

Aus der Arbeit des Testkuratoriums

MAILBOX '90

Ein computergestütztes Test- und Trainingsverfahren zur Personalentwicklung *

Joachim Funke

1. Testgrundlage

1.1 Diagnostische Zielsetzung

MAILBOX '90 von Roest, Scherzer, Urban, Gangl und Brandstätter (1989)¹ ist ein computergestütztes Verfahren zur Berufseignungsdiagnostik, bei dem die „Simulation einer möglichst realitätsnahen Umgebung...Rückschlüsse auf das entsprechende Arbeitsverhalten des Bewerbers“ erlauben soll. MAILBOX '90 realisiert zu diesem Zweck eine in Assessment-Centern gebräuchliche Aufgabenstellung vom Typ „Postkorb“ (aktuelle Übersichten: Schippmann, Prien & Katz, 1990; Schmidt, Ones & Hunter, 1992), bei der unter Zeitdruck eine Reihe unsortierter Dokumente durchzusehen und hinsichtlich eventuell notwendiger Entscheidungen zu bewerten ist. Da eines der Hauptprobleme in der praktischen Anwendung von Postkörben die anschließende, meist personal- und zeitintensive Auswertung darstellt, verwundert es wenig, daß zunehmend Anstrengungen unternommen werden, dieses Verfahren in eine computergestützte Version zu überführen, um damit einige der mit einem Rechneinsatz verbundenen Vorteile nutzen zu können.

Ziel der vorliegenden Implementation eines Postkorbs ist eine Objektivierung der Instruktion, des Verfahrensablaufs und der Auswertung durch folgende Aspekte (vgl. auch Horn, 1991; Roest & Horn, 1990, S. 215): (1) Aufhebung potentieller Chancenungleichheit für „Computerunerfahrene“ durch ausführliche Instruktion und Übungsphase; (2) Bereitstellung einer kontextsensitiven Hilfefunktion bei Bedienungsproblemen; (3) eingebaute Störungen des Arbeitsablaufs; (4) verzögertes Feedback bei falscher Delegation; (5) Protokollierung des exakten Testablaufes sowie automatische Vergabe von Positiv- und Negativpunkten pro bearbeiteter Aufgabe.

1.2 Theoretische Grundlagen

MAILBOX '90 basiert wie alle Postkorb-Verfahren auf dem klassischen Prinzip der Arbeitsprobe (vgl. Funke, 1993). Die Testperson sieht sich mit einem gefüllten Postkorb konfrontiert, dessen Inhalt in begrenzter Zeit und mit begrenzten Ressourcen bearbeitet werden muß. MAILBOX '90 besteht aus neun Vorgängen (Dokumenten), die innerhalb von 32 Minuten mausgesteuert am Rechner durchzugehen sind.

Die Grundkonzeption des Verfahrens zeigt sich in der Ableitung der Dimensionen, die die Testautoren zur Bewertung des Bewerberverhaltens heranziehen. Insgesamt sieben Skalen werden aufgrund inhaltlicher Überlegungen definiert, die ersten vier aufgabenorientiert, die drei folgenden verlaufsorientiert (vgl. S. 16ff).

* Roest, F., Scherzer, A., Urban, E., Gangl, H. & Brandstätter, C. (1989). MAILBOX '90. Ein computergestütztes Test- und Trainingsverfahren zur Personalentwicklung. Wien und Weinheim: Scicon Datenverarbeitungsgesellschaft mbH und Beltz Test GmbH. Preis für Einzelplatzversion: 9.000,- DM; Demo-Diskette: 100,- DM (wird bei späterem Kauf angerechnet).

¹ Alle Seitenangaben ohne näheren Quellenvermerk beziehen sich auf dieses Manual. Das dem Rezensenten vorgelegte Programm trägt die Versionsnummer 4.2.

Die vier aufgabenorientierten Skalen sind im einzelnen: (1) *Arbeitszeit*: Hier wird pro Dokument die Bearbeitungszeit gemessen; (2) *Arbeitseinteilung*: Auszählung der Häufigkeit, mit der jedes Dokument bearbeitet wurde (= eventuelles Wiederaufgreifen schon gelesener Vorgänge); (3) *Zielorientierung*: richtig/falsch-Klassifizierung jedes Dokuments hinsichtlich eines von den Testautoren vordefinierten Endzustands, in dem es sich am Testende befinden sollte; (4) *Arbeitsgüte*: pro Dokument werden 1 bis 4 Punkte gegeben für bestimmte damit zu verrichtende Tätigkeiten (z. B. je 1 Punkt für Einsichtnahme in das Dokument, Weiterleitung an eine andere Person, Überprüfung eines Termins im Kalender).

