

# MANAGEMENT-FALLSTUDIEN (MFA)<sup>1</sup>

Bernd Reuschenbach, Joachim Funke

## 1 TESTART

Computergestütztes Verfahren zur Personalauswahl und Personalentwicklung als Kombination aus Leistungs- und Persönlichkeitstests

## 2 TESTMATERIAL

Die Materialien zu den Management-Fallstudien werden in einem blauen Hartschalenkoffer geliefert. Zum Inhalt gehören eine 58 Seiten „starke“ Handanweisung, eine CD-ROM und ein Abrechnungs-Dongel (MFA-Decoder). Außerdem sind die vier Verfahren der MFA auch als Papier- und Bleistift-Version (PuB-Version) beigelegt. Hierzu gehören vier Testhefte und drei verschiedene Antwortbögen. Zur PuB-Bearbeitung des vierten Verfahrens (MFA-5) müssen von den Testleitenden leere Blätter gestellt werden.

Zur Installation der Computerversion wird ein handelsüblicher PC (486er oder Pentium) mit mindestens 16 MB Speicher (RAM), 20 MB freiem Speicherplatz auf der Festplatte und einem CD-ROM-Laufwerk benötigt. Der MFA-Decoder, der auf die Druckerschnittstelle aufgesteckt werden muss, dient als Kopierschutz und Abrechnungsadapter. Das Testverfahren kostet laut Hogrefe-Katalog 3450,- DM. Beim ersten Kauf sind 50 Anwendungen für jeweils eines der vier Teilverfahren freigeschaltet, was etwa 12 Komplettdurchläufen entspricht. Vier weitere Durchführungen (also die Nutzung aller vier Teilverfahren für eine Person) kosten 100,- DM. Weitere Durchführungslizenzen können durch Kauf von Freischaltungs-Codes erworben werden.

---

<sup>1</sup> Georg P. Fennekels & Simone D'Zouza. Göttingen: Hogrefe, 1999; Rezensenten: Bernd Reuschenbach & Joachim Funke, Psychologisches Institut der Universität Heidelberg. Für hilfreiche Anmerkungen zu einer Erstfassung danken wir Dr. Marlene Endepohls, Handschuhshheim. Besonderer Dank geht an Dr. Ernst Fay, Bad Godesberg, für seine zugleich konstruktiven wie humorigen Kommentare.

### 3 TESTGLIEDERUNG

Die Management-Fallstudien (MFA) bestehen aus vier einzelnen Verfahren, in denen typische Probleme des betrieblichen Alltags geschildert werden:

- MFA-K (Fallstudie zur Konfliktbewältigung),
- MFA-F (12 Kurzfälle zu Führungssituationen),
- MFA-S (9 Kurzfälle zu sozialen Situationen),
- MFA-Z (Fallstudie zur Selbstorganisation, Planung und zum Zeitmanagement).

Jedes dieser vier Verfahren beginnt in der PC- und in der PuB-Version mit einer Instruktion, die auf der ersten Seite des Testheftes bzw. einem separaten Fenster auf der Bildschirmoberfläche präsentiert wird. Die vier einzelnen Teile des Verfahrens werden nachfolgend vorgestellt.

#### 3.1 MFA-Konfliktsituation (MFA-K)

In der MFA-Konfliktsituation soll der Konflikt mit einem Mitarbeiter gelöst werden. Die Testperson (Tp) soll sich vorstellen, sie sei Abteilungsleiter in einer Tochtergesellschaft eines Unternehmens. Ein Mitarbeiter (Herr Köppler) zeige in der letzten Zeit deutliche Leistungsverlechterungen. Ausgehend von dieser Situation soll die Tp dann Entscheidungen treffen, so dass die Ursachen der Leistungsverlechterung – Fennekels und D'Zouza sprechen von einem „Konflikt“ – behoben werden. Nach welchen Kriterien Entscheidungen getroffen werden sollen oder wie das Ergebnis der Konfliktlösung ausfallen soll, wird in der Instruktion nicht erwähnt. Die Tp wird mit dem Satz: „Entscheiden Sie sich für eine Alternative“ (S.1 / Testheft) allein gelassen. Nach der Instruktion folgt die erste Aufgabe, eingeleitet mit dem Satz: „Sie beschließen, dem Leistungsabfall von Herrn Köppler auf den Grund zu gehen und die Angelegenheit zu klären. Was werden Sie tun?“ Es stehen zunächst 5 Antwortalternativen zur Verfügung. In der Computerversion klickt die Person den Buchstaben der ihrer Meinung nach sinnvollsten Alternative an und gelangt nach einem Klick auf „Weiter“ automatisch zu einer der insgesamt 40 möglichen Folgeseiten. In der PuB-Version steht hinter der Antwortalternative eine Seitenzahl, die als nächstes aufgeschlagen werden soll. Die gewählte Seite muss von der Tp in einen separaten Testbogen eingetragen werden. Analog zum adaptiven Testen bestimmt das Antwortverhalten der Tp, welche Fragestellung als nächstes folgt. Wird von der Tp beispielsweise von den fünf möglichen Reaktionen die Alternative A ausgewählt („Sie fragen Ihren eigenen Vorgesetzten, um herauszufinden, was in diesem Unternehmen bei solchen Fällen in der Regel unternommen wird“), so muss in der PuB-Version als nächstes die Sei-

te A7<sup>2</sup> aufgeschlagen werden.

In der Computerversion gelangt die Tp automatisch zur entsprechenden Seite. Hier warten erneut verschiedene Antwortalternativen, von der aus weitere Seiten aufgerufen werden müssen. Die Anzahl der Antwortalternativen pro Seite liegt zwischen eins und sechs. Unter den vierzig Seiten, die aufgerufen bzw. aufgeblättert werden können, befinden sich vier Abschlusseiten, die das Problem als gelöst erklären. Die Lösung des Konfliktes ist erreicht, wenn die Tp eine dieser vier Endseiten erreicht hat. Diese vier Lösungen werden im Testmanual beschrieben als

- „konstruktive Lösung des Problems unter Beachtung von Mitarbeiter- und Firmeninteressen,
- „Aussitzen“ der Problems,
- Ausweichen der Problemlösung durch Versetzung des Mitarbeiters,
- radikale Lösung durch Kündigung.“ (S.41).

Der Versuch, den Konflikt zu lösen, muss aber nicht auf einer dieser vier Seiten enden, da sich die Tp auch in Schleifen „festfahren“ kann. Solche Schleifen entstehen, wenn einmal aufgerufene Seiten später erneut erreicht werden und mit einer der bereits gewählten Alternativen beantwortet werden. Eine solche „Schleife“ wäre beispielsweise die Seitenfolge A1->A7->A1->A7->A1, oder – weniger trivial, da nicht sofort erkennbar – A22->A33->A9->A4->A22 usw.

