

Joachim Funke

# Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale<sup>1</sup>

Dokumentation: Funke, J. (1993). Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37. Jg. (N. F. 11) 3, 119–129.

Schlagwörter: Arbeitsprobe, Postkorb, Computereinsatz, Personalauswahl

## Zusammenfassung

Arbeitsproben sind klassische Instrumente der Eignungsdiagnostik mit hoher Validität. In Zeiten zunehmend komplexer werdender Anforderungen ist jedoch die traditionelle Definition einer Arbeitsprobe als getreue Vorwegnahme einer späteren Anforderung zu erweitern im Sinne eines Bündels von abstrakten Anforderungen, die in der späteren Arbeitssituation gestellt werden. Zunächst werden daher Standard-Definitionen von Arbeitsproben sowie klassische Beispiele prozeß- wie ergebnisorientierter Probendiagnostik betrachtet. Aus deren Begrenzung kommt es zu einem Begriffsverständnis, das von einem höheren Abstraktionsgrad ausgeht und lediglich eine strukturelle bzw. funktionelle Übereinstimmung zwischen Probe und späterer Tätigkeit verlangt. Beispiele für computergestützte Arbeitsproben in Form der Produkte MAILBOX '90 und PC-POSTKORB, die auf der klassischen Postkorb-Übung aufbauen, werden detailliert und kritisch beschrieben. Zukünftige Entwicklungsmöglichkeiten theoriegeleiteter Diagnostik auf der Basis allgemein- und sozialpsychologischer Konzeptionen und unter Verwendung von Rechnern werden abschließend skizziert.

*Computer-based Work Samples: Concept, Examples, and Future Developments*

## Abstract

Work samples are classic instruments of personnel selection with high validity. With job requirements becoming more complex, the traditional definition of work samples as true replications of later requirements has to be extended towards an ensemble of more abstract requirements which are necessary and representative for the later job requirements. In a first step, traditional definitions and classical examples of process and result oriented work samples are considered. Starting from their limitations, a new conceptual understanding is developed which implies a higher level of abstraction and requires only a structural or functional equivalence between work sample and later job requirements. Examples of computer-based work samples derived from the classical in-basket task (MAILBOX '90 and PC-POSTKORB) are introduced and described in more detail. Potentials for future developments based on theoretical assumptions from a cognitive as well as from a social psychological point of view are outlined.

## 1 Einleitung

Arbeitsproben stellen ein klassisches Instrument der Eignungsdiagnostik dar. Bereits in den 20er Jahren dieses Jahrhunderts hatte sich die Arbeitsprobe als zweites Stand-

bein der damals „Psychognostik“ genannten Diagnostik neben den gerade entwickelten Intelligenzprüfungen etabliert. Aber auch in aktuellen Arbeiten aus den 70er und 80er Jahren werden nach wie vor positive Angaben über den Einsatz dieses Verfahrens gemacht.

Angesichts des wachsenden Einsatzes von Computern in der eignungsdiagnostischen Praxis (vgl. Jäger, 1990) stellt sich die Frage, wie es mit der Übertragung von Arbeitsproben auf rechnergestützte Umgebungen aussieht. Dazu wird es erforderlich sein, zunächst den Begriff der Arbeitsprobe zu beleuchten. Dabei ist insbesondere zu klären, inwiefern nicht ein zu enges Begriffsverständnis vorherrscht. Am Beispiel von Situationsübungen aus Assessment-Centern wie z. B. dem „Postkorb“ wird gezeigt werden, daß es sich hierbei ebenfalls um Arbeitsproben handelt. Da diese Übungen inzwischen in computergestützter Form vorliegen, werden zwei ausgewählte Verfahren dieses Typs kurz vorgestellt und hinsichtlich ihrer Vorzüge, aber auch ihrer Schwächen diskutiert.

Schließlich geht es um weitere Entwicklungsmöglichkeiten, die sich durch den Rechnereinsatz für Arbeitsproben eröffnen. Dabei wird zu zeigen sein, daß durch den Einbezug von Medien wie etwa interaktivem Video interessante diagnostische Verfahren entstehen könnten, die in standardisierter Form nicht nur die Erfassung des Planungs- und Entscheidungsverhaltens eines einzelnen Individuums ermöglichen, sondern darüber hinaus auch Merkmale des sozialen Interaktionsverhaltens am Rechner zugänglich machen und damit über die gegenwärtige Praxis von Assessment-Centern hinausführen.

## 2 Zum Begriff der Arbeitsprobe

Eine verbindliche Definition des Begriffs der Arbeitsprobe ist nicht einfach zu finden. Zunächst soll daher die historische Sicht dargelegt werden, ehe neuere Ansätze zum Zuge kommen.

### 2.1 Die historische Sicht

In einer viel (und häufig falsch) zitierten Arbeit von Giese (1924) über „Die Arbeitsprobe in der Psychognostik“ wird folgende Definition gegeben:

1 Dank an Lisa Irmen für ihre Hilfe bei der Literaturbeschaffung und bei der Suche nach geschlechtsneutralen Bezeichnungen. Wertvolle Kommentare und Anregungen verdanke ich Ernst Fay, Hermann-Josef Fisseni, Uwe Funke, Klaus-Martin Klein, Walter Neubauer, Heinz Schuler und Heinrich Wottawa. Luise Greuel hat auf ihre Art das Schreiben dieser Arbeit gefördert.

„Unter Arbeitsprobe verstehe ich den Versuch, in geregelter Form und an neutralem Stoff dem Prüfling Aufgaben zu stellen, die uns aufgrund seines dabei zu beobachtenden werktätigen Verhaltens einen Einblick in seine charakterologische Struktur ermöglichen“ (S. 163).

Dabei wird die so definierte Arbeitsprobe von einer Probearbeit dadurch unterschieden, daß letztere Kenntnisse und Fertigkeiten voraussetze (wie in einer schulischen Klassenarbeit). Bei der Arbeitsprobe soll der Stoff „neutral“ sein – Giese denkt an das Sortieren von Holzfiguren, das Nachrechnen von teilweise fehlerhaften Additionen einstelliger Zahlen, oder an das Stanzen von Löchern an vorgegebene Stellen in Papierstreifen. Wesentlich ist für ihn die „Arbeit“ des Probanden (Pbn), der nicht nur rezeptiv hinnehmen soll, was ihm vorgesetzt wird: „Wir fordern Handlung, Aktivität, Schaffen“ (Giese, 1924, S. 164). Hierbei soll nicht eine einzige Funktion in Anspruch genommen werden, sondern die gesamte Persönlichkeit soll beteiligt sein. Deswegen wird sie auch als „Komplexprüfung“ (S. 170) bezeichnet, eine „unter gleichbleibenden Bedingungen zustande kommende Prüfsituation“ (S. 170). Ausdrücklich fordert er, nicht nur eine, sondern mehrere Arbeitsproben zu einem Prüfverfahren zusammenzustellen, um durch die Vielfältigkeit der Prüfungen die Sicherheit des Verfahrens zu erhöhen (S. 180). Diese Vielfalt war z. B. gegeben bei der „Einheitsprüfung der Deutschen Reichsbahn“ (nachzulesen bei Moede, 1943, S. 187 f.), die aus einer reichhaltigen Sammlung von Arbeitsproben bestand. Ziel war dabei weniger die Auslese als vielmehr die Empfehlung, in welchem Bereich ein bestimmter Bewerber oder eine Bewerberin plazierte werden könnte.<sup>2</sup>

In seinem „Lehrbuch der Psychotechnik“ kommt Moede (1930) zu folgender Klassifikation von Arbeitsproben, wobei er fließende Übergänge zuläßt: (1) die Wirklichkeitsprobe „unter den Bedingungen der Wirklichkeit“, (2) Proben im Sinne von Schemata der Wirklichkeit, bei denen die wesentlichen Funktionen angesprochen werden, sowie (3) abstrakte Proben, die eine reine Funktionsprüfung darstellen. Stellt z. B. bei der Auswahl von Bergführern das Bergsteigen eine Wirklichkeitsprobe dar, so ist das Treppensteigen nur eine Schema-Probe und die Funktionsprüfung des Herzens wäre die abstrakte Probe. Da die Arbeitsprobe nicht die tatsächliche Tätigkeit, sondern eben nur ein Schema der Wirklichkeit darstelle, müsse man „mit aller Vorsicht“ auswerten (Moede, 1930, S. 39).

