maßnahme eingesetzt, um *überhaupt* die wesentlichen Bereiche der Managementtätigkeit diagnostisch abdecken zu können. So betont Howard (1983), daß erst der Einsatz des Postkorbs dem Personalberater ein Instrument an die Hand gibt, um die mit anderen AC-Übungen nicht zu erfassenden "administrative skills" zu messen, die auch 20 Jahre nach der Messung noch immer statistisch bedeutsam ( $r = .19$ ) mit der erreichten Hierarchiestufe als Außenkriterium korrelieren. Unter diesem Gesichtspunkt, aber auch aufgrund der Notwendigkeit, differenzierte Empfehlungen für Weiterbildungsmaßnahmen im Anschluß an Personalentwicklungs-ACs abgeben zu können, ist der Einsatz möglichst vieler, vor allen Dingen aber verschiedenartiger AC-Übungen zu befürworten.

Rechtfertigung erfährt der Einsatz der Postkorbübung im AC dabei nicht nur durch seine erwiesene prädiktive Validität (siehe z.B. auch Brass & Oldham, 1976; Bray & Grant, 1966; Frederiksen, 1966; Meyer, 1970; Wollowick & McNamara, 1969), sondern auch durch die Beobachtung, daß der Postkorb offensichtlich von anderen Verfahren unabhängige Varianzanteile aufzuklären vermag (siehe Tab. 3). Hinweise auf einen eigenständigen Beitrag des Postkorbs finden sich auch bei Hinrichs und Haanpera (1976) sowie Neidig, Martin und Yates (1978). Bemerkenswert ist insbesondere die negative Korrelation zwischen dem Abschneiden beim Postkorb und dem Abschneiden in den übrigen Übungen: Hierbei werden offensichtlich gegenläufig verknüpfte Kompetenzaspekte erfaßt. Im Hinblick auf die Interpretation dieses Befundes ist eine Studie von Borman (1982) von besonderem Interesse, bei der der Postkorb als einziges Verfahren negativ mit den beiden Variablen "Erster Eindruck" und "Physische Attraktivität" korreliert war. Dies deutet darauf hin, daß der Postkorb ein wirksames Instrument sein kann, um Blender und Vertreter des von Neubauer (1980) beschriebenen AC-Erfolgstypus ("Vielredner mit eindrucksvollem Auftreten und Durchsetzungsvermögen", S. 154) zu identifizieren, die außer einem gewinnenden Wesen und kommunikativen Kompetenzen wenig Managementpotential erkennen lassen. Die hier geschilderte objektive Auswertungstechnik vermeidet dabei für das Postkorbverfahren die breiten Interpretationsspielräume, die andere komplexe situative Verfahren kennzeichnen und häufig zu Urteilsverzerrungen führen. Sie bietet ferner die Voraussetzung für die wünschenswerte Bildung standardisierter Normvergleichswerte.

Die am verhaltenstheoretisch-interaktionistischen Ansatz orientierte Assessment-Center-Technologie versteht sich letztlich als kriterienorientierte Messung berufsrelevanter Verhaltensweisen. Das zu erfassende Persönlichkeitsmerkmal wird dabei konzipiert als die „Eigenschaft eines Menschen, in einer gegebenen Klasse von Situationen eine bestimmte Klasse von Verhaltensweisen zu äußern“ (Klauer, 1987, S.13/14). Im Kontext des AC ist hierbei der Aufgabenbegriff von besonderer Bedeutung: Die kriterienorientierte Diagnostik steht und fällt mit der präzisen Definition der Aufgabenmenge. Nach Klauer (1987) ist für eine Aufgabe aber stets sowohl die Stimuluskomponente als auch die Responsekomponente entscheidend. Entgegen der allgemein verbreiteten AC-Praxis sollte die Erfassung der „Responses“ deshalb nicht allein auf freier Eindrucksbildung anhand beliebiger, frei flottierender Verhaltensweisen basieren, sondern auf einer objektiven Kodierung und Bewertung der zu beobachtenden Handlungen, unter Bezug auf eine Norm, die für eine genau definierte und standardisiert dargebotene Aufgabe gesetzt wird (vgl. Maukisch, 1989). Wie die durch die Aufgabensituation ausgelösten Verhaltensweisen zu bewerten sind, muß dabei notwen-