Um solchen Schleifen, die immer wieder zu den selben Fragestellungen zurückführen, zu entgehen, müssen andere Alternativen ausgewählt werden. Die Tp könnte beispielsweise mittels *trial-and-error*-Strategie nach anderen Lösungswegen suchen, die dann nicht mehr ihren ursprünglichen Absichten entsprechen und damit auch nicht mehr Spiegelbild ihrer eigentlichen Konfliktlösestrategie sind. Die Testpersonen in der PuB-Version haben es aus zwei Gründen wesentlich leichter, im Sinne einer solchen *trial-and-error*-Strategie zu einer der vier Abschlusseiten zu gelangen:

- Durch die Nummerierung der Seiten in der PuB-Version ist schneller zu erkennen, dass eine Seite schon einmal bearbeitet wurde. Ein Beispiel: Die Testperson befindet sich bei der Aufgabe A6, wählt dann Alternative B. Sie gelangt zur Aufgabe A22, dort wählt sie Alternative A und gelangt so zur Aufgabe A21. Den Weg dokumentiert sie im Lösungsblatt. Auf der Seite A21 angekommen, wird sie sich kaum für die Alternative D entscheiden, die sie

<sup>2</sup> Die Seitenzahl befindet sich in der rechten Ecke der Kopfzeile. Diese Seitenzahl muss in den Antwortbogen eingetragen werden. Unklar bleibt, ob diese Antworten von der Versuchsperson ausradiert und dann revidiert werden dürfen.

wieder zur Aufgabe A6 zurückführt, da sie A6 bereits auf ihrem Lösungsblatt markiert sieht. Demgegenüber weiß die Tp in der Computerversion nicht, bei welcher Aufgabe sie sich gerade befindet, da die Seiten nicht mit A6, A21 oder A22 usw. gekennzeichnet sind. Sie muss den gesamten Text durchlesen um festzustellen, dass sie die Seite bereits früher gelesen hat. Außerdem weiß die Tp in der Computerversion nicht explizit, zu welchen Folgeseiten die Auswahl einer Alternative führt.

- Es kann nicht ausgeschlossen werden, dass die Tp beim Durchblättern des Testheftes zu einer der vier Abschlusseiten gelangt. Hierbei könnte sie sich beispielsweise die Seitenzahl merken und diese Alternative auf einer der vorangehenden Seiten auswählen.

Innerhalb von 15 Minuten (in beiden Versionen) sollte eine der vier Schlusseiten erreicht werden. Die Anzahl der Seiten, die auf die jeweilige Lösungsseite führen, ist unterschiedlich. Zu der - laut Handanweisung - besten Lösung (Ax)<sup>3</sup> gelangt man nur über eine Seite. Auf die anderen drei Abschlusseiten verweisen ein bis drei vorangehende Seiten.

### 3.2 MFA-Soziale Situationen (MFA-S)

Bei der MFA-Soziale Situationen steht die „Effektivität der Konfliktbewältigung“ (S. 8)<sup>4</sup> und die „Sensibilität des Teilnehmers zwischenmenschliche Problem wahrzunehmen“ (S. 8) im Mittelpunkt. Es werden neun verschiedene Problemsituationen des betrieblichen Alltags beschrieben, die durch die Auswahl von Antwortalternativen (Computerversion) bzw. durch freie Schilderungen (PuB-Version) gelöst werden sollen. Die geschilderten Fälle reichen von Alkoholproblemen am Arbeitsplatz über Unzufriedenheit der Mitarbeiterinnen und Mitarbeiter bis hin zu einem Konflikt mit einem Mitarbeiter, der sich bei einem bedeutsamen Geschäftstermin ungünstig in Szene setzt (s.u.).

Zwischen der Computerversion und der PuB-Version bestehen wesentliche Unterschiede. Diese betreffen a) die Instruktion, b) die Antwortmöglichkeit und c) die Bearbeitungszeit.

- a) In der Computerversion soll die Tp laut Instruktion von den vorgegebenen Antwortalternativen die auswählen, die dem eigenen Verhalten am nächsten kommen würden. Hierbei sind Mehrfachantworten möglich. Die Anzahl der Antwortalternativen in der Computerversion schwankt zwischen 3 und 7. Als dezenter Hinweis auf die verlangten Leistungen ist die Instruktion mit „Bewältigung sozialer Situationen“ überschrieben, darüber hinaus werden keine Angaben zu den Maßstäben der Beurteilung gemacht. Die Instruktion in der PuB-Version ist – was die Anzahl der Wörter betrifft – genau doppelt so lang. Dort heißt es: „Bitte versetzen Sie sich in die Lage der Mitarbeiter und Führungskräfte und beschreiben Sie, wie Sie sich verhalten würden. Wichtig ist, dass Sie konkret angeben, was Sie veranlassen und entscheiden. Nehmen Sie Stellung! Ob Sie nun eine oder mehrere Maßnahmen ergreifen, bleibt Ihnen überlassen.“
- b) Während in der Computerversion zwischen drei und sieben Antwortalternativen ausgewählt werden können, muss in der PuB-Version die Problembewältigung in Form eines freien Textes geschildert werden. Die Aufgabe der Testleiterin/des Testleiters (TL) besteht darin, diese freie Schilderung später den Antwortalternativen der Computerversion zuzuordnen.
- c) Für die Bearbeitung sind in der Computerversion 30 Minuten vorgesehen. Ob die lange Bearbeitungszeit von 30 Minuten im Vergleich zu den 15 Minuten bei der MFA-K gerechtfertigt ist, muss offen bleiben. Die Bearbeitungszeit für die PuB-Version der MFA-S ist in der Handanweisung mit 60 Minuten und im Testheft mit 45 Minuten angegeben.

Die Computerversion erlaubt das Rückblättern zu den einzelnen Problemen und auch die Neuauswahl von Alternativen. Darauf deutet das anklickbare Wort „Zurück“ hin. Auch die Handanweisung verweist auf die Möglichkeit, die Antworten beim Zurückblättern nochmals zu verändern. „Soll eine Entscheidung wieder rückgängig gemacht werden, klickt man einfach nochmals auf den Buchstaben, so dass die Markierung verschwindet“ (S. 28). In der Instruktion der Computerversion wird diese Möglichkeit aber nicht erläutert. Vermutlich können nur erfahrene Computerbenutzer diese Option erahnen und dann auch nutzen. Ein expliziter Hinweis könnte gerade für Computernovizen hilfreich sein.

### 3.3 MFA-Führungssituationen (MFA-F)

Bei den MFA-Führungssituationen werden zwölf Fälle geschildert, bei denen der Einsatz „diverser Führungstechniken“ (S. 8) verlangt wird. Laut Handanweisung werden folgende Anforderungen gestellt: „Initiative ergreifen und Abläufe gestalten / Führungstechniken angemessen einsetzen“ (S. III). Die Tp soll sich dazu in

<sup>3</sup> Um keine Lösungswege offenzulegen, wird die ideale Lösungsseite im Folgenden mit Ax, alle übrigen Seiten mit Aa, Ab, Ac usw. bezeichnet.

<sup>4</sup> Alle Angaben von Seiten, Tabellen oder Abbildungen ohne nähere Quellenangabe beziehen sich auf die Handanweisung (Fennekels & D'Zouza, 1999).

die Lage der jeweiligen Führungskraft versetzen und bei der Auswahl der Alternativen die „beschriebenen personellen und/oder sachlichen Bedingungen“ (S. 1) beachten. Ziel ist es, die Leistungsfähigkeit der Mitarbeitenden zu verbessern bzw. abzusichern und dabei die unterschiedlichen Kompetenzen der Teammitglieder zu beachten. „Unter Kompetenz verstehen wir die momentan erkennbare Bereitschaft und Fähigkeit aller Teammitglieder, fachlich und menschlich angemessen zu handeln“ (S. 1). Nach der Instruktion folgen zwölf Situationen, in denen es bevorzugt um Teamkonflikte und Leistungsrückgänge<sup>5</sup> der Mitarbeitenden geht.

In der Computerversion wählt die Tp per Mausclick eine der vier Antwortalternativen und bestätigt die Wahl mit „Weiter“. Es erscheint dann der nächste Fall. Ein Zurückblättern ist möglich, wird aber in der Instruktion nicht explizit erwähnt. In der PuB-Version muss die Tp die ausgewählten Alternativen auf einem gesonderten Testbogen eintragen.