Die drei verkaufsorientierten Skalen unterteilen die halbstündige Gesamtbearbeitungszeit in sechs Fünf-Minuten-Intervalle, in denen jeweils die Frequenz bestimmter Ereignisse ausgemittelt wird: (5) *Aktivität*: Häufigkeit der Verwendung aller für die Dokumentenbearbeitung möglichen Funktionen (inklusive Seitenumblättern, etc.); (6) *Delegieren*: Häufigkeit der im jeweiligen Zeitabschnitt vorgenommenen Delegationen an andere Personen; (7) *Ordnen*: Häufigkeit, mit der die Option „Dokumente ordnen“ genutzt wird.

Wie ersichtlich wird, liegt damit keine spezifische Fundierung des Verfahrens „auf den Ergebnissen der wissenschaftlichen Psychologie“ (Testkuratorium, 1986) vor. Vielmehr wird – wie bei anderen Postkörben auch – die Augenschein-Validität des Verfahrens als Rechtfertigung für dessen Konstruktion herangezogen (S. 8).

1.3 Nachvollziehbarkeit der Testkonstruktion

Über die Konstruktion von MAILBOX '90 wird im Manual nichts weiter berichtet außer dem Hinweis darauf, daß „zunächst ein inhaltsorientierter Testkonstruktionsansatz im Vordergrund“ stand (S. 14). Wie das Material konkret ausgewählt und ob es modifiziert wurde (und falls ja, aufgrund welcher Befunde), ist dem Manual wie auch den beiden anderen Veröffentlichungen zum Verfahren nicht zu entnehmen. Die vom Testkuratorium (1986) geforderte „Nachvollziehbarkeit der einzelnen Arbeitsschritte bei der Erstellung des Testmaterials“ ist damit nicht erfüllt.

2. Testdurchführung

2.1 Durchführungsobjektivität

Durch seine vollständige Darbietung auf dem Bildschirm eines Rechners und durch die rechnergestützte Ablaufkontrolle besitzt MAILBOX '90 maximale Durchführungsobjektivität. Sowohl die Vorabinstruktion zur Rechnerhandhabung als auch die kurze Instruktion zum Szenario selbst sind verständlich abgefaßt, verlangen allerdings von der Testperson den Erwerb vieler neuer Konzepte (wie z. B. Arbeitsfläche, Befehlsspalte, Statuszeile, Hilfsmittel, Anweisung, Meldung; auch die Bedienung von Rechner, Notizblock und Kalender muß gelernt werden). Alle Texte werden direkt am Bildschirm vorgegeben. Die Kurzinformation benötigen nach Angaben der Autoren etwa 18–20 Minuten, die längere Version 30–35 Minuten.

2.2 Transparenz

MAILBOX '90 ist der Testperson gegenüber offen in Hinblick auf die gestellten Anforderungen. Da es sich um einen Leistungstest unter Zeitdruck handelt, weiß die getestete Person, worauf es im wesentlichen ankommt, ohne daß im Detail erkennbar wäre, welche Entscheidungsalternative die richtig bewertete ist. Die Instruktion verlangt, in der gegebenen Zeit möglichst viele Dokumente zu bearbeiten, macht auf die Delegationsmöglichkeiten aufmerksam und erläutert das Prinzip der Dokumentenrückgabe bei fehlerhafter Bearbeitung bzw. falscher Delegation. Hinweise auf die Auswertung können ansonsten dem Verfahren selbst nicht entnommen werden. Intransparent ist das Verfahren insofern, als die Vorabinstruktion explizit ohne zeitliche Begrenzung vorgegeben wird, jedoch aus der benötigten Zeit Rückschlüsse auf die „Auffassungsgabe“ der Person gezogen werden (S. 79).