Ein klassischer Arbeitsversuch ist der Pauli-Test (Arnold, 1951; siehe auch Christiansen, 1983), bei dem anhand eines einstündigen fortlaufenden Addierens einstelliger Zahlen nicht nur Leistungsmenge und -güte, sondern auch Leistungsschwankungen erfaßt werden. Es handelt sich hierbei wohl um eine „abstrakte Probe“ sensu Moede, wenngleich der Untersuchungszweck diese Bewertung ändern mag: Während der Pauli-Test eine abstrakte Probe darstellt, wenn man die Vigilanz eines Piloten erfassen möchte, rückt er in die Nähe einer Wirklichkeitsprobe,

wenn es um die Leistungsbeurteilung eines Revisions-Sachbearbeiters geht. Obzwar eine Zeitlang diese Art der Arbeitsprobe als „Universaltest“ eingesetzt wurde („ein sicheres Diagnostikum zur Ermittlung konstitutiver Unterschiede, z. B. des Unterschieds der Geschlechter“ – so Arnold, 1951, S. 7), deckt sie nicht das klassische Verständnis des Begriffs ab. Hierzu zählt z. B. eher eine Arbeit von Schulte (1921), der die „Berufseignung des Damenfriseurs“ mittels speziell entwickelter Arbeitsproben (Sortieren von Haarproben, Messung der Zugstärken beim Kämmen von Perücken usw.) bestimmt hat. Diese eher amüsanten Proben sind inzwischen durch situativ orientierte Methoden abgelöst worden (vgl. Schuler & Moser, 1990), zu denen insbesondere die Assessment-Center-Methode zählt.

## 2.2 Moderne Sichtweise

Im Rahmen der heute üblichen Assessment-Center-Methode zur Eignungs- bzw. Potentialdiagnose geht es darum, eine bestimmte Menge von als relevant erachteten Dimensionen möglichst gut auf eine Reihe von Übungen abzubilden, um dann die individuellen Ausprägungen auf diesen Dimensionen in Hinblick auf ein definiertes Anforderungsprofil zu bewerten.

Während Leistungstests (zur Erfassung von Intelligenz, Konzentration usw.) unspezifische Verfahren darstellen, wird die Arbeitsprobe „ganz gezielt zur Prognose der Leistung an einem konkreten Arbeitsplatz entwickelt“ (Rösler, 1988, S. 69). Dieses Vorgehen besitzt – so Rösler – gegenüber den unspezifischen Tests den Vorteil höherer „Augenscheinvalidität“ für Bewerber und Arbeitgeber und somit eine höhere Akzeptanz. Hinsichtlich des Kriteriums der „sozialen Validität“ sensu Schuler (1990), das aus den vier Komponenten „Information“, „Partizipation“, „Transparenz“ sowie „Ergebniskommunikation“ besteht, schneiden Arbeitsproben und Simulationen gut ab: Sie werden „als transparent und belastend erlebt und positiv bewertet“ (Schuler, 1990, S. 190). Schuler und Funke (1993) sehen in Arbeitsproben standardisierte Aufgaben, die inhaltlich valide und erkennbar äquivalente Stichproben des erfolgsrelevanten beruflichen Verhaltens darstellen. In recht großzügiger Auslegung des Begriffs werden von den genannten Autoren etwa auch biografische Fragebögen als „retrospektive Arbeitsproben“ bezeichnet; eine Probezeit gilt als zeitlich ausgedehnte Arbeitsprobe.

Nach einer von Jeserich (1982) beschriebenen Klassifikation von Übungen gehören Szenarien, Postkörbe und Fallstudien zu den „Situationsübungen“, mit denen sowohl analytische Fähigkeiten als auch soziale Kompetenzen diagnostiziert werden können. Situationsübungen und Arbeitsproben sind einander jedoch sehr ähnlich:

„Situational exercises are roughly the white-collar counterpart of work samples; that is, they are used mainly to select people for managerial and professional jobs. Unlike work samples, which are

2 Moede (1943, S. 182) spricht stolz von einer „Totalbegutachtung“ der Bewerber aus der Rotte zum Zwecke der Besteinweisung in die einzelnen Reichsbahndienstsparten nach dem Einheitsverfahren“.

designed to be replicas of the job, situational exercises only mirror part of the job“ (Muchinsky, 1990, S. 124).

Zu den Situationsübungen zählt Muchinsky das Postkorb-Verfahren und die führerlose Gruppendiskussion. Deren insgesamt geringere Validität (zwischen 0.15 und 0.35 im Vergleich zu 0.40 bis 0.60 bei Arbeitsproben) führt er auf die größeren Schwierigkeiten bei der Festlegung von Kriterien für Management-Erfolg zurück, verglichen etwa mit den Kriterien für Mechaniker-Erfolg.

Zusammenfassend läßt sich eine Einordnung der Instrumente „Arbeitsprobe“ und „Situationsübung“ im Vergleich zu anderen Auswahlverfahren gemäß den von Muchinsky (1990) angegebenen evaluativen Standards vornehmen wie in Tabelle 1 gezeigt. Während die Spalte „Validität“ in Tabelle 1 das wohl wichtigste Kriterium darstellt, das die Güte der Vorhersagbarkeit der Kriteriumsleistung betrifft, bezieht sich die Spalte „Fairneß“ auf die Vorhersagbarkeit von Berufserfolg unbeeinflusst von der Subgruppen-Zugehörigkeit von Bewerbern und Bewerberinnen; die Spalte „Anwendbarkeit“ fragt nach möglichen Einschränkungen in Hinblick auf Stellen- bzw. Bewerbungstypen, die Spalte „Kosten“ bewertet den finanziellen Aufwand der Verfahrensimpementation. Vor allem in Hinblick auf den zusätzlich zu Muchinsky eingeführten Evaluationsmaßstab „Akzeptanz“ sind hohe Ausprägungen für die beiden genannten Verfahren zu verzeichnen; das Kostenargument fällt für Arbeitsproben stärker ins Gewicht als für Situationsübungen, deren Fairneß allerdings unklar ist.

In der mehrere tausend Untersuchungen zusammenfassenden Metaanalyse von Hunter und Hunter (1984; vgl. auch Asher & Sciarrino, 1974) hinsichtlich der Validität verschiedener Prädiktoren für die spätere Berufsleistung (gemessen über Vorgesetztenurteil, Aufstieg, Trainingserfolg oder Beschäftigungsdauer) wird unterschieden zwischen Berufsanfängern und Berufserfahrenen: Während die Berufsleistung von Berufsanfängern mit Abstand am besten durch Tests zu kognitiven Fähigkeiten vorhergesagt werden kann (mittlere korrigierte Validität 0.53), ist bei Berufserfahrenen die Arbeitsprobe mit einer mittleren korrigierten Validität von 0.54 sogar noch um einen Prozentpunkt besser als die auch hier gut geeigneten Fähigkeitstests.

Tabelle 1

Vergleich von zehn Personalauswahlmethoden hinsichtlich wichtiger evaluativer Standards (nach Muchinsky, 1990, S. 134; erweitert um den Akzeptanz-Standard)

Auswahlverfahren	Validität	Fairneß	evaluativer Standard		
			Anwendbarkeit	Kosten	Akzeptanz
Arbeitsproben	hoch	hoch	niedrig	hoch	hoch
Situationsübungen	mittel	(unbekannt)	niedrig	mittel	hoch
Simulationen	??	(unbekannt)	??	??	hoch
Intelligenztest	mittel	mittel	hoch	niedrig	niedrig
Fähigkeits- und Leistungstests	mittel	hoch	mittel	niedrig	mittel
Persönlichkeits-/Interessentests	mittel	hoch	niedrig	mittel	niedrig
Interviews	niedrig	mittel	hoch	mittel	hoch
Biografische Informationen	hoch	mittel	hoch	niedrig	mittel
Kolleg(inn)enurteile	hoch	mittel	niedrig	niedrig	mittel
Empfehlungsschreiben	niedrig	(unbekannt)	hoch	niedrig	niedrig

Anmerkung: Detailinformationen über die Art der genannten Auswahlverfahren wie auch über die Definition der Bewertungskategorien findet man bei Muchinsky (1990).

### 2.3 Probleme von Arbeitsproben

Arbeitsproben im strengen Sinn scheinen zunächst auf „blue-collar“-Arbeiter beschränkt und erfassen kein Potential. Dieses Defizit gleichen Arbeitsproben mit Lernfähigkeitstests aus, wie sie etwa in der Meta-Analyse von Robertson und Downs (1989) beschrieben werden. Die sogenannten „trainability tests“ stellen dabei Arbeitsproben für untrainierte und unerfahrene Personen dar (mit einem „miniature job training“), wohingegen die gängigen Arbeitsproben für erfahrene und kompetente Personen konzipiert sind. Die Beschränkung von Arbeitsproben auf handwerkliche Fähigkeiten wird durch das oben erwähnte Begriffsverständnis von Schuler und Funke (1993) allerdings aufgehoben. In diesem Verständnis rücken Arbeitsproben und Situationsübungen enger zueinander als in der traditionellen Sicht – eine erweiterte Auffassung also, der wir uns hier anschließen wollen.