### 3.4 MFA-Zeitmanagement (MFA-Z)

Bei der MFA-Zeitmanagement handelt es sich um eine Planungsaufgabe, bei der innerhalb einer vorgegebenen Zeit (30 Minuten am Computer bzw. 45 Minuten bei der PuB-Version) verschiedene Termine unter Beachtung der Rahmenbedingungen in eine zeitliche Reihenfolge gebracht werden sollen. Gemessen werden soll die „Planungseffizienz und das Strukturierungsvermögen des Teilnehmers unter Zeitdruck“ (S. 7).

Die Tp wird instruiert, dass sie einen halben Tag im Arbeitsalltag eines Mitarbeiters einer Außenstelle im Ausland planen soll. Der Tag beginnt um 8.30 in der Geschäftsstelle und soll auch dort um spätestens 12.00 Uhr enden. Zwischendurch sind verschiedene Erledigungen zu machen (z.B. Abholen von Devisen in der Bank, ein Besuch des Arztes, ein Besuch im Krankenhaus usw.). Diese Erledigungen können nicht frei aneinander gereiht werden, sondern werden durch feste Terminvorgaben mitbestimmt. So muss beispielsweise der Impfpass um 11.10 Uhr abgeholt werden und um 11.27 Uhr eine Warenübergabe am Bahnhof stattfinden. Die Transferzeiten von einem Ort zum anderen sind auf einem Stadtplan eingetragen, ebenso die Verweildauern an den verschiedenen Stellen.

<sup>5</sup> Wie bei den MFA-Konfliktsituationen stehen auch hier Leistungsrückgänge der Mitarbeitenden im Mittelpunkt, was eine inhaltliche Abgrenzung der beiden Teilverfahren erschwert.

Es werden 3 bis 4 Stunden eines Arbeitsalltages simuliert. Die Tp soll herausfinden, „in welcher Reihenfolge die notwendigen Besorgungen effektiv zu erledigen sind“ (S. 1/ Testheft). Für die Planung hat sie 30 bzw. 45 Minuten Zeit. In der Computerversion klickt die Tp dazu auf das anvisierte nächste Ziel und der Computer errechnet automatisch die Zeit bis zur Ankunft am ausgewählten Ort zuzüglich der dortigen Verweildauer, daher verwundert etwas der Hinweis „Runden Sie (sic!) die Zeiten für die Fahrten mit dem Motorroller“ (Instruktion MFA-Z Computerversion). So können nacheinander alle 11 Termine der „to do“-Liste in die Liste der erledigten Termine überführt werden. In der PuB-Version muss die Tp die Wegezeiten selbständig ausrechnen und auch die Verweildauer beachten. Während einige Termine laut Instruktion eingehalten werden „müssen“, wird ein entscheidender Termin, die Reparatur des Rollers, als eher fakultative Maßnahme dargestellt: „Wegen der chaotischen Verkehrsbedingungen bleibt Ihnen nichts anderes übrig, als zu Fuß zu gehen. Der Hausmeister hat zwar angeboten, Ihnen seinen Motorroller auszuleihen. Dieser hat aber einen Kupplungsschaden und kann z.Z. nur im Schrittempo gefahren werden. Es sei denn, Sie lassen ihn reparieren. Die Reparatur dauert eineinhalb Stunden“ (S. 1/ Testheft MFA-Z).

Die Instruktion bleibt während des ganzen Verfahrens abrufbar. Der zu berücksichtigende Stadtplan ist in der Computerversion hinter der Instruktion positioniert. Er kann durch einen Klick auf die entsprechende Schaltfläche in den Vordergrund gebracht werden. In der PuB-Version liegen Instruktion und Stadtplan nebeneinander. Für die PuB-Version bleibt die Frage ungeklärt, ob ein leeres Blatt zum Rechnen verwendet werden darf und ob eine Änderung des vorläufigen Plans durch Nutzung eines neuen Blatts oder durch Ausradieren des bisherigen Lösungsweges erlaubt ist. Eine Revision des Plans in der Computerversion ist möglich, wird aber nicht explizit erläutert.

Wichtig gerade in Hinblick auf Planrevisionen ist, dass in der PuB-Version der gesamte Lösungsweg samt den Zwischenzeiten auf dem Papier sichtbar bleibt, wohingegen in der Computerversion auf dem Bildschirm nur noch die Reihenfolge der Orte und die jeweils erreichte Endzeit sichtbar sind.

## 4 GRUNDKONZEPT

Die Entwicklung dieser computergestützten Fallanalysen wird von Fennekels und D'Zouza (1999) mit den bisherigen Defiziten bei anderen Fallstudien begründet. „Wir benötigen verhaltensnahe Situationen, die den Rahmenbedingungen und Kriterien der Eignungsdiagnostik genügen“ (S. 36). Ob die Fallanalysen zur Selektion, zur Potentialeinschätzung oder zur Beurteilung von Mitarbeiterinnen und Mitarbeitern eingesetzt werden sollen, dürfen oder können, bleibt unklar.

Um das erste Kriterium der Verhaltensnähe zu erfüllen, wurden die Fälle mittels „critical incident technique“ (CIT; Flanagan, 1954) entwickelt. An der Auswahl der Items waren Mitarbeitende aus dem Bereich Personalwesen/Personalentwicklung und Linienmanager aus unterschiedlichen Bereichen beteiligt. Um den „Kriterien der Eignungsdiagnostik“ gerecht zu werden, wurden die Verfahren in einem Assessment-Center eingesetzt. Anschließend wurden die Ergebnisse mit Potential-einschätzungen und Daten über die Karriereentwicklung verglichen. Die MFA sollen Aussagen zu folgenden drei Bereichen ermöglichen:

- *Soziale Kompetenz*: „Aktives Gestalten von zwischenmenschlichen Konfliktsituationen und Abläufen“.
- *Führungsverhalten*: „Initiative ergreifen und Abläufe gestalten/Führungstechniken angemessen einsetzen“.
- *Systematisches Denken und Handeln*: „Aufgaben systematisch angehen/praxisgerechte Lösungen finden“.

Es handelt sich also um eine Kombination von Leistungstests (z.B. MFA-Z) und Persönlichkeitstest (z.B. MFA-F).

Die inhaltliche Füllung und Präzisierung der Dimensionen bleibt (wie auch die Auswertung zeigt) auffällig theoretisch und leitet sich allein aus den Konstruktionsprinzipien (CIT) und der prognostischen Validität (Beförderungspotential) ab. Fennekels und D'Zouza (1999) verfolgen also eine eher external geleitete Strategie der Testkonstruktion. Das genaue Vorgehen wird im Manual nur sehr kurz angesprochen. Zwingende Erklärungen zur Testkonstruktion (Itemkonstruktion und Itemselektion) fehlen, so dass es schwerfällt, die Arbeitsschritte bei der Erstellung der Testmaterialien kritisch zu bewerten.

Die Einbindung in eine Computerversion ist in erster Linie für den praktischen Einsatz vor Ort sinnvoll. Die Auswertung erfolgt automatisch und die Ergebnisse werden in Textform inhaltlich dargestellt. Diesem Vorteil steht der Nachteil gegenüber, dass eine Zuordnung der vier Verfahren zu den drei eben genannten Facetten nur unzureichend theoretisch und ebensowenig empirisch begründet wird. Weiterhin bleibt bei einigen Verfahren unklar, welche Items und Antwortalternativen welchen Dimensionen zugeordnet werden.