2.3 Zumutbarkeit

Die getestete Person wird mit MAILBOX '90 zeitlich, psychisch und körperlich in zumutbarer Weise belastet. Potentielle Widerstände gegenüber dem Medium Computer sind denkbar, allerdings sprechen Akzeptanzstudien (z. B. Schuler, 1990) für einen eher vorteilhaften Effekt dieses Instruments.

2.4 Verfälschbarkeit

Da es sich nicht um Einstellungsfragen handelt, sondern Entscheidungen auf Verhaltens-ebene gefordert sind, erscheint die Verfälschbarkeit des Verfahrens gering. Die in anderen Postkörben mögliche Verfälschung durch Antworten im Sinne sozialer Erwünschtheit auf Fragen, bei denen etwa im Konflikt zwischen Dienst- und Privatangelegenheiten Prioritäten zu setzen sind, ist bei MAILBOX '90 auf ein Minimum reduziert, da hier bis auf eine Ausnahme nur firmenrelevante Vorgänge aufgenommen wurden.

2.5 Störanfälligkeit

Das Verfahren verlangt infolge des vorgegebenen Zeitdrucks eine von äußeren Störungen freie, konzentrierte Bearbeitung durch die Testperson. Die eingebauten Störungen in Form simulierter Anrufe und zurückgegebener Dokumente stellen einen kontrollierten Test der Belastbarkeit dar, der die Reaktion auf derartige Überraschungen zu prüfen gestattet. Eine spezifische Auswertung dieser Reaktionen erfolgt allerdings nicht standardmäßig.

3. Testverwertung

3.1 Auswertungsobjektivität

Die Auswertung der in einer Protokolldatei abgelegten Eingriffsaktivitäten erfolgt nach Aufruf eines separaten Programms (wird dies nicht sofort im Anschluß an die Datenerhebung aufgerufen, muß die Protokolldatei zunächst umbenannt werden, um versehentliches Überschreiben zu verhindern). Als Auswertungsoption kann zwischen einer Kurzauswertung (siehe einen Ausschnitt davon in Abb. 1) und einer ausführlichen Auswertung gewählt werden, bei der zusätzlich ein minutiöses Protokoll aller Probanden-Eingriffe ausgedruckt wird.

Neben einer Statistik der einzelnen Vorgänge werden die Rohwerte zu jedem Vorgang ausgegeben und grafisch als Balkendiagramm veranschaulicht; zusätzlich werden die Rohwerte in Staninewerte transformiert². Hohe Staninewerte indizieren offensichtlich hohe Leistungen. Außerdem wird für die drei verkaufsorientierten Skalen ein Balkendiagramm des Leistungsverlaufs ausgegeben. Da die Auswertung vollständig unter Kontrolle des Programms steht und sich auf die von der Testperson erzeugte Protokolldatei bezieht, ist die Auswertungsobjektivität maximal.

3.2 Zuverlässigkeit

Zur Überprüfung von Reliabilität und Validität des Verfahrens wurden Erhebungen an 105 Grazer Psychologie-Studenten sowie an 193 Bewerbern für eine Trainee-Position in einer Firma für Unternehmensberatung vorgenommen³. Von 159 der 193 Teilnehmenden liegen die Urteile des umfassenderen Auswahlverfahrens vor (73 angenommen, 86 abgelehnt). Inwiefern das Postkorb-Ergebnis in die Auswahlentscheidung einfließt, bleibt unklar. Methodisch wurde so vorgegangen, daß für jede der sieben Skalen Clusteranalysen über die Gesamtstichprobe

² Es wäre sicher sinnvoller, anstelle der Rohwerte die Staninewerte im Balkendiagramm zu veranschaulichen (vgl. Abb. 1).

³ Ich übernehme hier die männliche Form, obwohl aus dem Text nicht eindeutig hervorgeht, ob wirklich nur männliche Studierende bzw. nur Bewerber untersucht wurden.

aus testtheoretischer Sicht kritisch bewertet werden, da schon bei Wechsel des Mediums von Papier-Bleistift auf Computer ein Übertragen der Testgütekennwerte nicht ohne weiteres vorgenommen werden darf (vgl. Klieme & Stumpf, 1990). Ein Wechsel der Inhalte muß daher zwangsläufig eine erneute Überprüfung der Kennwerte nach sich ziehen.