Eine andere Kritik an Arbeitsproben stammt von Wernimont und Campbell (1968). Danach sei die häufig beobachtbare hohe Korrelation zwischen Arbeitsprobe und Kriterium darauf zurückzuführen, daß es sich bei der Arbeitsprobe um eine Art Paralleltest handle. Diese Kritik scheint aber leicht zurückweisbar zu sein: Wenn die Arbeitsprobe tatsächlich eine derartige Paralleltest-Eigenschaft zur Arbeitswirklichkeit aufweist, rechtfertigt dies m.E. umso mehr ihren Einsatz als Diagnostikum.

Es zeigt sich somit, daß die vermuteten Probleme von Arbeitsproben sich im wesentlichen als „Scheinprobleme“ darstellen und somit nicht als Argument gegen ihren eignungsdiagnostischen Einsatz dienen können.

## 3 Beispiele für computergestützte Arbeitsproben

In diesem Kapitel wird eine bestimmte Arbeitsprobe im Sinne einer inhaltlich validen und erkennbar äquivalenten Stichprobe des erfolgsrelevanten beruflichen Verhaltens (vgl. Schuler & Funke, 1993) genauer dargestellt: die sogenannte „Postkorb“-Übung. Der Postkorb klassischer

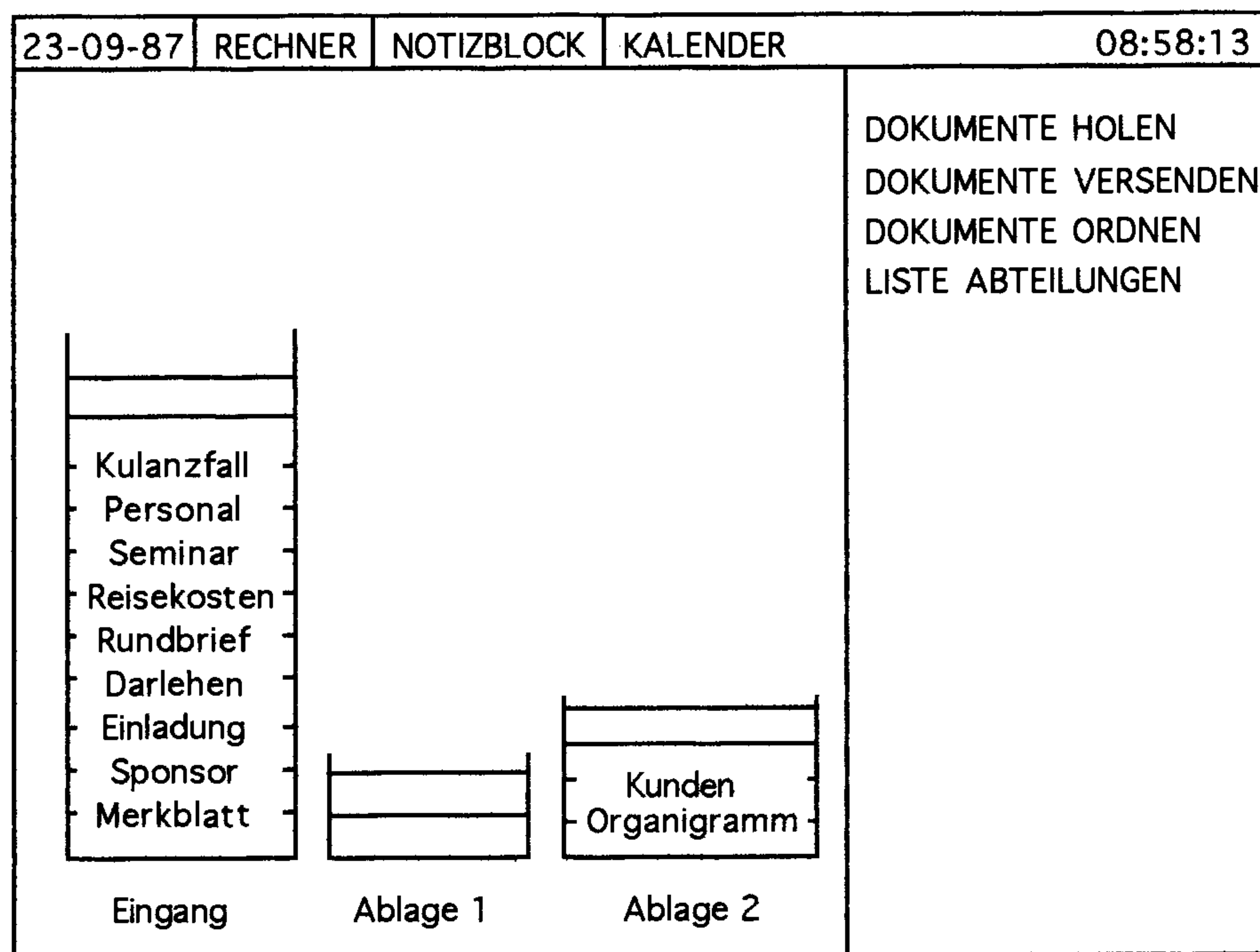


Abbildung 1

*MAILBOX '90: Display des Schreibtisches mit den verschiedenen Körben und Hilfsmitteln; in der Kopfzeile: Hilfsmittel; rechts: Eingriffsmöglichkeiten*

Form (vgl. z. B. das bei Jeserich, 1981, S. 172–193, veröffentlichte Beispiel) verlangt als zeitbegrenzte schriftliche Einzelübung die Bearbeitung einer größeren Menge an Vorgängen unterschiedlicher Herkunft. Die bearbeitende Person selbst wird in die Rolle eines Managers und Familienoberhaupts versetzt, der/das gerade von einer längeren Dienstreise zurückkehrt und gleich zur nächsten aufbrechen soll. Außer Papier und Bleistift stehen keine anderen Hilfsmittel zur Verfügung, um mit den verschiedenen schriftlich dokumentierten Vorgängen aus dem beruflichen wie privaten Bereich umzugehen. Aufgabe der bearbeitenden Person ist das rasche Treffen von Entscheidungen, die etwa darin bestehen können, schriftlich Anweisungen zu erteilen, die Bearbeitung eines Vorgangs zu delegieren oder die Entscheidung zurückzustellen. Häufig werden der bearbeitenden Person Informationen über die interne Struktur der Firma in Form eines Organigramms zur Verfügung gestellt. In den meisten Fällen wird übrigens auch von Bearbeiterinnen verlangt, sich in die männliche Rolle hineinzusetzen.

Die inzwischen zahlreiche Literatur zum „Postkorb“ hat sich mit vielfältigen Aspekten dieses Instruments auseinandergesetzt, das nach Ansicht vieler Autoren zur Standardübung eines „Assessment Centers“ gehört (zum Überblick siehe Schippmann, Prien & Katz, 1990). Im deutschsprachigen Raum liegen jedoch nur wenige Untersuchungen zum Postkorb-Verfahren vor (z. B. Didi, 1992; Hasselmann, 1993).

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. Allerdings sollten bei einer Umsetzung des Verfahrens auf den Rechner auch noch andere Vorteile genutzt werden als die bloße Auswertungsersparnis. Nachfolgend werden zwei Computerversionen von Postkörben, MAILBOX '90 und PC-POSTKORB, genauer vorgestellt und kritisch analysiert.

### 3.1 MAILBOX '90

Ziel dieser 1986 begonnenen, seit 1988 als Prototyp und seit 1990 als fertiges Produkt vorliegenden Implementation eines Postkorbs ist eine Objektivierung der Instruktion, des Verfahrensablaufs und der Auswertung durch folgende Aspekte (vgl. 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 (z. B. simulierter Anruf); (4) verzögertes Feedback bei falscher Delegation; (5) Protokollierung des exakten Testablaufes sowie automatische Vergabe von Positiv- und Negativpunkten pro bearbeiteter Aufgabe.