## 5 DURCHFÜHRUNG

In der Handanweisung wird eine durchschnittliche Durchführungsdauer von 2 Stunden genannt. Die Bearbeitungszeit für die PuB-Version ist länger (MFA-S: 60 Minuten laut Handanweisung, laut Testheft 45 Minuten, gegenüber 30 Minuten

bei der PC-Version; MFA-Z: 45 Minuten versus 30 Minuten bei der PC-Version). Als Begründung für die Zeitunterschiede zwischen PC- und PuB-Version wird folgendes angegeben: „Bei den MFA-S geschieht dies [Anm.: die Zeitverlängerung], da der Teilnehmer bei der PuB-Form im Gegensatz zur PC-Form die Entscheidung in einem selbst verfassten Text beschreibt. (...) Bei der MFA-Z wird bei der PuB-Bearbeitung mehr Zeit eingeräumt, da der Teilnehmer hier die Ankunftszeiten an den einzelnen Stationen selbst berechnen muss, während dies bei der PC-Bearbeitung durch das Programm automatisch geschieht“ (S. 15).

In der Computerversion kann die Zeitbegrenzung unter dem Menü „Option“ vom TI frei festgesetzt werden. Hierzu wird der Hinweis gegeben: „Bitte nehmen Sie dies nur in Abstimmung mit einem Testdiagnostiker vor“ (S. 14). Ein Hinweis auf die in einem derartigen Fall fehlenden Normen fehlt allerdings. In allen Verfahren der Computerversion wird in einem Zeitfenster die ablaufende Zeit standardmäßig eingeblendet<sup>6</sup>, allerdings spielt die Lösungszeit bei keinem der Verfahren in der Auswertung eine Rolle.

Über die Reihenfolge der Bearbeitung der einzelnen Verfahren liegen keine Empfehlungen vor, ebenso fehlt die Angabe, auf welche Bearbeitungssequenz sich die Normen beziehen. Möglicherweise ist es in bestimmten Fällen angebracht, nur Teile des Gesamtverfahrens zu bearbeiten. Auch hierzu liegen keine Empfehlungen vor.

Der Übergang von einem Testteil zum nächsten verlangt in der Computerversion die erneute Eingabe von Vor- und Nachnamen der Tp. Diese Prozedur behindert die Anwendung aller vier Teile als Gruppentest. Für jeden Testteil wird eine eigene Datei angelegt, die zur Abbuchung einer Abrechnungseinheit führt. Der jeweilige Stand noch verfügbarer Anwendungen wird bei jeder erneuten Abbuchung angezeigt.

## 6 Auswertung und Interpretation

Für die Auswertung der PuB-Version müssen die Ergebnisse zunächst im sogenannten „Auswertermodus“ in die Computerform überführt werden. Eine Auswertung der PuB-Ergebnisse ohne den Computer ist nicht möglich, da der Aus-

<sup>6</sup> Die Zeiteinblendung kann in den Voreinstellungen unterdrückt werden. Um eine bessere Vergleichbarkeit mit der PuB-Version zu erreichen, wäre es besser gewesen, keine Zeit einzublenden und dies zur Grundeinstellung zu machen.

wertungsalgorithmus nicht ersichtlich ist und selbst durch Testläufe nicht vollständig erschlossen werden kann. Die Auswertung erfolgt in der Computerversion (Teilnahmemodus und Auswertungsmodus) für alle vier Verfahren automatisch. Für jedes der vier Teilverfahren besteht die Auswertung aus der Darlegung der folgenden fünf Teile:

- 1) dem gewählten Lösungsweg bzw. einer Auflistung von richtigen und falschen Antworten,
- 2) dem Rohwert und der Angabe der maximal möglichen Punkte,
- 3) einem Staninewert (eine Erklärung des Staninewertes befindet sich in der Fußnote der Auswertung),
- 4) einem quantitativen Teil, in dem die Anzahl der richtigen und positiven Antworten in den einzelnen Dimensionen in Prozentwerten dargestellt wird (außer bei der MFA-Z),
- 5) einem qualitativen Teil, in dem die Ergebnisse aus 4) nochmals in Textform beschrieben werden.

Fragwürdig ist, ob die Übersetzung der Zahlenwerte in Textform den Begriff „Gutachten“ rechtfertigt, wie es im Anhang des Testmanuals geschehen ist. Hierzu fehlt der dem Gutachten inhärente Bezug zu einer konkreten Fragestellung, die Aufnahme von Aspekten der Verhaltensbeobachtung und die eigentliche Stellungnahme, um nur einige definierende Merkmale des Gutachtens zu nennen. Wegen der Besonderheiten der vier Teilverfahren werden Auswertung und Interpretation jeweils separat dargestellt.

### 6.1. MFA-Konfliktsituation (MFA-K)

Der Testverlauf bei diesem Verfahren wird durch die Wahl der jeweiligen Antworten und damit durch die Tp selbst bestimmt. Die Gesamtzahl der bearbeiteten Items ist daher von Person zu Person unterschiedlich. Durch die Addition aller Antworten auf den 40 Seiten des Testheftes kann man aber die Anzahl der möglichen Antwortalternativen ermitteln. Diese beträgt 138. „Jede der Gesprächsalternativen bezieht sich auf eine der beiden Operationalisierungen der Dimension „Soziale Kompetenz““ (S. 42). Diese beiden Dimensionen lauten:

- Bedürfnisse, Interessen anderer erkennen und berücksichtigen,
- Unterstützung anbieten und Konflikte lösen.

Die Handanweisung spricht dagegen von 157 Entscheidungen, wobei unklar bleibt, wie der Differenzbetrag von 19 Antwortalternativen zustande kommt, wenn man die Unabhängigkeit der beiden Dimensionen voraussetzt. 76 dieser möglichen Antworten sollen im Sinne von „Bedürfnisse, Interessen erkennen und

berücksichtigen“ bewertet werden, die restlichen 81 im Sinne der zweiten Dimension. Jede gewählte Alternative kann auf den beiden Dimensionen, die sie mitdefinieren, als geeignet (+) oder ungeeignet (-) eingestuft werden.

Es bleibt unklar, welche der 36 Ausgangssituationen (40 abzüglich der vier Lösungsseiten) im Sinne einer der beiden Dimensionen beantwortet werden sollen. Ebenso ist unklar, welche Alternativen positiv und welche negativ bewertet werden. Aus der Musterlösung im Anhang ist ersichtlich, dass es eine optimale Lösung gibt, wenn in 80 Prozent der Fälle die Bedürfnisse und Interessen anderer erkannt und berücksichtigt werden und in den restlichen 20 Prozent der Fälle dem Mitarbeiter Alternativen und Unterstützung angeboten werden. Hierbei bleibt offen, welche Items ausgewählt werden müssen, um diese Prozentsätze zu erreichen. Weiterhin ist unklar, ob es auch bei anderen Verteilungen der Antworten auf den beiden Dimensionen möglich ist, einen maximalen Punktwert zu erreichen.

Um die zwischen den Tp unterschiedliche Anzahl bearbeiteter Items bei der Auswertung zu berücksichtigen, wird der Anteil an positiven und negativen Antworten im Sinne der oben genannten Dimensionen berechnet und zur Gesamtzahl der Antworten in Beziehung gesetzt. „Die Summe der gewählten positiven Entscheidungen wird in Prozent umgerechnet und bildet den Rohwert“ (S. 42).