digerweise a priori und eindeutig festgelegt werden (Klauer, 1987). Denn bei einer fehlenden Strukturierung und Explikation des Bezugssystems, auf das hin Handlungen bewertet werden sollen, können beobachtete Verhaltensweisen kaum als „sample“ (Wernimont & Campbell, 1968) von Handlungen verstanden werden, die das Vorhandensein oder das Fehlen einer bestimmten Kompetenz (im Sinne Klauers) indizieren. Viele AC-Übungen leiden unter der Inkonsistenz und Beliebigkeit des Bewertungssystems und der mangelnden Standardisierung der Situationsvariablen; wer Intransparenz als varianzerweiterndes Mittel einsetzt, verkennt zudem, daß die so erzeugte Varianz Fehlervarianz ist.

Neben der Notwendigkeit einer möglichst weitgehenden Standardisierung des Bewertungsmaßstabes durch eine vorab eindeutige Festlegung der als korrekt zu bewertenden Handlungsalternativen lassen sich einige weitere Folgerungen für die AC-Praxis ziehen.

Zunächst scheint der Postkorb deutlich von anderen Verfahren zu unterscheidende Varianzanteile zu erfassen. Auch bei einer aus ökonomischen Gründen wünschenswerten Reduktion der Zahl der eingesetzten Übungen wird man deshalb neben einem die Kommunikationsfähigkeit erfassenden Verfahren wie dem Mitarbeitergespräch oder der Gruppendiskussion sinnvollerweise immer auch eine stärker analytisch geprägte Übung wie den Postkorb einsetzen.

Eine weitere interessante Einzelbeobachtung ergab die Auswertung der Trennschärfen aller eingesetzten Items. Das für das Postkorbverfahren schon fast klassische Item „Irrläufer“ - der Bearbeiter muß erkennen, daß ein Schriftstück nur irrtümlich an ihn geschickt wurde - wies nämlich mit  $r = -.17$  eine negative Trennschärfe auf. Allerdings war es mit einer relativen Lösungshäufigkeit von nur  $P = .07$  in einem extremen Schwierigkeitsbereich angesiedelt. Es wäre deshalb wünschenswert, die Qualität dieses Items in weiteren Untersuchungen genauer zu prüfen.

Zusammenfassend empfehlen wir den Einsatz der Postkorbübung als wichtige Ergänzung des Verfahrensarsenals einer verhaltenswissenschaftlich fundierten Assessment-Center-Technologie. Die Vorteile der Verhaltens- und Situationsbezogenheit sollten dabei jedoch nicht durch Inkaufnahme unkontrollierter Fehlerquellen aus der Hand gegeben werden. Auswertungsobjektive Postkörbe der hier beschriebenen Art erlauben eine kriterienorientierte Messung der Bewerberleistung in einer a priori definierten Aufgabe unter Zugrundelegung eines eindeutig definierten Bezugssystems.