3.6 Informationsausschöpfung

Die in der Protokolldatei enthaltene Information wird naturgemäß nur partiell ausgeschöpft. Allerdings liegt gerade hierin die Crux des Verfahrens: die Kondensation dieser vielen protokollierten Kleindetails (wie z.B. dem genauen Zeitpunkt, zu dem eine bestimmte Seite eines Dokuments aufgerufen und auch wieder verlassen wurde) zu bedeutungshaltigen Indikatoren des Arbeitsstils und der Arbeitsgüte erweist sich als schwieriges Unterfangen.

Die Verlaufsangaben der drei Skalen „Aktivität“, „Delegieren“ und „Ordnen“ sind nicht einfach zu interpretieren, verlangt doch nicht jeder Vorgang in gleicher Weise etwa eine Delegation. Dies hängt stark von der-freigestellten-Reihenfolge der Bearbeitung ab. Die in Abb. 1 gezeigte Seite beschreibt das Ergebnis einer Testperson, die nach etwa 20 Minuten die Bearbeitung der Dokumente beendet hatte und in der verbleibenden Zeit (das Programm endet nicht vor Ablauf der 32 Minuten) sich aus Langeweile verschiedene Dokumente noch einmal durchgesehen hat – der Arbeitsverlauf in den Intervallen 5 und 6 steigt daraufhin auf einen Spitzenwert!

Für den Diagnostiker entsteht damit eine schwierige Situation: Auf der Basis des kompakten zweiseitigen Ergebnisausdrucks ist eine direkte Beurteilung der Testperson nicht möglich. Da die Konfiguration der vielen Kennwerte aber ohne größere Erfahrung nicht sinnvoll interpretiert werden kann, erweist sich die Informationsausschöpfung als zu umfangreich. Die verschiedenen Aggregationsebenen (einzelne Dokumente und Funktionen; aufsummierte Werte für Dokumente; gemittelte Werte über alle Dokumente; nach Arbeitsphasen getrennte Kennwerte) erschweren die Arbeit der Testinterpretation.

3.7 Änderungssensitivität

Hierzu liegen derzeit keine Informationen vor. Eine wiederholte Bearbeitung von MAILBOX '90 sollte allerdings bei lernfähigen Pbn zu Leistungssteigerungen führen. Da im Untertitel des Tests explizit von einem „Trainingsverfahren“ gesprochen wird, überrascht die Nichtbehandlung dieses Themas im Manual bzw. in den vorliegenden Veröffentlichungen. Inwiefern MAILBOX '90 zum Training verwendet werden kann und woran Trainingserfolge abzulesen sind, müssen die Anwender selbst herausfinden.

4. Testevaluation

4.1 Ökonomie

Inwiefern das Verfahren als „kostengünstig“ zu bezeichnen ist, hängt von den Bewertungsgrundlagen ab: Für eine normale psychologische Praxis oder eine öffentliche Beratungsstelle ist dieses Verfahren sicher als kostenaufwendig zu bezeichnen. Da ein Postkorb wie MAILBOX '90 jedoch häufig als Bestandteil eines Assessment Centers im Rahmen von Maßnahmen zur Personalauswahl oder Personalentwicklung Verwendung findet, kann man dafür berichtete Zahlen zur Bewertung heranziehen. Nach einer von Barell (1992, S. 250f.) vorgenommenen Kosten-Nutzen-Rechnung eines Assessment Centers der Schweizerischen Kreditanstalt kommt es dort z. B. zu einer verbesserten Führungsleistung von 120.000,- SFr pro beförderten Mitarbeiter. An diesem Assessment ist ein Postkorb nicht unmaßgeblich beteiligt. Vor dem Hintergrund dieser Kosten-Nutzen-Betrachtung ist der Produktpreis sicherlich angemessen, zumal unbegrenzt viele Anwendungen des Verfahrens möglich sind.

4.2 Fairneß

Computererfahrene Personen können insbesondere durch die erforderliche Mausbedienung einen Geschwindigkeitsvorteil haben. Diese Gefahr sehen auch die Testautoren und empfehlen bei Computernovizen die vorherige Durchführung eines Maustrainings. Die von den Testautoren durchgeführten Vergleiche zwischen EDV-erfahrenen bzw. EDV-unerfahrenen Personen liefern jedoch in beiden vorliegenden Stichproben keinerlei signifikante Unterschiede.