Die Ideallösung kommt zustande, wenn man dem Entscheidungsbaum aus Abbildung 34 im Manual folgt (Aa, Ab, Ac, Ad, Ae, Ax). Das Ziel wäre dann nach fünf Entscheidungen erreicht. Das Erreichen der Abschlusseite Ax wäre demnach das Ende des besten Lösungsweges.

In der Handanweisung wird verschwiegen, dass es noch kürzere Wege geben kann, die zu der besten Zielseite Ax führen. Beispielsweise ist es möglich, dass nach nur vier Entscheidungen der Konflikt gelöst wird. Offenbar reicht aber in diesem Fall die Anzahl der gewählten Alternativen für eine quantitative Auswertung nicht aus, denn es wird nur eine qualitative Auswertung vorgenommen, dennoch wird ein Staninewert angegeben. Weiterhin gibt es Lösungswege, die mit nur drei Entscheidungen zu einer der anderen vier Zielseiten führen und damit das Programm beenden. In einem solchen Fall gibt das Programm folgende Fehlermeldung aus: „Die Auswertung kann nicht vorgenommen werden.“ Ein Zugriff auf die Daten der Tp ist dann nicht mehr möglich!

Unter welchen Umständen eine vollständige Auswertung, eine nur qualitative oder gar keine Auswertung stattfindet, bleibt den Rezensenten verborgen. Jedenfalls ist es ärgerlich, wenn einerseits Lösungen möglich sind, die nicht ausge-

wertet werden können und auf deren Rohdaten man nicht mehr zugreifen kann, andererseits dafür eine Einheit im Wert von 25,- DM vom Abrechnungsdongel abgebucht wird.

Bei der Auswertung des MFA-K zeigten sich neben den eben dargestellten Problemen unter anderem folgende Besonderheiten:

- a) Folgt man dem idealen Lösungsweg, wie er in der Abbildung 34 der Handanweisung beschrieben ist, mit Ausnahme des letzten Schritts, so erstellt das Programm keine Auswertung mit der Begründung, dass zu wenige Entscheidungen getroffen wurden.
- b) Wird der „ideale“ Lösungsweg (Aa, Ab, Ac, Ad, Ae, Ax) durch Schleifen mit falschen Alternativen ergänzt (z.B. Aa, Af, Aa, Af, Aa, Ab, Ac, Ad, Ae, Ax), bekommt man dennoch in der qualitativen Auswertung bescheinigt: „Frau Y findet schnell eine tragfähige Konfliktlösung. Sie setzt sich mit den anstehenden Konflikten mit ihren Mitarbeitern auseinander und sucht ein klärendes Gespräch. Frau Y erfährt durch offene Fragen an den Mitarbeiter die Hintergründe für sein Verhalten. Auf den Kritikpunkt des Mitarbeiters geht sie ein. Der Konflikt wird erfolgreich angegangen und erfolgreich ohne Umwege gelöst.“ Trotz vorheriger konstruktiver Schleifen wird der Konflikt also als „ohne Umwege gelöst“ bezeichnet.
- c) Wird der Test aus Zeitgründen abgebrochen, so dominiert in der qualitativen Auswertung, trotz des bis zum Abbruch richtigen Lösungsweges, eine negative Schilderung, die bevorzugt den Abbruch kommentiert. Wird beispielsweise der Lösungsweg wie unter (b) eingeschlagen, aber nicht die letzte Lösungsseite (Ax) ausgewählt (Aa, Af, Aa, Af, Aa, Ab, Ac, Ad, Ae, Abbruch), dann heißt es dazu im qualitativen Teil: „Herr X kommt zu keiner abschließenden Lösung des Problems. Ansatzweise wird die Situation erfasst und es werden teilweise angemessene Ansätze zur Klärung gefunden. Diese werden nicht konsequent fortgeführt. Es ist keine klare Linie in der Gesprächsführung erkennbar.“

Die Prozentwerte werden in der qualitativen Auswertung für die beiden Dimensionen getrennt aufgeführt. Wie die Zuordnung dieser Prozentwerte zu den inhaltlichen Schilderungen im qualitativen Teil geregelt ist, bleibt unklar. Als statistische Kennwerte werden für die MFA-K ein Mittelwert von  $M=80,83$  und eine Standardabweichung von  $SD=19,47$  angegeben. Angaben zur Reliabilität der Skala fehlen ganz. Durch Testläufe kann man erschließen, dass sich diese Zahlen ( $M$  und  $SD$ ) auf die Summen der beiden Prozentwerte der positiven Antworten in allen Dimensionen beziehen. Auch die Berechnung der Stanine-Werte legt anscheinend die Summe der Prozentzahlen positiver Antworten zugrunde.

## 6.2. MFA-Soziale Situationen (MFA-5)

Laut Handanweisung soll auch mit diesem Verfahren die soziale Kompetenz erfasst werden. Die beiden Subdimensionen der MFA-K werden jedoch um andere inhaltliche Aspekte ergänzt:

- Bedürfnisse und Interessen anderer berücksichtigen /Informationen offen austauschen,
- Unterstützung und Hilfe anbieten/Konflikte lösen.

In der quantitativen Auswertung werden die vier Aspekte getrennt aufgeführt. Weiterhin heißt es bezüglich der Auswertungskriterien: „Die Lösung des Probanden zeichnet sich durch die Bearbeitungstiefe und den Variantenreichtum der Vorschläge aus“ (S. 43). Auf diese Facette wird in der Auswertung allerdings nicht eingegangen. Der theoretische Hintergrund der Dimensionen sowie die Frage, ob die Dimensionen überhaupt homogen sind, werden nicht diskutiert. Zur Frage, welche Antworten welchen Dimensionen zugeordnet sind, heißt es lediglich: „Die Zuordnung der Antwortalternativen zu den Operationalisierungen wurde von mehreren Ratern unabhängig vorgenommen“ (S. 43).

Auf die 9 Problemfälle sollen 44 Antwortalternativen verteilt sein. Mehrfachantworten sind möglich. „Alle Antworten, die bei den Problemfällen gegeben werden können, sind einem der vier Bereiche zugeordnet“ (S. 30). Die Zuordnung der Antwortalternativen zu den beiden Dimensionen soll Tabelle 6 des Manuals leisten. Die dort angegebene Anzahl der Antwortalternativen von 44 ist jedoch falsch, tatsächlich gibt es 45 Antwortalternativen. Es sind 14 statt 15 negative Antworten möglich. Dies ist jedenfalls die Anzahl der Items, die bei Ankreuzen aller Alternativen unter der Rubrik „Negative Antworten“ erscheinen.

Bei Testdurchläufen fiel weiterhin in Hinblick auf die Auswertung folgendes auf:

- Werden alle 45 Alternativen ausgewählt, so erreicht man in allen vier Dimensionen 100 Prozent der richtigen Antworten. Ebenso werden bei einem solchen Antwortverhalten auch 100 Prozent der negativen Antworten erreicht, mit einer Ausnahme: Für die Dimension „Unterstützung und Hilfe geben“ wird bei den negativen Antworten ein Wert von 0 Prozent angegeben. Hier scheinen also nur positive Antworten möglich, während bei den anderen Dimensionen auch negative Antworten vorkommen. Auf diese Besonderheit wird im Handbuch nicht hingewiesen, vielmehr wird die unipolare Dimension „Unterstützung und Hilfe anbieten“ in der Tabelle 6 mit der bipolaren Dimension „Konflikte lösen“ zusammengefasst.
- Da in die Stanine-Berechnung nur die positiven Werte eingehen, ist es für die Tp günstig, wenn sie möglichst viele Items ankreuzt. Es wundert daher

nicht, dass die qualitative Auswertung beim Ankreuzen aller Items extrem positiv ausfällt. Bei einem solchen Ankreuzverhalten werden die 14 negativen Antworten gegenüber der positiven Darstellung nur in einem Satz zusammengefasst: „Auffällig ist, dass schwerwiegende Fehlentscheidungen getroffen werden.“

Die bedeutsamsten Auswertungsprobleme dürften sich bei der Übertragung der PuB-Version in die Computerversion ergeben. Die Probleme entstehen (a) durch die Schwierigkeit, die freien Antworten der PuB-Version in die vorgegebenen Antwortalternativen zu überführen, und (b) durch softwarebedingte Mängel im Auswertungsmodus. Auf beide Problembereiche soll näher eingegangen werden.