## 7 Literatur

- Bentz, V. (1968). The Sears experience in the investigation, description and prediction of executive behavior. In J.A. Meyer (Ed.), *Predicting managerial success* (pp. 59-152). Ann Arbor, MI: Foundation for Research on Human Behavior.
- Bentz, V. (1971). Validity studies at Sears. *Symposium on Validity of Assessment Centers*. Washington DC: American Psychological Association.
- Borman, W. (1982). Validity of Behavioral Assessment for Predicting Military Recruiter Performance. *Journal of Applied Psychology*, 67, 3-9.
- Brannick, M., Michaels, C., & Baker, D. (1989). Construct Validity of In-Basket Scores. *Journal of Applied Psychology*, 74, 957-963.
- Brass, D., & Oldham, G. (1976). Validating an In-Basket Test Using an Alternative Set of Leadership Scoring Dimensions. *Journal of Applied Psychology*, 61, 652-657.
- Bray, D.W., & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, 80 (17, Whole No. 625).
- Bycio, P., Alvares, K., & Hahn, J. (1987). Situational Specificity in Assessment Center Ratings: A Confirmatory Factor Analysis. *Journal of Applied Psychology*, 72, 463-474.
- Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Didi, H.-J. (1992). *Konstruktvalidierung und Weiterentwicklung einer Postkorb-Aufgabe im Rahmen eines Assessment-Centers*. (Unveröffentlichte Diplomarbeit). Bonn: Psychologisches Institut der Universität Bonn.
- Dukerich, J.M., Milliken, F.J., & Cowan, D.A. (1990). In-basket exercises as a methodology for studying information processing. *Simulation & Gaming*, 21(4), 397-410.
- Frederiksen, N. (1962). Factors in In-Basket performance. *Psychological Monographs: General & Applied*, 76 (Whole No. 541).
- Frederiksen, N. (1966). Validation of a simulation technique. *Organizational Behavior and Human Performance*, 1, 87-109.
- Frederiksen, N., Saunders, D.R., & Wand, B. (1957). The In-Basket test. *Psychological Monographs: General & Applied*, 71 (Whole No. 438).
- Funke, J. (1993). Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 119-129.
- Funke, J., & Rasche, B. (1988). *PC-Postkorb. Manual zur PC-Version der Postkorb-Übung*. (Unveröffentlichtes Manuskript). Bonn: Psychologisches Institut der Universität Bonn.
- Gaugler, B., Rosenthal, D., Thornton, G., & Bentson, C. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology*, 72, 493-511.
- Ginsbury, L., & Silverman, A. (1972). The leaders of tomorrow: Their identification and development. *Personnel Journal*, 51(9), 662-666.
- Hemphill, J.K. (1961). The in-basket technique: Overview. In Educational Testing Service (Ed.), *Proceedings of the Conference on the Executive Study* (pp. 1-11). Princeton, N.J.
- Hess, U., & Schmitt-Planert, A. (1985). Das Assessment Center, ein eignungsdiagnostisches Hilfsmittel für die betriebliche Praxis - Verlauf einer Entdeckung. In W.F. Kugemann, S. Preiser & K.A. Schneewind (Hrsg.), *Psychologie und komplexe Lebenswirklichkeit. Festschrift zum 65. Geburtstag von Walter Toman*. Göttingen: Hogrefe.
- Hinrichs, J.R., & Haanpera, S. (1976). Reliability of measurement in situational exercises: An assessment of the assessment center method. *Personnel Psychology*, 29, 31-40.
- Horn, R. (1991). Mailbox: A computerized in-basket task for use in personnel selection. *European Review of Applied Psychology*, 41, 325-327.