Daß Mailbox '90 in Abhängigkeit von der Geschlechtsangabe der Person die geschlechtspassende Variante auswählt, sei lobend erwähnt, da es für Postkorb-Bearbeiterinnen erfahrungsgemäß häufig Identifikationsprobleme mit der fiktiven, standardmäßig männlichen Hauptrolle gab.

Welchen Einfluß Vorerfahrungen mit anderen Postkörben haben, ist unklar; allerdings muß man annehmen, daß bestimmte Verhaltensweisen (z.B. Delegation, Kalendernutzung) bei Assessment-Center-erfahrenen Personen häufiger gezeigt werden. Diesem mit zunehmender Verbreitung des Verfahrenstyps wachsenden Problem sollte entsprechende Forschungsaktivität gewidmet werden.

4.3 Akzeptanz

Angaben zur Akzeptanz des Verfahrens finden sich im Manual und in den Veröffentlichungen zum Verfahren nicht. In den einleitenden Bemerkungen des Manuals wird auf die hohe „face validity“ und auf die generell hohe Akzeptanz der Postkorb-Methode verwiesen. Nach einer von Schuler (1990, S. 190) vorgenommenen zusammenfassenden Bewertung verschiedener Untersuchungen zu dieser Frage werden tatsächlich gerade Arbeitsproben und Simulationen von Teilnehmern eines Assessment Centers als transparent und belastend erlebt sowie insgesamt positiv bewertet. Klieme und Stumpf (1990, S. 30) meinen sogar, von einer generell besseren Akzeptanz computerisierter Verfahren sprechen zu können. Mit diesen allgemeinen Aussagen im Hintergrund dürfte die Akzeptanz des Verfahrens als positiv angenommen werden. Demnach sollten empirische Belege speziell zur Akzeptanz dieses Verfahrens nachgeliefert werden.

4.4 Vergleichbarkeit

Das Verfahren stimmt von seiner Konstruktion her im wesentlichen mit dem Aufbau klassischer Postkörbe (vgl. Jeserich, 1981) überein. Direkt vergleichbar ist MAILBOX '90 den ebenfalls computergestützten Verfahren OFFICE von Fennekels und Nathan (1992), PC-POSTKORB von Funke und Rasche (1988) sowie PC-POSTKORB von Scharley (1992). Vergleichbare veröffentlichte Papier-Versionen stammen von Hasselmann (1992; POSTMAPPE) sowie Jeserich (im Druck; POSTKORB-ÜBUNG).

Zu beachten ist allerdings, daß allein durch die computerisierte Darbietung eines Leistungstests völlig andere Arbeitsbedingungen als bei einer Papier-Bleistift-Version vorliegen (vgl. Klieme & Stumpf, 1990, S. 25): das Durchblättern und Vorsortieren geht mit Zetteln einfach rascher und ist anschaulicher als das Blättern am Bildschirm.

4.5 Bewährung

Außer von den Testautoren selbst liegen derzeit keine weiteren Veröffentlichungen vor, aus denen auf die Bewährung des Verfahrens geschlossen werden könnte. Firmeninterne Bewertungen, die eventuell existieren könnten, dringen jedoch nur selten bis in Fachzeitschriften vor. Hier sollten Erfahrungsberichte unbedingt veröffentlicht werden.

5. Äußere Testgestaltung

Testleiter-Material. Neben dem knapp 100seitigen Manual erhält der Anwender zwei Dreieinhalb-Zoll-Disketten, auf denen sich die Software mitsamt einem Installationsprogramm befindet. Zusätzlich gehört ein Kopierschutzstecker („hardlock“) zum Lieferumfang, der ein