**(a) Übertragung der freien Antworten aus der PuB-Version.** Das Problem der Übertragung der freien Antworten in die vorgegebenen Alternativen kann beispielhaft an Fall 7 der MFA-S verdeutlicht werden, ließe sich aber auch an einem der anderen Fälle demonstrieren. Der Fall:

„Als Abteilungsleiter haben Sie zu wichtigen Verhandlungen mit dem Personalrat Ihren Mitarbeiter und Stellvertreter, Herrn Mommsen, mitgenommen. Sie sind jedoch sehr erstaunt, als Ihr Mitarbeiter jede auftretende Gesprächspause benutzt, sich selbst ins Geschehen einzuschalten und dabei mit taktisch ‚kritischen‘ Informationen und nicht immer intelligenten Vorschlägen nur so um sich wirft. Die Situation wird Ihnen langsam peinlich, und der Gesprächsverlauf droht Ihnen aus dem Ruder zu geraten. Wie verhalten Sie sich?“

Im „Teilnehmermodus“ wählt die Tp selbst eine der folgenden Alternativen aus:

- Bei der nächsten kritischen Bemerkung von Herrn Mommsen unterbrechen Sie bestimmt, damit nicht noch mehr passiert.
- Sie führen unter einem Vorwand eine Gesprächspause herbei.
- Sie erklären Herrn Mommsen unter vier Augen das Problem und machen ihn auf die Wichtigkeit der Verhandlung aufmerksam.
- Sie führen das Gespräch zusammen mit Herrn Mommsen weiter.
- Sie fordern Herrn Mommsen auf, sich an die Gesprächsregeln zu halten und ihm [Anm.: vermutlich „Ihnen“ gemeint] sein Verhalten zu erklären.

Für die Überführung der freien Antworten der PuB-Version in die Computerversion werden im Auswertungsmodus die ausführlichen Antwortmöglichkeiten der Teilnehmerversion kurz zusammengefasst. Die Reihenfolge der Antwortalternativen ist gegenüber der PuB-Version vertauscht. Für den gerade dargestellten Fall 7 muss die freie Antwort folgenden Alternativen zugeordnet werden:

- Gesprächspause herbeiführen.

- Problematik erklären.
- Klare Regelung herbeiführen, Aussage zum erwarteten Verhalten treffen.
- Verhandlung mit Herrn Mommsen weiterführen.
- Vor allen blossstellen?

Hieran schliessen sich zwei Kritikpunkte an:

*Erstens:* Sind die Alternativen im Auswertungsmodus und im Teilnehmermodus wirklich *identisch*? Während einige Itemzuordnungen einleuchtend sind (z.B. (b) = (1)), ist fraglich, ob die Option (e) im Teilnehmermodus („Sie fordern Herrn Mommsen auf, sich an die Gesprächsregeln zu halten und ihm sein Verhalten zu erklären“) inhaltlich der Option (5) im Auswertungsmodus („vor allen blossstellen“) entspricht.

*Zweitens:* Weiterhin könnte man sich in der PuB-Version freie Antworten vorstellen, die *nicht* in den Alternativen abgebildet werden, aber dennoch im Sinne der beiden Dimensionen abbildbar sind. Eine Tp könnte beispielsweise schreiben: „Ich würde versuchen, das Gespräch an mich zu reißen, so dass Herr Mommsen gar nicht mehr dazu kommt, etwas zu sagen. Wenn er dennoch irgendwelche kritischen Einwände hat, würde ich sie überhören, ihn sozusagen mit meinen eigenen Worten überfahren, damit er keine Kommentare dazu geben kann.“

Wie soll diese Antwort nun klassifiziert werden? Einerseits würde sie in die Alternative „Verhandlung mit Herrn Mommsen weiterführen“ passen, andererseits würde dies auch nicht die ganze Antwort abbilden, denn die Tp bemüht sich, nicht die Kontrolle zu verlieren. Ihre Bemühungen lassen sich aber auch nicht in die Kategorie „Vor anderen blossstellen“ oder „Klare Regelung herbeiführen, Aussage zum erwarteten Verhalten treffen“ einordnen. So können auch in anderen Fällen Antworten vorkommen, die schwer zu klassifizieren sind. Bei Testdurchläufen zeigte sich eine nur geringe Übereinstimmung zwischen den Auswertenden.

**(b) Software-bedingte Mängel im Auswertungsmodus.** Neben dem Überführungsproblem bestehen auch Risiken bei der Eingabe der Daten im Auswertungsmodus. Zwei Probleme sind aufgefallen:

- Eine einmal angeklickte Alternative kann, wenn man die Seite bereits verlassen hat und dann mit „Zurück“ die Seite erneut aufruft, nicht mehr verändert werden. Eine Änderung der Eingabe beim Blättern (beispielsweise zur Kontrolle der Eingaben) ist also nicht mehr möglich.
- Weiterhin ist es verwirrend, dass nach dem Klicken auf „Weiter“ die einge-

<sup>7</sup> Die Antworten im Teilnehmermodus und Auswertungsmodus sind wie folgt zugeordnet: (a) = (3), (b) = (1), (c) = (2), (d) = (4), (e) = (5).



gebenen Antworten in die Buchstabenfolge der Teilnehmerversion umcodiert werden. Wenn man bei der Auswertung des Falls 7 die Alternativen A und B auswählt, dann auf „Weiter“ klickt und später wieder zu der Seite zurückkommt, dann sind statt der ursprünglichen Eingaben A und B nun die Umkodierungen A und C aktiviert, was den Antworten im Teilnehmermodus entspricht. Da man im Handbuch nicht über diese Umkodierung informiert wird, ist es verwirrend, wenn bei der Kontrolle der Eingabe plötzlich automatisch veränderte Angaben auftauchen.

### 6.3 MFA-Führungssituationen (MFA-F)

Bei den MFA-Führungssituationen werden 12 Problemfälle vorgegeben, bei denen es darum geht, in Abhängigkeit vom vorliegenden Problem ein mitarbeiter- oder aufgabenbezogenes Verhalten auszuwählen. Abbildung 33 der Handanweisung stellt die beiden dichotomen Anforderungen in einer Vierfeldertafel dar, aus der dann vier „Facetten des Führungsverhaltens“ abgeleitet werden: (1) *Fordern und Fördern*: stark aufgaben- und mitarbeiterbezogen; (2) *Partizipation*: wenig aufgaben- und stark mitarbeiterbezogen; (3) *Kontrolle*: stark aufgaben- und wenig mitarbeiterbezogen; (4) *Delegation*: wenig aufgaben- und wenig mitarbeiterbezogen.