- Howard, A. (1983). Measuring management abilities and motivation. In R.B. Ekstrom (Ed.), *Measurement, technology, and individuality in education. New directions for testing and measurement*, 17, 31-44. San Francisco: Jossey-Bass.
- Jaffee, C.L. (1971). *Effective management selection: The analysis of behavior by simulation techniques*. Reading: Addison-Wesley.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern. Assessment-Center-Verfahren*. München: Hanser.
- Kesselman, G., Lopez, F., & Lopez, F. (1982). The Development and Validation of a Self-Report In-Basket Test in an Assessment Center Setting. *Public Personnel Management*, 11, 228-238.
- Klauer, K.J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Kliem, O. (1983). Die Postkorbübung als Test- und Trainingsinstrument. *Personal*, 5, 193-196.
- Klieme, E., & Stumpf, H. (1990). Computereinsatz in der pädagogisch-psychologischen Diagnostik. In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 8. Jahrbuch der Pädagogischen Diagnostik*, 13-63. Weinheim: Beltz.
- Leichner, R. (1979). *Psychologische Diagnostik - Grundlagen, Kontroversen, Praxisprobleme*. Weinheim: Beltz.
- Lopez, F.M., Jr. (1966). *Evaluating executive decision making: The in-basket technique*. AMA Research Study 75. New York: American Management Association, Inc.
- Maukisch, H. (1989). Informationswert und Ökonomie der diagnostischen Prinzipien von Assessment Center Systemen zur Erfassung von Management Potential. In C. Lattmann (Hrsg.), *Das Assessment-Center-Verfahren der Eignungsbeurteilung*, 251-291. Heidelberg: Physica.
- Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meyer, H.H. (1970). The validity of the In-Basket test as a measure of managerial performance. *Personnel Psychology*, 23, 297-307.
- Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.
- Mitchell, J.O. (1975). Assessment center validity: A longitudinal study. *Journal of Applied Psychology*, 60, 573-579.
- Neidig, R.D., Martin, J.C., & Yates, R.E. (1978). *The FBI's Management Aptitude Program assessment center: Research report no. 1* (TM-78-3). Washington, DC: U.S. Civil Service Commission. (NTIS No. PB-286 181).
- Neidig, R., & Neidig, P. (1984). Multiple Assessment Center Exercises and Job Relatedness. *Journal of Applied Psychology*, 69, 182-186.
- Neubauer, R. (1980). Die Assessment Center Technik. Ein verhaltensorientierter Ansatz zur Führungskräfteauswahl. In R. Neubauer & L.v. Rosenstiel (Hrsg.), *Handbuch der angewandten Psychologie*, 122-158. München: Moderne Industrie.
- Richards, S.A., & Jaffee, C.L. (1972). Blacks supervising whites: A study of interracial difficulties in working together in a simulated organization. *Journal of Applied Psychology*, 56, 234-240.
- Robertson, I., Gratton, L., & Sharpley, D. (1987). The Psychometric Properties and Design of Managerial Assessment Centers: Dimensions into Exercises won't go. *Journal of Occupational Psychology*, 60, 187-195.
- Roest, F., & Horn, R. (1990). Mailbox-90: Computerunterstützte Diagnostik im Assessment-Center. *Diagnostica*, 36, 213-219.
- Sackett, P. (1987). Assessment centers and content validity: Some neglected issues. *Personnel Psychology*, 40, 13-25.

- Sackett, P., & Dreher, G. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401-410.
- Sackett, P., & Dreher, G. (1984). Situation Specificity of Behavior and Assessment Center Validation Strategies: A Rejoinder to Neidig and Neidig. *Journal of Applied Psychology*, 69, 187-190.
- Schippmann, J., Prien, E., & Katz, J. (1990). Reliability and Validity of In-Basket performance measures. *Personnel Psychology*, 43, 837-859.
- Schmitt, N., & Noe, R.A. (1983). Demonstration of content validity: Assessment center example. *Assessment Center Technology*, 6, 5-11.
- Schuler, H. (1989). Die Validität des Assessment Centers. In C. Lattmann (Hrsg.), *Das Assessment-Center-Verfahren der Eignungsbeurteilung*, 223-250. Heidelberg: Physica.
- Silverman, W.H., Dalessio, A., Woods, S.B., & Johnson, R.L. (1986). Influence of assessment center methods on assessor's ratings. *Personnel Psychology*, 39, 565-578.
- Strunz, C. (1994). *Assessment Center Technik: Postkorb „Mobilfunk-Hotline Fasanenhof“*. Unveröffentlichtes Manuskript.
- Thornton, G., & Byham, W. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Turnage, J.J., & Muchinsky, P.M. (1984). A comparison of the predictive validity of assessment center evaluations versus traditional measures in forecasting supervisory job performance: Interpretative implications of criterion distortion for the assessment center paradigm. *Journal of Applied Psychology*, 69, 595-602.
- Ward, L. (1959). *The business in-basket test: A method of assessing certain administrative skills*. (Research Bulletin RB-59-8). Princeton, NJ: Educational Testing Service.
- Ward, L. (1960). The performance of business school students on an in-basket test. In J. Hemphill (Ed.), *Proceedings of the conference on the executive study: The in-basket technique* (pp.91-105). Princeton, NJ: Educational Testing Service.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.
- Wernimont, P.F., & Campbell, J.P. (1968). Signs, samples and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Wollowick, H., & McNamara, W. (1969). Relationship of the components of an assessment center to management success. *Journal of Applied Psychology*, 53, 348-352.