unberechtigtes Betreiben der Software verhindert. Das Manual berichtet ausführlich über die beiden durchgeführten Untersuchungen an Studenten und Trainees, beschreibt die Aufgabenstellung im Detail, gibt Hinweise zur Installation der Software und zu ihrer Handhabung. Außerdem wird das Auswertungsprogramm beschrieben und es wird erläutert, wie der Ergebnisausdruck zu interpretieren ist. Dieser Teil fällt allerdings vergleichsweise kurz und unbefriedigend aus (das Installationskapitel ist umfangreicher). So heißt es etwa auf S. 77 bezüglich der Interpretation der tabellarisch aufgeführten Verwendungshäufigkeiten der verschiedenen Befehle und der Rohwerte zu den Aufgaben: „Außer bei der Arbeitsgüte ... kann allerdings keine einheitliche Bewertung dieser Daten vorgegeben werden.“ Entsprechend verwirrend sind die gegebenen Hinweise: „Die Verwendung des Terminkalenders gibt Aufschluß darüber, ob der Bewerber sich über vorhandene Termine informiert hat“ (S. 78). „Zu beachten ist auch, in welcher Reihenfolge die Dokumente bearbeitet und welche Prioritäten gesetzt wurden“ (S. 78). „Ein wichtiges Kriterium für die Auffassungsgabe des Bewerbers ist die Bearbeitungszeit der Instruktion. Auch die Verteilung der Bearbeitungszeit über die Dokumente liefert Hinweise“ (S. 79). Diese Hinweise sind keine guten Entscheidungshilfen für den Testleiter, der mit einem unübersichtlichen Ergebnisprotokoll alleingelassen wird. Der hohen Auswertungsobjektivität tritt somit eine hohe Interpretationsunsicherheit entgegen.

Benutzeroberfläche. Den Testpersonen stehen auf dem Bildschirm ein einfacher Taschenrechner, ein Notizblock und ein Kalender zur Verfügung. Zusätzliche Informationen können sie aus einer Kundenliste sowie einem Organigramm des Unternehmens entnehmen. Die Bedienung erfolgt mausgesteuert, gehorcht aber nicht dem ergonomischen Prinzip der „direct manipulation“. Neben dem Eingangskorb stehen auf dem Bildschirm zwei weitere Ablagekörbe zur Verfügung (vgl. Abb. 2).

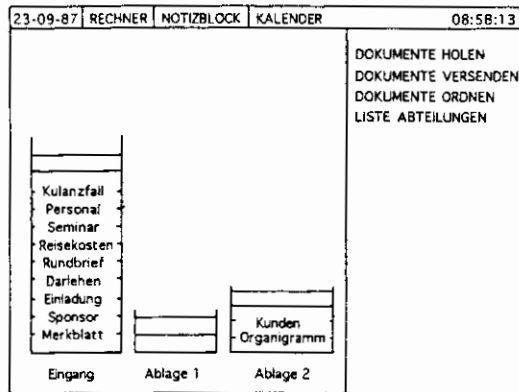


Abbildung 2: Bildschirmdisplay des simulierten Schreibtisches mit verschiedenen Körben; in der Kopfzeile: Hilfsmittel; rechts: Eingriffsmöglichkeiten.

Außerdem steht ein kontextsensitives Hilffsystem bereit, mit dem sich die Testpersonen während der Bearbeitung Informationen über die Bedienung verschaffen können. Dies wird auf die Testzeit angerechnet; die vielen Hilfstexte dürften daher nur selten zum Einsatz kommen.

Installation und Auswertung. Bei der Installation des Programms ist sehr genau den Vorschriften des Manuals zu folgen. Eine erste Installation scheiterte daran, daß der Rezensent eine handelsübliche und angeblich Microsoft-kompatible Genius-Maus nicht verwenden konnte. Bei der Installation des Kopierschutzsteckers ist darauf zu achten, daß an diesen Stecker zusätzlich angeschlossene Peripherie-Geräte eingeschaltet werden müssen, andernfalls bricht das Programm mit Fehlermeldung ab (wird nicht im Manual erwähnt). Bei einer der probeweise durchgeführten Testanwendungen geriet das Testprogramm nach fünf Minuten in einen blockierten Zustand, der nur noch durch Ausschalten des Computers zu überwinden war. In einem anderen Testlauf wurde das eilige Dokument des Stellenvorgängers über Unregelmäßigkeiten in der Finanzabteilung nicht angezeigt. In anderen Probeläufen erschien der bereits angekündigte Blanko-Scheck nicht. Auch das Auswertungsprogramm hatte Schwierigkeiten, mit den Non-Standard-Verzeichnissen des Rezensenten umzugehen, und ist nicht sehr benutzerfreundlich gestaltet. Hier sollten die Softwareproduzenten Verbesserungen vornehmen.