Bei den 12 Problemfällen stehen jeweils vier Alternativen zur Auswahl, Mehrfachantworten sind nicht möglich. Die Testpersonen müssen also zwölfmal eine der vier Alternativen auswählen. Dabei ist jeweils eine Alternative die optimale Lösung, die mit zwei Punkten bewertet wird. Weiterhin gibt es eine „teilweise angemessene“ Antwort, die mit einem Punkt bewertet wird. Die restlichen zwei Alternativen sind unangemessen und erreichen keinen Punkt. Maximal sind so über die 12 Problemfälle hinweg 24 Punkte möglich. Die 12 Problemfälle sind vier Anforderungsdimensionen zugeordnet. Jeweils drei Fälle verlangen eine Antwort im Sinne von (1) „Partizipation“, (2) „Kontrolle“, (3) „Delegation“ beziehungsweise im Sinne von (4) „Fordern und Fördern“. Für die Rohwerte wird ein Mittelwert von  $M=13,82$  und eine Standardabweichung von  $SD=2,90$  angegeben. Kritisch ist auch hierbei, dass Anwendende wie Testleitende über die Zuordnung der Items zu den Dimensionen im Unklaren gelassen werden.

### 6.4 MFA-Zeitmanagement (MFA-Z)

Das MFA-Zeitmanagement unterscheidet sich von den anderen MFA-Verfahren dadurch, dass hier nur eine Dimension erfasst wird, die als Zeitmanagement be-

schrieben wird. Daher fehlt in der quantitativen Auswertung auch eine differenzierte Darstellung des Lösungsverhaltens. Über eine Abstandsmatrix (siehe weiter unten) werden die getroffenen Entscheidungen eindimensional bewertet. Der dabei entstehende Rohwert soll Aussagen über die „Planungseffizienz und das Strukturierungsvermögen unter Zeitdruck“ (S. 7) erlauben.

Erste Auswertungsprobleme werden bei der Überführung der Antworten der PuB-Version in den Auswertungsmodus deutlich: Die Computerversion berechnet immer den kürzesten Weg zwischen zwei Zielen. Falsch berechnete Wegzeiten in der PuB-Version können im Auswertungsmodus nicht berücksichtigt werden. Beispielsweise könnte die Tp für die Strecke zwischen Bahnhof und Geschäftsstelle einen Weg über die Buchhandlung und das Reisebüro nehmen, was vermutlich häufig passieren wird, da dies der optisch kürzeste Weg ist, der aber 31 Minuten dauert. Der Computer wählt automatisch die zeitlich kürzeste Strecke (über die Bank) als Wegstrecke aus; diese Strecke ist optisch länger, aber die benötigte Zeit beträgt hier nur 26 Minuten. Ein solcher Rechenfehler in der PuB-Version verändert natürlich auch die Ausgangssituation für die Berechnung der Folgetermine, die dann aber unter Umständen wieder die Kriterien der Instruktion (Einhaltung der Termine, „effektive Erledigung“) erfüllen könnten. Leider können diese Termine dann nicht mit den neuen Zeiten eingegeben und ausgewertet werden.

Zur Auswertungslogik: Die in der Instruktion vorgegebenen Restriktionen (Abschluss um 12.00 Uhr, verschiedene andere fixe Termine) schränken die Anzahl der möglichen Lösungswege ein. Unter den möglichen Lösungen, die unter Beachtung aller Restriktionen die Terminvorgabe 12.00 Uhr erfüllen, wird eine Musterlösung als Auswertungsmaßstab festgesetzt, denn „ferner kann gezeigt werden, dass es nur eine Lösung gibt, die Besorgungen in minimaler Zeit zu erledigen“ (S. 44). Dieser vermeintlich beste Lösungsweg wird in Abbildung 35 des Manuals dargestellt.

Gibt man die dort dargestellte Sequenz im Auswertungsmodus ein, so fällt zunächst auf, dass mehr als die Hälfte der dort angegebenen Zeiten nicht richtig berechnet wurden. Teilweise werden falsche Wegezeiten zugrunde gelegt und/oder die Ankunftszeiten stimmen nicht mit den Computerberechnungen überein. Legt man aber als Berechnungsgrundlage den Stadtplan oder die Angaben des Computers zugrunde, dann wird klar, dass dies nicht die ideale Lösung sein kann, denn die Tp würde demnach erst um 10:01 in der Geschäftsstelle ankommen, obwohl laut Instruktion 10:00 Uhr gefordert war.

Die Performanz der Tp wird ausgehend von einer Musterlösung durch eine Abstandsmatrix bewertet. Diese sieht folgendes vor: Für die Auswahl einer Anlauf-

stelle, die dem Rangplatz dieser optimalen Sequenz entspricht, erhält die Tp die volle Punktzahl. Beispielsweise ist der Ort X<sup>4</sup> in der Musterlösung die erste Anlaufstelle. Wird dieser auch von der Tp als erste Stelle gewählt, dann erhält sie die volle Punktzahl (3 Punkte), wird X als zweite Anlaufstelle gewählt, erhält sie zwei Punkte. Wird der Ort X als dritte Anlaufstelle gewählt, erhält sie einen Punkt und wenn X an vierter oder späterer Stelle gewählt wird, gibt es keinen Punkt. Maximal sind so 33 Punkte (11 Entscheidungen mit maximal 3 Punkten) möglich. Als statistische Kennwerte werden ein Mittelwert von  $M=19,42$  und eine Standardabweichung von  $SD=7,09$  angegeben. Da die Auswahl eines Ortes den weiteren Verlauf des Verfahrens mitbestimmt, kann nicht von einer stochastischen Unabhängigkeit der Items ausgegangen werden. Aus diesem Blickwinkel scheint die Verwendung einer Auswertungsmatrix befremdlich. Hier wäre es sicherlich günstiger gewesen, ein Globalmaß für die Lösungsgüte zu verwenden, z.B. die Anzahl der erreichten Ziele oder die Lösungszeit.

Problematisch ist weiterhin, dass die in der Abbildung 36 dargestellte Abstandsmatrix nicht der tatsächlichen Berechnung zugrunde liegt. Gibt man nämlich die angegebene optimale Sequenz - gemäß den Abbildungen 35 und 36 des Manuals - ein, so erreicht man nur 24 der maximal 33 Punkte.

Weder die in den Abbildungen 35 und 36 aufgeführten Lösungen noch die in dem Beispielgutachten genannte optimale Lösung („muster7zma“) sind wirklich die „besten“ Lösungen. Es kann eine Lösung konstruiert werden, die in kürzerer Zeit unter Einhaltung aller Restriktionen dazu führt, dass die Geschäftsstelle 3 Minuten früher erreicht wird (Reisebüro, Buchhandlung, Bank, Werkstatt, Krankenhaus, Geschäftsstelle, EDV-Händler, Werkstatt [ab jetzt verkürzte Wegzeiten], Arzt, Konsulat, Bahnhof, Geschäftsstelle). Für diese zeitbeste Lösung erzielt man nach der Auswertungslogik von Fennekels und D'Zouza (1999) gerade einmal 13 der 33 Punkte, wenn man ihre Abstandsmatrix einer Auswertung per Hand zugrunde legt. In der Computerauswertung werden sogar nur 11 Punkte angegeben. Selbst wenn man die Lösung des Beispielgutachtens („muster7zma“) zugrunde legt, bei der die volle Punktezahl erreicht wird, und daraus eine Abstandsmatrix erstellt, bleibt unklar, wie der Punktwert 11 für die obige Lösung zustande kommt. Vermutlich liegen noch weitere Musterlösungen zugrunde, von denen aus die Berechnungen des Punktwertes vorgenommen werden. Da aber nicht nachvollziehbar ist, wie diese Unstimmigkeiten entstanden sind, ist eine differenzierte Beurteilung des Auswertungsmodus nicht möglich. Die zeitoptimale Lösung er-