*Beschreibung der Aufgabenstellung und der Benutzeroberfläche.* Die Aufgabe besteht aus neun Vorgängen (Dokumenten), die innerhalb von 32 Minuten bearbeitet werden müssen (vgl. Roest, Scherzer, Urban, Gaugl & Brandstät-

Tabelle 2

Die sieben aufgrund inhaltlicher Überlegungen abgeleiteten Skalen von MAILBOX '90 (linke Spalte: Dimension; rechte Spalte: Verhaltensindikator)

(1) Arbeitszeit	Bearbeitungszeit pro Dokument
(2) Arbeitseinteilung	Auszählung der Häufigkeit, mit der jedes Dokument bearbeitet wurde (= eventuelles Wiederaufgreifen schon gelesener Vorgänge)
(3) Zielorientierung	richtig/falsch-Klassifizierung jedes Dokuments hinsichtlich einer vordefinierten Endposition, an der es sich am Testende befinden sollte
(4) Arbeitsgüte	pro Dokument 1 bis 4 Punkte für bestimmte Tätigkeiten (z. B. für Einsichtnahme in das Dokument, Weiterleitung an eine andere Person, Überprüfung eines Termins im Kalender)
(5) Aktivität	Häufigkeit der Verwendung aller für die Dokumentenbearbeitung möglichen Funktionen (inkl. 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

Anmerkung: Die aufgabenorientierten Skalen sind (1) bis (4). Die verlaufsorientierten Skalen (5) bis (7) unterteilen die Gesamtbearbeitungszeit in sechs 5-Minuten-Intervalle, in denen jeweils die Frequenz bestimmter Ereignisse ausgezählt wird.

ter, 1989). Dem Pbn stehen dabei auf dem Bildschirm ein Rechner, ein Notizblock und ein Kalender zur Verfügung. Zusätzliche Informationen kann er aus einer Kundenliste sowie einem Organigramm des Unternehmens entnehmen. Die Bedienung erfolgt mausgesteuert. Neben dem Eingangskorb stehen auf dem Bildschirm zwei weitere Ablagekörbe zur Verfügung (vgl. Abbildung 1).

*Auswertung der Testleistung und Beschreibung der zugrundeliegenden Skalen.* Insgesamt sieben Skalen werden aufgrund inhaltlicher Überlegungen abgeleitet, vier davon aufgabenorientiert und drei verlaufsorientiert (siehe Tabelle 2).

*Untersuchungen zu Dimensionalität, Reliabilität und Validität des Verfahrens.* Empirische Untersuchungen zur Dimensionalität des Verfahrens liegen nicht vor. 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 vorgenommen. Von 159 der 193 Teilnehmer 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 nach Ward über die Gesamtstichprobe durchgeführt wurden; zusätzlich wurden Mittelwertvergleiche zwischen den Ergebnissen der Studenten und der Bewerber vorgenommen. Hinsichtlich Arbeitszeit ist z. B. die Bewerberstichprobe schneller als die der Studenten, angenommene Bewerber sind schneller als abgelehnte. Für die Skala Arbeitseinteilung wird eine split-half-Reliabilität von .75 angegeben, für Arbeitsgüte von .79. Sonstige Reliabilitätsangaben werden nicht gemacht. Während sich bei den aufgabenorientierten Skalen die Analysen jeweils auf die erreichte Punktzahl (bzw. die verbrauchte Zeit) bezogen, beziehen sich die Clusteranalysen bei den prozeßorientierten Skalen auf die Auftrittshäufigkeit der definierten Verhaltensweisen in sechs Fünf-Minuten-Abschnitten. Die hier vorgelegten Drei-Cluster-Lösungen sind allerdings inhaltlich schwerer zu interpretieren: was es etwa bedeutet, wenn eine Personengruppe im dritten Zeitabschnitt im Mittel fünfmal

ordnet, während die anderen Cluster dort im Mittel unter 0.5 liegen, bleibt unklar.

Validierungsinformationen stammen aus Interkorrelationen zu anderen Verfahren. Von insgesamt 392 mitgeteilten Korrelationen erweisen sich 41 als statistisch bedeutsam (beim zugrundegelegten Irrtumsniveau von  $\alpha = .05$  sollten allein 20 der 392 *per Zufall* signifikant sein). In der Studentenstichprobe zeigt sich so z. B. ein bedeutsamer Zusammenhang ( $r = .20$ ) zwischen der Skala „Arbeitszeit“ und dem Score der „Advanced Progressive Matrices“, die zahlenmäßig höchste Korrelation besteht zwischen „Zielorientierung“ und der Skala „Selbstbeschuldigung“ aus einem Streßverarbeitungsfragebogen ( $r = .41$ , d. h. 16% gemeinsame Varianz). So folgern die Autoren nicht ganz zu Unrecht, daß „die MAILBOX '90 ein von anderen Verfahren nicht erfaßtes, unabhängiges Verhalten mißt“ (Roest et al., 1989, S. 40). Die daraus im nächsten Satz abgeleitete Schlußfolgerung teile ich allerdings nicht: „Die Ähnlichkeit der Ergebnisse mit Validierungsstudien anderer Pos[t]korbverfahren deutet darauf hin, daß es gelungen ist, das Postkorbverfahren am Computer so zu operationalisieren, daß ähnliche Verhaltensmerkmale im Test zum Ausdruck kommen wie in der Papier-Bleistift-Version.“

*Bewertung.* Der letzte Abschnitt machte bereits deutlich, daß hinsichtlich psychometrischer Ansprüche Forscher wie Anwender nicht zufriedengestellt werden. Weder über Dimensionalität noch Reliabilität noch Validität des Verfahrens ist genaues bekannt. Die vorgelegten Daten sind als unzureichend zu bewerten. 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 allzusehr 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, aber dies sind eher oberflächliche Änderungen. Die Ableitung von Parametern ist noch nicht zufriedenstellend

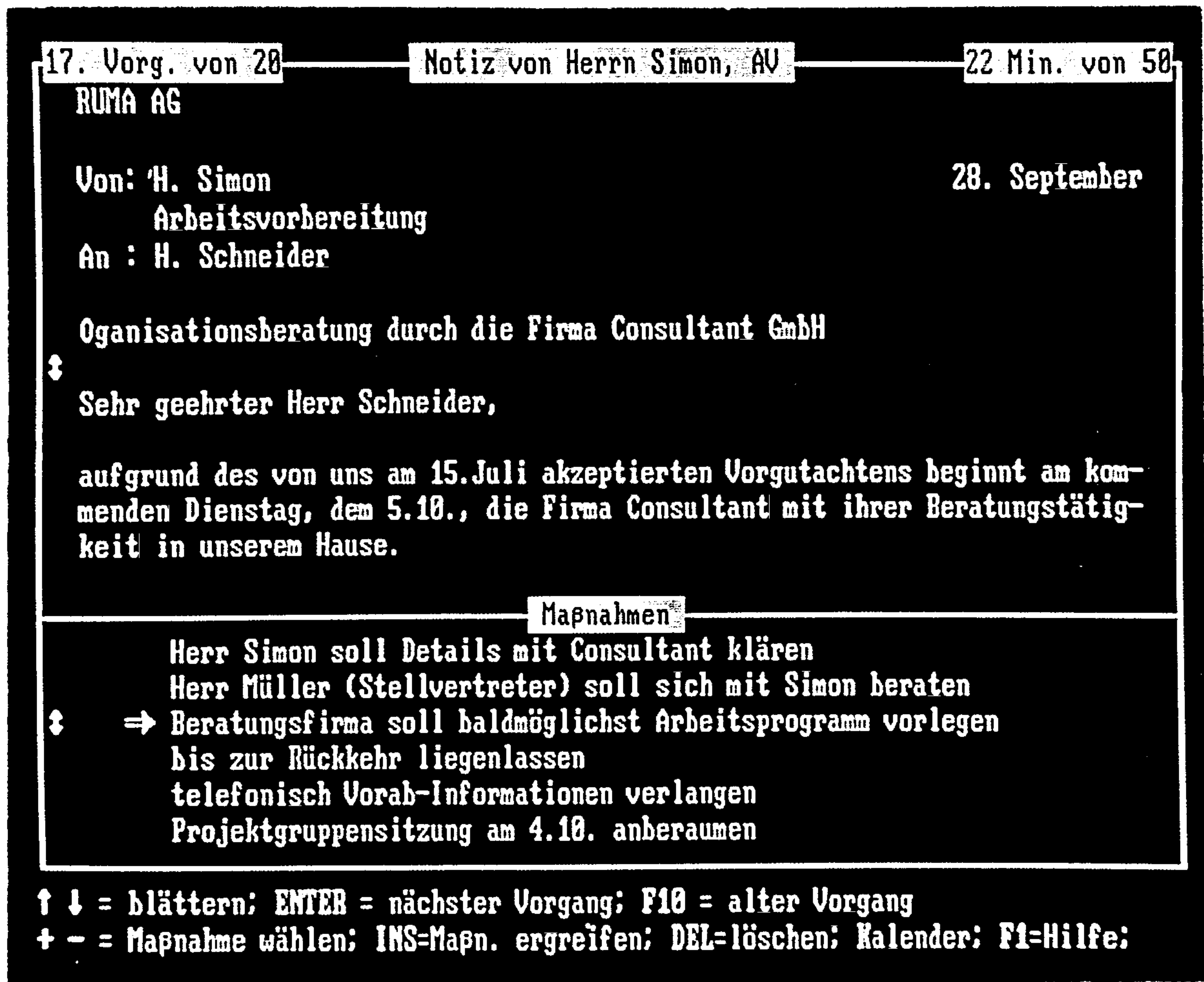


Abbildung 2

PC-POSTKORB: Display eines Teils eines Vorgangstextes im oberen Teil und einiger Entscheidungsmöglichkeiten dazu im unteren Teil des Bildschirms