6. Abschließende Bewertung

Wie bereits deutlich geworden sein dürfte, ist das Verfahren hinsichtlich psychometrischer Ansprüche nicht völlig zufriedenstellend. Dies gilt sowohl für die Konstruktion des Verfahrens als auch für die Ableitung von Leistungsparametern, deren Reliabilität und Validität noch weitgehend ungeprüft ist. Schippmann, Prien und Katz (1990) sehen einen Grund für diesen unbefriedigenden Zustand darin, daß die hohe Augenschein-Validität dieser Art von Arbeitsproben psychometrische Analysen fast überflüssig erscheinen lasse. Die von diesen Autoren nach Durchsicht vorliegender Postkorb-Studien gemachte „final conclusion that the technique has become institutionalized based more on belief and conviction than on compelling evidence from the research“ (Schippmann et al., 1990, S. 857) gilt zwar zunächst nur für Papier-Bleistift-Versionen, aber die oben gegebene Darstellung rechtfertigt wohl die Ausdehnung dieser Argumentation auf Computerversionen. Natürlich ist auch zu beachten, daß die hier vorgetragene Kritik hinsichtlich fehlender teststatistischer Informationen auf die meisten Papier-Bleistift-Varianten gleichfalls Anwendung findet und daher kein Spezifikum der Computervariante darstellt.

bleibt die Frage, wie die Testkonstruktion in ihrem jetzigen Stadium zu bewerten ist. Ganz offensichtlich haben die Autoren von MAILBOX '90 im Auge gehabt, eine vereinfachte Präsentation und vereinfachte Auswertung anzubieten, ohne den Charakter der Übung dabei allzu sehr zu verändern. Im Vordergrund steht das Bemühen um eine möglichst direkte Übertragung der Papier-Bleistift-Vorgehensweise auf den Rechner. Einige medienspezifische Anpassungen (simulierter Telefonanruf; Rückgabe falsch bearbeiteter oder falsch delegierter Dokumente nach einer gewissen Zeit) sind dabei gemacht worden. Von den neuen Möglichkeiten des Mediums wird somit ansatzweise Gebrauch gemacht, vor allem von der Möglichkeit der raschen Ergebnisrückmeldung. Dieser Vorteil ist jedoch nur dann einer, wenn die Ergebnisse tatsächlich für eine Entscheidung hilfreich sind.

Die Ableitung von Parametern ist jedoch noch nicht zufriedenstellend begründet (warum z. B. bei den verlaufsorientierten Skalen sechs Zeitabschnitte und nicht drei, fünf oder zehn? Warum sollte die Häufigkeit, mit der eine Umblätter-Funktion genutzt wird, etwas mit Aktivität im Sinne einer Führungseigenschaft zu tun haben? etc.). Hier liegt das Problem darin, die mit dem neuen Medium möglich gewordenen massenhaften Datenerfassung (jeder Tastendruck wird auf der Zeitachse protokolliert) auf ein sinnvolles Maß zu reduzieren und die gewählten Indikatoren vernünftig zu interpretieren. Verglichen damit hatten es die Konstrukteure der

Papier-Bleistift-Versionen wesentlich einfacher. Der von den Testentwicklern vertretene Anspruch (S. 14), komplexere Verhaltensmuster wie Zielorientierung oder Planung auf Verhaltenzebene beobachtbar zu machen und zu operationalisieren, muß zumindest in bezug auf die Operationalisierung als noch nicht eingelöst bezeichnet werden. Die Forderung von Maukisch (1990, S. 43) nach „interaktiver Prozeßdiagnostik“, bei der enger umrissene Aufgabenkomplexe verwendet werden, „die die späteren Aufgaben und Trainingsziele in vereinfachter, miniaturisierter Form darstellen und bei denen interaktiv auf die Reaktionen des Kandidaten eingegangen wird“, ist hinsichtlich der Möglichkeiten interaktiver Reaktionen beim vorgestellten Verfahren erst ansatzweise realisiert.