begründet. Warum werden z. B. bei den verlaufsorientierten Skalen sechs Zeitabschnitte und nicht drei, fünf oder zehn unterschieden? Warum sollte die Häufigkeit, mit der eine Umblättern-Funktion genutzt wird, etwas mit Aktivität im Sinne einer Führungseigenschaft zu tun haben? Das Problem liegt darin, die mit dem neuen Medium möglich gewordenen massenhaften Datenspeicherungen (jeder Tastendruck wird auf der Zeitachse protokolliert) auf ein *sinnvolles* Maß zu reduzieren und die gewählten Indikatoren sinnvoll zu interpretieren.

### 3.2 PC-POSTKORB

PC-POSTKORB ist eine von Funke und Rasche (1988) entworfene PC-gestützte Variante eines Postkorbs, der zuvor in einer Papier-und-Bleistift-Form bei ca. 120 Führungskräften aus verschiedenen Firmen eingesetzt wurde.<sup>3</sup> Diesem Postkorb zugrunde liegen 20 Vorgänge, unter anderem aus der bei Jeserich (1981) veröffentlichten Fassung,

die sowohl firmeninterne als auch private Vorgänge betreffen.

Entstanden ist dieser Postkorb aus einer kritischen Bewertung der Objektivität vorliegender Papier-Bleistift-Versionen, die sich im wesentlichen auf vier Punkte stützt: (1) Die fragliche Qualität der „Musterlösung“, die als Bewertungsmaßstab herangezogen wird: Die damit vorgegebenen Richtiglösungen erscheinen in hohem Maße willkürlich und beliebig. (2) Die mangelnde Auswertungsobjektivität: vergebene Punktzahlen sind interpretationsabhängig – folglich kommen verschiedene Auswerter bei gleicher Antwort der Testperson zu unterschiedlichen Punktwerten. (3) Die Bewertung von kombinierten Maßnahmen: Die Auswertung muß nicht nur einzelne Lösungsvorschläge, sondern häufig das Ensemble einer Lösung beachten – auch dies mindert die Objektivität des Verfahrens. (4) Die vom Konzept der Musterlösung her suggerierte Eindimensionalität vieler Postkörbe – *ein* pauschaler „Postkorb“-Wert bildet die Leistung der Testperson ab – bleibt unge-

3 Die Papierversion stammt von Bernd Ahlbrecht, der zusammen mit Elfriede Hillejan die Anregung zur Erstellung einer PC-Version gegeben hat.

prüft und auch angesichts der verschiedenen Anforderungen zweifelhaft. Aus diesen vier Kritikpunkten heraus wurde die Konstruktion des neuen Verfahrens gespeist.

*Beschreibung der Aufgabenstellung und der Benutzeroberfläche.* In seinem Kern ist PC-POSTKORB als „multiple choice“-Verfahren konzipiert: Zu jedem der insgesamt 20 Vorgänge wird eine Liste von Antwortmöglichkeiten vorgegeben, aus denen eine oder mehrere Lösungen auszuwählen sind. Die Distraktoren stammen aus realen Antworten zu den in paper-and-pencil-Form präsentierten Items aus der Vorläuferversion. Die Bearbeitung erfolgt über Funktionstasten. Insgesamt stehen einem Pbn für die reine Vorgangsbearbeitung 50 Minuten zur Verfügung; das Lesen der am PC dargebotenen Instruktion sowie die Bearbeitung eines Übungsteils zur Handhabung der Benutzeroberfläche (Vorgang lesen, Handlungsalternativen auswählen, Kalenderführung) erfolgt ohne Zeitbegrenzung. Abbildung 2 zeigt einen typischen Bildschirmaufbau während einer Vorgangsbearbeitung mit einigen der Wahlmöglichkeiten im unteren Bildschirm Drittel.

*Auswertung der Testleistung und Beschreibung der zugrundeliegenden Skalen.* Bewertet wird anhand einer durch Fachleute festgelegten Musterlösung, die insgesamt vier a priori aufgrund einer inhaltlichen Aufgabenanalyse festgelegte Dimensionen heranzieht: (1) Analysefähigkeit, (2) Delegationsvermögen, (3) Soziales Geschick sowie (4) Organisationsfähigkeit. Eine Skala (5) Entscheidungsqualität stellt den Summenwert der vier genannten Teilleistungen dar, die nachfolgend etwas genauer charakterisiert werden. Jede Skala wird durch richtige (Plus-Items) und falsche (Minus-Items) Antworten beschrieben, die sich über die verschiedenen Vorgänge verteilen.

*Analysefähigkeit (A)* bezieht sich auf das korrekte Einschätzen und Beurteilen der durch einen Vorgang präsentierten Sachlage (Itempool der Plus- und Minus-Items: 21 A+, 84 A-). Korrekt analysiert wird etwa, wenn man sich bei einem Eurocard-Angebot für die Ablage im Papierkorb entscheidet. *Delegationsvermögen (D)* betrifft die Fähigkeit zu erkennen, welche Aufgaben nicht selbst erledigt werden müssen, sondern an andere dafür qualifizierte Personen weiterzureichen sind (Itempool: 9 D+, 21 D-). Eine falsche Delegation ist es z. B., die Schwiegermutter mit der Betreuung der Kinder zu beauftragen, da ein eigenes Kindermädchen zur Verfügung steht. *Soziales Geschick (S)* bezieht sich auf die Kompetenz, die durch einen Vorgang geschaffenen Verhältnisse in „menschlichem“ Umgangston zu klären. Mit einer fristlos gekündigten Person ein klärendes Gespräch zu führen wird z. B. positiv bewertet (Itempool: 7 S+, 7 S-). *Organisationsfähigkeit (O)* liegt vor, wenn verschiedene zeitliche Anforderungen in einem Plan konfliktfrei untergebracht werden können. Dies manifestiert sich überwiegend an der Führung des bildschirmgestützten Kalenders, wo wichtige Eintragungen vorgenommen und bereits vorgegebene unwichtige Termine gestrichen werden sollen (Itempool: 11 O+, 11 O-).

*Untersuchungen zu Dimensionalität, Reliabilität und Validität des Verfahrens.* Publierte empirische Untersuchungen

zu Dimensionalität, Reliabilität und Validität dieses Postkorbs liegen nicht vor. Funke und Rasche (1988) verweisen auf die Notwendigkeit derartiger Untersuchungen und berichten nur teststatistische Kennwerte der alten Version (61 Führungskräfte im Vergleich zu 18 Studenten; für 19 Items in beiden Stichproben bei deutlichem Leistungsunterschied zugunsten der Führungskräfte split-half-Reliabilität von .40, Spearman-Brown-korrigiert .56). Im neuen Manual hat sich daran nichts geändert. Allerdings werden dort Normwerte abgedruckt, die sich auf insgesamt 115 Personen beziehen und eine Transformation der jeweiligen Rohpunktwerte in Prozentränge ermöglichen.

*Bewertung.* Mit dieser Version ist eine grundsätzliche Änderung der Anforderung verbunden: von der kreativen Suche nach einer Lösung hat sich die Anforderung in Richtung auf die Auswahl der vermutlich korrekten Antworten hin verschoben. Der Grund für diese Änderung ist im radikalen Streben nach Objektivität der Auswertung zu sehen: nicht nur die Vorgangsbearbeitung, sondern auch die Kalenderführung sind – immer in Hinblick auf eine durchaus kritisierbare Musterlösung – problemlos in Leistungswerte zu überführen. Durch die realistische Gestaltung der Distraktoren sowie durch die Tatsache, daß mehrere Antworten richtig sein können, ist die multiple-choice-Situation damit auf ein erhebliches Schwierigkeitsniveau zu heben, die den Verdacht einer zu einfachen Aufgabe, bei der man ja „nur noch“ die richtige Antwort herausfinden müsse, entkräftet. Allerdings sind sowohl die fehlenden psychometrischen Informationen als auch die ausschließlich auf Objektivierung abgestellte Nutzung des Mediums „Computer“ zu kritisieren.

### 3.3 Zusammenfassende Bewertung

Die Darstellung der beiden ausgewählten computergestützten Postkörbe zeigt in beiden Fällen ein erhebliches Defizit im Hinblick auf psychometrische Informationen. Dies gilt sowohl für die Konstruktion der Verfahren als auch für die Ableitung von Leistungsparametern, deren Reliabilität und Validität weitgehend ungeprüft ist. Schippmann, Prien und Katz (1990) sehen einen Grund für diesen unbefriedigenden Zustand darin, daß die hohe Augenschein-Validität jeder Art von Arbeitsprobe psychometrische Analysen fast überflüssig erscheinen lassen. Die von den genannten 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-und-Bleistift-Versionen, aber die oben gegebene Darstellung rechtfertigt wohl die Ausdehnung dieser Argumentation auf Computerversionen. Allerdings ist zu beachten, daß die Kritik an den computergestützten Varianten hinsichtlich fehlender teststatistischer Informationen auf die meisten Papier-und-Bleistift-Varianten gleichfalls Anwendung findet und daher kein Spezifikum der Computervarianten darstellt.

Dabei ist zu bedenken, daß bestimmte Besonderheiten der Postkorb-Übung eine einfache Übertragung klassischer testtheoretischer Konzepte verhindern. So ist z. B. das Prinzip der Bestimmung von Retest-Reliabilitäten nicht ohne weiteres zu übernehmen, da eine simple Testwiederholung wegen dabei auftretender Lerneffekte kaum in Frage kommen dürfte. Aber auch andere Arten der Reliabilitätsbestimmung (z. B. split-half) kommen wegen der gewollten partiellen Abhängigkeit verschiedener Vorgänge (= Items) nicht ohne weiteres in Frage. Dies rechtfertigt keineswegs das Fehlen von Angaben, sondern verlangt im Gegenteil nach einfallsreichen Vorgehensweisen zur Behebung dieser Probleme.

Natürlich ist bei einer Bewertung auch zu fragen, inwieweit bestimmte Aspekte des Mediums „Computer“ innovativ genutzt werden. Hierzu zählen zum einen die Möglichkeiten der Dynamisierung des Testens wie auch die des raschen Ergebnis-Feedbacks. Von MAILBOX '90 wird z. B. die Möglichkeit genutzt, die aktuelle Bearbeitung eines Vorgangs durch einen simulierten Anruf (= Telefonläuten sowie anschließende Einblendung von Text auf dem Schirm) zu stören. Während dieses Detail eher einen oberflächlichen Aspekt betrifft, ist die Nutzung der Rückgabe falsch bearbeiteter Dokumente sicher sinnvoll und kann als kleiner Schritt in Richtung Lerntest (siehe Guthke, 1982) gesehen werden. Bei PC-POSTKORB ist eine „biedere“ Umsetzung der Papier-und-Bleistift-Variante realisiert, die von den neuen Möglichkeiten des Mediums noch keinen Gebrauch macht, von der raschen Ergebnismeldung einmal abgesehen, die beide Verfahren anbieten. Hierbei hat PC-POSTKORB allerdings den Vorzug der automatischen Kalender-Auswertung: während die von der bearbeitenden Person vorgenommenen Kalendereinträge bei MAILBOX '90 als beliebige Klartext-Eingabe durch den Pbn erfolgen, die nicht ausgewertet werden können, sondern dem Testleiter bzw. der Testleiterin auf dem Ergebnisausdruck als eben diese Klartext-Eingabe zur manuellen Bewertung vorgelegt werden, erlaubt PC-POSTKORB durch die multiple-choice-Kalenderführung eine automatisierte Bewertung der Kalendereinträge, die als separater Leistungsindex („Organisationsfähigkeit“) ausgegeben wird.

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 jedenfalls hinsichtlich der Möglichkeiten interaktiver Reaktionen bei den vorgestellten Verfahren nur ansatzweise realisiert.

Die Verwendung *computergestützter* Verfahren im Kontext von Personalauswahl und Personalentwicklung im Bereich leitender Führungskräfte hat gerade erst begonnen. Von daher verwundert es kaum, daß erst wenige Instrumente im Sinne akzeptabler Testverfahren entwickelt wurden und sich erst wenige wissenschaftliche Studien dieses Themas angenommen haben. Um weitere Entwicklungsarbeiten zu

ermuntern, sollen nachfolgend einige noch nicht ausgeschöpfte Möglichkeiten skizziert werden.

## 4 Entwicklungsmöglichkeiten

Unter der Prämisse, daß computergestützte Arbeitsproben einen guten Zugang zur Potentialbeurteilung von Arbeitnehmern und Arbeitnehmerinnen darstellen, sollen nachfolgend Perspektiven zukünftiger Forschung sowohl aus allgemein- wie aus sozialpsychologischer Sicht dargestellt werden. In beiden Fällen erlaubt der Einsatz von Computern neue Wege der Eignungsdiagnostik.

### 4.1 *Allgemeinpsychologische Perspektive*

Aus allgemeinpsychologischer Perspektive möchte ich zwei Entwicklungslinien hervorheben, die zukünftig mehr Bedeutung erlangen werden. Zum einen handelt es sich dabei um die computergestützte Diagnostik von Planungsfähigkeiten, zum anderen um das computergestützte Training zum besseren Umgang mit komplexen Situationen, das auf dem Einsatz von Modellbildungssystemen beruht.

#### 4.1.1 Planungsdiagnostik als Schemadiagnostik

Ganz allgemein kann in Übereinstimmung mit Schuler und Moser (1990, S. 26) festgehalten werden, daß kognitive Leistungstests noch nicht den festen Platz in der Managementdiagnostik erreicht haben, den sie eigentlich aufgrund ihrer prädiktiven Potenz einnehmen sollten. Ein nach wie vor bestehendes diagnostisches Desiderat betrifft dabei die Planungskompetenz. Hierfür liegen bis heute keine brauchbaren Indikatoren vor. Hauptproblem für die unbefriedigende Lage nicht nur im eignungsdiagnostischen Kontext stellt das Fehlen von brauchbaren Planungstheorien dar. Da jedoch bereits seit den frühen Arbeiten von Miller, Galanter und Pribram (1960) klar war, daß Planen auf dem Einbezug bereits vorliegender Erfahrungen beruht, wird die Rolle von im Gedächtnis gespeicherten Schemata und Skripten deutlich.

Für das Thema „Arbeitsproben“ bedeutet dies: in einer typischen Arbeitsprobe hängen die Entscheidungen der handelnden Person von ihren verfügbaren Schemata ab. So mag es etwa ein „Post-Bearbeitungsskript“ geben, wonach zunächst einmal alle eingegangene Post grob gescannt und dabei nach den zwei Dimensionen „Wichtigkeit“ und „Dringlichkeit“ sortiert wird. Die Skriptverfügbarkeit könnte dabei durch klassische gedächtnispsychologische Meßverfahren (z. B. Reaktionszeitmessung) rechnergestützt abgefragt werden.

#### 4.1.2 Modellbildung als Lernerfahrung für komplexe Zusammenhänge

Daß gerade Führungskräfte in der Lage sein sollten, komplizierte Zusammenhänge rasch zu erkennen und darin zielgerichtet und erfolgreich einzugreifen, scheint unbestritten. So ist es auch wohl zu erklären, daß die zunächst



zu Forschungszwecken entwickelten Paradigmen des „Komplexen Problemlösens“ (Überblicke hierzu bei Dörner, 1989; J. Funke, 1991) sowohl in diagnostischer Hinsicht als auch mit Blick auf den Trainingsaspekt sehr bald Eingang in den Anwendungsbereich fanden (siehe den Beitrag von U. Funke in diesem Heft). Hauptgedanke war es, daß die Teilnehmer aus ihren Fehlern beim spielerischen Umgang mit einem vorgegebenen simulierten System lernen sollten.

Doch schon frühzeitig hat Neuhauser (1976) in Hinblick auf „Business Games“ konstatiert, daß die größten Lerngewinne nicht bei den Teilnehmern auftraten, die mit diesen Systemen ihre Erfahrungen sammeln dürfen, sondern bei den Konstrukteuren solcher Systeme. Folglich ist es nur konsequent, die zu trainierenden Personen in die Rolle von Modellbildnern zu bringen. Dabei kann man insofern von einer abstrakten Arbeitsprobe sprechen, als das Bilden und Überprüfen von Modellen durchaus zu den zentralen kognitiven Leistungen einer Person gezählt werden kann. Wenngleich erste Versuche eines solchen Vorgehens bei Schülerinnen und Schülern im naturwissenschaftlichen Unterricht nicht zu dem gewünschten Erfolg führten (vgl. Klieme & Maichle, 1991), sind Berichte aus amerikanischen Studien bei Management-Studierenden wie auch bei erfahrenen Führungskräften erfolgverheißend (vgl. Senge & Sterman, 1992).