Die Verwendung *computergestützter* Verfahren im Kontext von Personalauswahl und Personalentwicklung im Bereich leitender Führungskräfte hat gerade erst begonnen (vgl. Funke, 1993; Kluwe, 1990; Schuler & Moser, 1990). Von daher verwundert es kaum, daß erst wenig marktreife Instrumente entwickelt wurden und sich erst wenige wissenschaftliche Studien dieses Themas angenommen haben. Angesichts des hohen Bedarfs und der mit dem Verfahren wohl gegebenen Vorteile hoher Akzeptanz bei Anwendern wie bei Testpersonen sollte mehr Evaluationsforschung hierzu betrieben werden. MAILBOX '90 zeigt, daß lohnenswerte Themen in Kontext computergestützter Eignungsdiagnostik auf ihre Bearbeitung warten – sie sollten an die richtigen Adressaten delegiert werden.

Literatur

- Barell, G. (1992). *Bewährungskontrollen von Assessment Centern mittels testtheoretischer Validitätsmodelle. Unter besonderer Berücksichtigung der verhaltensorientierten „Within exercise-Variante“ zur Prognose allgemeiner Führungseignung in einer Schweizer Großbank.* Frankfurt: Lang.
- Fennekels, G. P. & Nathan, J. (1992). OFFICE. Ein Postkorb zur Diagnose von Führungsverhalten (unveröffentlichtes Manuskript). Bonn: Fennekels Personal- und Führungskräfteentwicklung.
- Funke, J. (in Druck). Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie.*
- Funke, J. & Rasche, B. (1988). *PC-POSTKORB. Manual zur PC-Version der Postkorb-Übung* (unveröffentlichtes Manuskript). Bonn: Psychologisches Institut der Universität Bonn.
- Hasselmann, D. (1992). *POSTMAPPE* (unveröffentlichtes Manuskript). Kiel: Hasselmann & Strauß GbR.
- Horn, R. (1991). MAILBOX: A computerized in-basket task for use in personnel selection. *Revue Européenne de Psychologie Appliquée*, 41, 325–327.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern. Assessment-Center-Verfahren.* München: Hanser.
- Jeserich, W. (im Druck). *Die Postkorb-Übung. Ein Instrument für Diagnose und Training.* Bergisch-Gladbach: Gesellschaft für Qualitative Personalarbeit.
- Klieme, E. & Stumpf, H. (1990). Computereinsatz in der pädagogisch-psychologischen Diagnostik. In K. Ingenkamp & R. S. Jäger (Eds.), *Tests und Trends 8. Jahrbuch der Pädagogischen Diagnostik* (pp. 13–63). Weinheim: Beltz.
- Kluwe, R. H. (1990). Computergestützte Systemsimulationen. In W. Sarges (Ed.), *Management-Diagnostik* (pp. 458–463). Göttingen: Hogrefe.
- Maukisch, H. (1990). Eignungsdiagnostik als prognostische Hilfe bei der Auswahl, Plazierung und Entwicklung von Führungskräften. In W. Sarges (Ed.), *Management-Diagnostik* (pp. 32–46). Göttingen: Hogrefe.

- Roest, F. & Horn, R. (1990). MAILBOX '90: Computerunterstützte Diagnostik im Assessment-Center. *Diagnostica*, 36, 213–219.
- Roest, F., Scherzer, A., Urban, E., Gangl, H. & Brandstätter, C. (1989). *MAILBOX '90. Ein computergestütztes Test- und Trainingsverfahren zur Personalentwicklung.* Wien und Weinheim: Scicon Datenverarbeitungsgesellschaft mbH und Beltz Test GmbH.
- Scharley, T. (1992). *PC-POSTKORB. Ein effizientes Beurteilungs- und Trainingsinstrument* (unveröffentlichtes Manuskript). Konstanz: Scharley & Partner Management Development Consultants.
- Schippmann, J. S., Prien, E. S. & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology*, 43, 837–859.
- Schmidt, F. L., Ones, D. S. & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology*, 43, 627–670.
- Schuler, H. (1990). Personenauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184–191.
- Schuler, H. & Moser, K. (1990). Geschichte der Managementdiagnostik. In W. Sarges (Ed.), *Management-Diagnostik* (pp. 18–28). Göttingen: Hogrefe.
- Testkuratorium der Föderation deutscher Psychologenverbände (1986). Beschreibung der einzelnen Kriterien für die Testbeurteilung. *Diagnostica*, 32, 358–360.

Anschrift des Verfassers

PD Dr. Joachim Funke
 Psychologisches Institut
 Universität Bonn
 Römerstraße 164
 W-5300 Bonn 1
 Email: Joachim.Funke@uni-bonn.de