Ähnlich sieht dies auch Dörner (1989, S. 305), der das bloße Agieren mit einer „Mikrowelt“ für nicht ausreichend hält, sondern die zu trainierenden Personen einer „Anforderungssymphonie“ verschiedener Systeme aussetzen möchte (vgl. die ähnliche Argumentation von Schmidt & Bjork, 1992, über die Bedingungen effizienten Transfers aus Trainingssituationen). Graham, Morecroft, Senge und Sterman (in Druck) schlagen vor, für das Training von Führungskräften eine Kombination von klassischer Fallstudie und entsprechendem Simulationsmodell zu verwenden. Diese Art der Unterrichtung „systemischen Denkens“ hat ihrer Ansicht nach weitaus größeren Wert als das reine Üben an einer vorgegebenen Mikrowelt. Die aktive Auseinandersetzung mit einem Realitätsbereich, über den ein Modell erst noch zu formulieren und zu testen ist, hat etwas mit wissenschaftlicher Aktivität gemeinsam: das Aufstellen und anschließende Prüfen von Hypothesen. Der daraus zu ziehende Lerngewinn ist allemal größer als die bloße Rezeption vorgegebener Inhalte und macht daher den Wert einer Modellbildung aus. Neben dem diagnostischen Aspekt hat diese Art von Arbeitsprobe somit auch noch einen pädagogischen Nebeneffekt.

#### 4.2 Sozialpsychologische Perspektive

Aus sozialpsychologischer Perspektive scheint mir die Entwicklung computergestützter Verfahren dort erwartbar zu sein, wo soziale Situationen simuliert werden können; aber auch der vergleichsweise wenig erforschte Bereich der verteilten Entscheidungsfindung kann hier vielversprechende neue Wege aufzeigen.

##### 4.2.1 Simulierte soziale Situationen

Die klassische Fallstudie (vgl. Kosiol, 1961) verlangt geradezu nach einer Computerimplementation, insbesondere für diejenigen Varianten, in denen mehrstufige Entscheidungen in verzweigter Form abverlangt werden. Wie Kaiser (1976) berichtet, wurde das Verfahren der Fallstudie zunächst als rein didaktisches Instrument zu Beginn dieses Jahrhunderts an der Harvard Business School entwickelt (weshalb das Verfahren auch gelegentlich „Harvard-Methode“ genannt wird). Ihr Einsatz zu diagnostischen Zwecken wurde erst später vorgeschlagen.

Eine mögliche Präsentationstechnik besteht in Form eines „Entscheidungs-Labyrinths“. Dabei wird verlangt, sich ausgehend von einem Start-Szenario durch die jeweils angebotenen Optionen hindurchzuarbeiten, indem man je nach gewählter Option zu einer bestimmten Seite des meist in Heftform dargebotenen Seiten-Labyrinths verzweigt, um dann letztlich zur Lösung zu gelangen. Nach einer Untersuchung von Bund (1990) über die Qualität eines derartigen Entscheidungs-Labyrinths, das im Rahmen einer Personalentwicklungsmaßnahme von 141 Probanden bearbeitet wurde, ist der Beitrag einer solchen Übung zur Gesamtpotentialaussage bisher gering. Dies sei vor allem auf die standardmäßige Auswertungsmethode zurückzuführen, die nur quantitative Aspekte der Lösungsfindung (z.B. Schrittzahl) berücksichtige. Qualitative Bearbeitungsaspekte bleiben unberücksichtigt. Zudem wird kritisiert, daß bei Entscheidungs-Labyrinthen herkömmlicher Art ein nicht regelgemäßes Durchblättern des Lösungsheftes kaum verhindert werden kann. Hier hat die rechnergestützte Präsentation deutliche Vorteile durch die genaue Kontrolle der dargebotenen Information und der Protokollierung des eingeschlagenen Weges. Zusätzlich ist denkbar, die Entscheidungen unter Zeitdruck abzuverlangen bzw. die Zeiten zu registrieren. Verglichen mit Kurzfällen, in denen für eine kurze Problemskizze eine einzige multiple-choice-Antwort erfaßt wird, ist das Entscheidungs-Labyrinth – so Bund (1990, S. 169) – auch durch den möglichen Einbezug wörtlicher Rede lebendiger und kann von der Art der Konfliktdarstellung auch für höhere Führungskräfte verwendet werden. Einer Übertragung dieser Testmaterialien auf einen Rechner steht m. E. nichts im Wege. Wie der nächste Abschnitt zeigt, muß man sich dabei nicht auf die reine Textpräsentation am Bildschirm beschränken.

Computergestütztes Interaktives Video (CIV, auch „Dialogvideo“ genannt) erlaubt über ein entsprechendes Interface die Ansteuerung eines Videorekorders bzw. einer Bildplatte durch einen Rechner und damit die Präsentation der audiovisuellen Information auf dem Rechnermonitor in Form stehender oder bewegter Bilder mit Mono- bzw. Stereoton. In Abhängigkeiten von Benutzereingaben (über Tastatur, Maus, Lichtgriffel, etc.) kann auf beliebige, exakt auffindbare Positionen verzweigt werden. Alle Video-Funktionen sind vom Rechner aus programmierbar. Je nach Interaktion des Benutzers am Bildschirm des Steuerrechners kommen somit andere Sequenzen zustande.

Die Verwendung lebensnaher Situationen als Testreize, die von angehenden Führungskräften zu beurteilen sind, und deren weiterer situativer Verlauf – ganz im Sinne des Entscheidungs-Labyrinths – von deren Entscheidungen abhängig ist, stellt eine Bereicherung der diagnostischen Möglichkeiten dar (vgl. Schuler, Diemand & Moser, 1993). Führungsverhalten in lebensnahen Situationen (Szene: Der Chefsekretär wird von einer Untergebenen beschimpft, die beobachtende Person – die Testperson – greift ein gemäß einer der fünf folgenden Optionen: ...) wie Verhandlungsstile (Szene: Die Verhandlungspartnerin will die Entscheidung aufschieben; die beobachtende Person rettet die Situation durch eine der folgenden Optionen: ...) können damit auf eine neue Weise zugänglich gemacht werden und müssen nicht ausschließlich über vorgestellte oder durch Texte beschriebene soziale Situationen angeregt werden.<sup>4</sup>

#### 4.2.2 „Distributed Decision Making“

In dem kürzlich erschienen Reader zum Thema „Distributed Decision Making“ von Rasmussen, Brehmer und Leplat (1991) wird argumentiert, daß sich im letzten Jahrzehnt die Anforderungen der Arbeitswelt geändert haben. Häufig genug gebe es heute Anforderungen, die nur noch im Team zu bewältigen sind. Obwohl die genannten Autoren dabei in erster Linie an Prozeßsteuerungen in industriellen Kontexten denken, die von Teams (Mannschaften, Crews) bearbeitet werden, nennt Brehmer (1991) aber auch klassische Führungssituationen etwa im militärischen Bereich, wo die Notwendigkeit zu Abstimmungen des Vorgehens durch die zunehmend rascher veränderlichen Situationen unerlässlich geworden ist.

Untersuchungen darüber, wie verteiltes Entscheiden erfolgt und wie man die Bereitschaft einer einzelnen Person erfaßt, solche Gruppenentscheidungen vorzubereiten und in der Umsetzung mitzutragen, sind noch rar, werden aber sicher verstärkte Aufmerksamkeit finden müssen. Die von Brehmer vorgeschlagene Untersuchungsmethodologie, wonach mehrere Personen an einem PC-Netzwerk ein simuliertes Problem gemeinsam bearbeiten und jedem einzelnen Teilnehmer nur ein bestimmter Ausschnitt des Szenarios verfügbar ist, stellt eine mögliche Vorgehensweise dar.

## 5 Abschließende Bemerkungen

Der gegenwärtige Zustand computerunterstützter Diagnostik auf dem Feld der Arbeitsproben erweist sich als noch unzureichend, wenn man seine empirischen Fundamente betrachtet. Anstelle eindeutiger psychometrischer Gütekennwerte wird vielfach nur das Argument der Augenscheinvalidität bzw. das der hohen Akzeptanz solcher Verfahren durch Teilnehmerinnen und Teilnehmer strapaziert.

Dieser unbefriedigende Zustand ist aber kein prinzipielles Problem, sondern weist allenfalls auf den hohen Forschungsbedarf hin. Angesichts der generell positiven Einschätzung von Arbeitsproben als diagnostischem Instrument und auch angesichts möglicher Entwicklungslinien, die weiter oben aufgezeigt wurden, dürfen an diesen Bereich hohe Erwartungen gestellt werden. Dieser Optimismus wird allenfalls gemindert durch die Beobachtung (vgl. Ahlburg, 1992), wonach Personen mit Personalentscheidungskompetenz sich nicht an den Ergebnissen wissenschaftlicher Untersuchungen zur differentiellen Effektivität verschiedener Auswahlverfahren (etwa Hunter & Hunter, 1984) orientieren. Aber dieses eher generelle Problem der Übertragung wissenschaftlicher Erkenntnisse in die konkrete Praxis der Personalauswahl kann nicht mit wissenschaftlichen Mitteln gelöst werden.

## Literatur

- Ahlburg, D. A. (1992). Predicting the job performance of managers: What do the experts know? *International Journal of Forecasting*, 7, 467–472.
- Arnold, W. (Hrsg.). (1951). *Der Pauli-Test. Seine sachgemäße Durchführung und Auswertung*. München: Johann Ambrosius Barth.
- Asher, J. J. & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, 27, 519–533.
- Brehmer, B. (1991). Distributed decision making: Some notes on the literature. In J. Rasmussen, B. Brehmer & J. Leplat (Eds.), *Distributed decision making: Cognitive models for cooperative work* (pp. 3–14). New York: Wiley.
- Bund, U. (1990). *Die Fallstudie im Assessment-Center* (Unveröffentlichte Diplomarbeit). Bonn: Psychologisches Institut der Universität Bonn.
- Christiansen, E. (1983). *Die Arbeitskurve nach Emil Kraepelin und Richard Pauli. Manual*. Weinheim: Beltz.
- Didi, H.-J. (1992). *Konstruktvalidierung und Weiterentwicklung einer Postkorb-Aufgabe im Rahmen eines Assessment-Centers* (Unveröffentlichte Diplomarbeit) Bonn: Psychologisches Institut der Universität Bonn.
- Dörner, D. (1989). *Die Logik des Mißlingens*. Hamburg: Rowohlt.
- Fricke, R. (1990). Videotests: „True-to-life“-Testsituationen durch interaktives Video. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 463–466). Göttingen: Hogrefe.
- Funke, J. (1991). Solving complex problems: Human identification and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–224). Hillsdale, NJ: Erlbaum.
- Funke, J. & Rasche, B. (1988). *PC-POSTKORB. Manual zur PC-Version der Postkorb-Übung* (Unveröffentlichtes Manuskript). Bonn: Psychologisches Institut der Universität Bonn.
- Giese, F. (1924). Die Arbeitsprobe in der Psychognostik. *Zeitschrift für angewandte Psychologie*, 23, 162–187.
- Graham, A. K., Morecroft, J. D. W., Senge, P. M. & Sterman, J. D. (in Druck). Model-supported case studies for management education. *European Journal of Operational Research*.
- Guthke, J. (1982). The learning test concept – An alternative to the traditional static intelligence test. *The German Journal of Psychology*, 6, 306–324.

<sup>4</sup> An der University of Wisconsin in Madison, Graduate School of Business, ist von Randall Dunham ein Programm namens „Manager's Workshop“ für didaktische Zwecke entwickelt worden, das diese Vorstellungen bereits realisiert (vgl. „Syllabus for the Macintosh“, March/April 1991, Number 16).

- Hasselmann, D. (1993). Eignungsdiagnostische Validität des computersimulierten Szenarios TEXTIL-FABRIK. In A. Gebert & U. Winterfeld (Hrsg.), *Arbeits-, Betriebs- und Organisationspsychologie vor Ort. Bericht über die 34. Fachtagung der Sektion Arbeits-, Betriebs- und Organisationspsychologie des BDP in Bad Lauterberg* (S. 541–550). Bonn: Deutscher Psychologen Verlag.
- Horn, R. (1991). MAILBOX: A computerized in-basket task for use in personnel selection. *Revue Européenne de Psychologie Appliquée*, 1, 325–327.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.
- Jäger, R. S. (1990). Computerdiagnostik – Ein Überblick. *Diagnostica*, 36, 96–114.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern. Assessment-Center-Verfahren*. München: Hanser.
- Jeserich, W. (1982). Das Assessment-Center-Verfahren. *Zeitschrift für Betriebswirtschaftliche Forschung*, 34, 365–373.
- Kaiser, F.-J. (1976). *Entscheidungsstraining. Die Methoden der Entscheidungsfindung (Zweite, erweiterte und verbesserte Auflage)*. Bad Heilbrunn: Julius Klinkhardt.
- Klieme, E. & Maichle, U. (1991). *Erprobung eines Modellbildungssystems im Unterricht. Bericht über eine Pilotstudie zur Unterrichtsevaluation im Auftrag des Landesinstituts für Schule und Weiterbildung, Soest*. Bonn: Institut für Test- und Begabungsforschung.
- Kosiol, E. (1961). Praktische Fälle im betriebswirtschaftlichen Hochschulunterricht. *International Review for Management and Managerial Sciences*, 2, 91–103.
- Maukisch, H. (1990). Eignungsdiagnostik als prognostische Hilfe bei der Auswahl, Plazierung und Entwicklung von Führungskräften. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 32–46). Göttingen: Hogrefe.
- Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.
- Moede, W. (1930). *Lehrbuch der Psychotechnik, I. Band*. Berlin: Julius Springer.
- Moede, W. (1943). *Eignungsprüfung und Arbeitseinsatz*. Stuttgart: Ferdinand Enke.
- Muchinsky, P. M. (1990). *Psychology applied to work An introduction to industrial and organizational psychology*. Pacific Grove, CA: Brooks/Cole.
- Neuhauser, J. (1976). Business games have failed. *Academy of Management Review*, 1, 124–129.
- Rasmussen, J., Brehmer, B. & Leplat, J. (Eds.). (1991). *Distributed decision making: Cognitive models for cooperative work*. New York: Wiley.
- Robertson, L. T. & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. *Journal of Applied Psychology*, 74, 402–410.
- 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 und Beltz.
- Rösler, F. (1988). Personalauslese, Training und Personalentwicklung in Organisationen. In D. Frey, C. G. Hoyos & D. Stahlberg (Hrsg.), *Angewandte Psychologie. Ein Lehrbuch* (S. 65–91). München: Psychologie Verlags Union.
- Schippmann, J. S., Prien, E. P. & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology*, 43, 837–851.
- Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–217.
- Schuler, H. (1990). Personenauswahl aus der Sicht der Bewerber. Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184–191.
- Schuler, H., Diemand, A. & Moser, K. (1993). Filmszenen. Entwicklung und Konstruktvalidierung eines neuen eignungsdiagnostischen Verfahrens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 3–9.
- Schuler, H. & Funke, U. (1993). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 235–283). Bern: Huber.
- Schuler, H., & Moser, K. (1990). Geschichte der Managementdiagnostik. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 18–28). Göttingen: Hogrefe.
- Schulte, R. W. (1921). *Die Berufseignung des Damenfriseurs. Methoden und Ergebnisse eines psychotechnischen Prüfsystems auf der Grundlage einer Funktionsanalyse (= Schriften zur Psychologie der Berufseignung und des Wirtschaftslebens, Band 17)*. Leipzig: Johann Ambrosius Barth.
- Senge, P. M. & Serman, J. D. (1992). Systems thinking and organizational learning: Acting locally and thinking globally in the organization of the future. *European Journal of Operational Research*, 59, 137–150.
- Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372–376.

*Anschrift des Verfassers:* PD Dr. Joachim Funke, Universität Bonn, Psychologisches Institut, Römerstr. 164, 53117 Bonn.

Eingegangen: 2. 9. 1992

Revision eingegangen: 29. 3. 1993

## Erratum

Im Beitrag von Schuler, H., Diemand, A. & Moser, K. (1993). Filmszenen. Entwicklung und Konstruktvalidierung eines neuen eignungsdiagnostischen Verfahrens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37. Jg. (N. F. 11), 3–9, fehlt auf S. 8 in Tabelle 1 bei „Kognitionsaufgaben“ und „Konzentrations-Arb.-pr.“ in der Spalte „Kundenorientierung“ die Anmerkung „\*\*“; in Tabelle 2 fehlt die Anmerkung „<sup>1</sup> Nur Ladungen > .40 sind wiedergegeben.“