<sup>4</sup> Wie bereits zuvor wird hier aus Gründen des Testschutzes die konkrete Darstellung des Lösungsweges durch die abstrakte Bezeichnung verschleiert.

zielt übrigens einen Staninewert von 3 und wird in der qualitativen Auswertung wie folgt kommentiert:

*„Herr X erfasst nur einige wesentliche Aspekte der Problemstellung. Vorgaben werden bei der Problemlösung nicht immer beachtet. Herr X kommt zu einigen folgerichtigen Teillösungen. Seine Zeitplanung ist nicht immer schlüssig. Es gelingt ihm mitunter nicht, Termine zeitgerecht zu planen. Die Gesamtaufgabe wird nicht gelöst.“*

Prinzipiell muss man die Frage stellen, ob die Berechnung mittels Abstandsmatrix überhaupt sinnvoll ist. Dies könnte nur dann bejaht werden, wenn eine Passung zwischen den durch die Instruktion implizierten Kriterien, den Bewertungskriterien der Handanweisung und dem Auswertungsalgorithmus im Sinne der Abbildung 36 (Manual) besteht.

Zunächst ist auffällig, dass keine Übereinstimmung zwischen Instruktion und Handanweisung besteht. Die Instruktion betont, dass es wichtig ist, die Termine einzuhalten und „die Besorgungen effektiv zu erledigen“, was immer damit konkret gemeint ist (es gibt mehrere mögliche Optimierungskriterien). Die Begründung für die Musterlösung in der Handanweisung betont neben dem Einhalten der Termine auch eine zeitoptimale Lösung. Gemeinsam ist beiden, dass die Einhaltung der Termine als wesentlich betrachtet wird. Es bleibt die Frage, ob die Einhaltung der Termine durch die Abstandsmatrix auch entsprechend bewertet wird. Das scheint nicht der Fall zu sein, wie folgende Lösungssequenz zeigt: Werkstatt, EDV-Händler, Reisebüro, Krankenhaus, Buchhandlung, Geschäftsstelle [ab hier weiter wie in der Musterlösung mit Ausnahme einer Abweichung]. Folgt man der Auswertungstabelle, so erzielt man 24 von 33 Punkten. Die Computerauswertung gibt aber nur 17 Punkte an. Das würde einem Staninewert von 4 bzw. 6 entsprechen. Mit dieser Lösung wird aber keiner der Termine eingehalten: Die Person würde zu spät zur Besprechung in der Geschäftsstelle ankommen, der Termin beim Arzt wäre um etwa 20 Minuten verpasst worden und der Kunde, mit dem man am Bahnhof verabredet war, wäre schon längst abgefahren. Umso erstaunlicher die Rückmeldung für die Tp:

*„Herr X. erfasst meistens die Problemstellung und das Ziel einer Aufgabe. Er berücksichtigt dabei weitgehend die relevanten Informationen und beachtet die Vorgaben bei der Problemlösung. Herr X. kommt meist zu vielen folgerichtigen und konsequenten Lösungen. Seine Zeitplanung ist überwiegend schlüssig. Herr X. zeigt in der Regel ein systematisches Vorgehen. Es wird insgesamt eine angemessene Lösung erreicht.“*

An dieser Auswertung verwundert zweierlei: (1) Die doch recht positiv wirkende qualitative Auswertung, trotz des rechnerisch ungeklärten Staninewertes von 4.

(2) Die Tatsache, dass im „Gutachten“ trotz verpasster Termine von einem systematischen Vorgehen, der Beachtung der Vorgaben und einer angemessenen Lösung gesprochen wird.

Bei Testdurchläufen sind folgende weitere Mängel in der Auswertung aufgefallen:

- Im Teilnahmemodus ist das unmittelbare Abholen des Rollers nicht möglich, da die Reparatur des Rollers 1,5 Stunden dauert. Es folgt der Hinweis: „Der Roller ist noch nicht (sic) fertig.“ Demgegenüber ist es im Auswertungsmodus möglich, das sofortige Abholen des Rollers als Option einzugeben.
- Es können Fälle zustande kommen, in denen ein Ort vor der Terminvorgabe erreicht wird. In der oben dargestellten zeitbesten Lösung wäre man beispielsweise schon um 11:26 Uhr am Bahnhof, um 11:27 soll das Treffen mit dem Kunden stattfinden, der um 11:29 Uhr weiterfährt. Hierbei ist fraglich, ob man tatsächlich bis 11:29 bleiben muss, oder-ob es schon möglich ist, um 11:28 den Bahnhof zu verlassen, denn die Verweildauer von 2 Minuten hat man ja eingehalten und es ging inhaltlich nur um die Übergabe einer Warenprobe. Ebenso stellt sich auch bei anderen Terminen die Frage, ob die im Stadtplan angegebenen Verweildauern (Geschäftsstelle: 30 Minuten, Arzttermin: 1 Minute) zum fixen Termin (10:00 Uhr, bzw. 11:10 Uhr) addiert werden müssen oder ob das frühere Erreichen der Orte eine vorzeitige Abfahrt ermöglicht. Diese Einwände sind aber nur für die PuB-Version von Relevanz, da in der Computerversion die Berechnung automatisch erfolgt.
- Bei einem Punktwert von 14 wird ein Staninewert von 3 angegeben. Die Berechnung mit den angegebenen Werten  $((14-19,42/7,09)+5)$  ergibt aber einen Wert von 4.2. Diese Verzerrung bleibt ungeklärt.

### 6.5 Berücksichtigung nicht-bearbeiteter Items

In der Handanweisung heißt es zwar: „In der Auswertung wird berücksichtigt, dass das Verfahren nicht vollständig bearbeitet wurde“ (S. 15), dies stimmt aber nicht ganz:

- Eine unvollständige Bearbeitung der MFA-K führt entweder zu keiner Auswertung oder im quantitativen Teil zu dem Hinweis, dass nur einige wenige Items ausgewählt wurden und so keine Berechnung des Staninewertes möglich ist.
- Bei der MFA-S erfolgt ein entsprechender Hinweis, wenn weniger als fünf von neun Items bearbeitet wurden.
- Bei der MFA-F wird bei drei oder weniger bearbeiteten Fällen (12 insgesamt) in der Auswertung darauf hingewiesen, dass „weniger als 50% der Aufgaben“ bearbeitet wurden. Zusätzlich wird in der qualitativen Auswertung er-

wähnt, dass sich der Staninewert auf die Bearbeitung aller Aufgaben bezieht. Werden dagegen vier (das entspricht 25 Prozent der Aufgaben) oder mehr Aufgaben gelöst, so fehlt in der Auswertung der Hinweis auf die unvollständige Bearbeitung. Der unkommentierte Staninewert wird in die Auswertung aufgenommen und in der inhaltlichen Auswertung wird die Anzahl der nicht-bearbeiteten Aufgaben nicht erwähnt.

- Bei der MFA-Z wird ein Hinweis aufgenommen, wenn weniger als fünf der elf Termine ausgewählt wurden.

Mit Ausnahme der MFA-S wird in keinem der anderen Verfahren der Rohwert zur Anzahl der bearbeiteten Items in Beziehung gesetzt. Ein Hinweis, dass im Falle eines Abbruchs aus Zeitgründen der Staninewert unter Vorbehalt zu interpretieren ist, fehlt im Testmanual wie auch in den exemplarisch abgedruckten „Gutachten“.

## 7 GÜTEKRITERIEN

Die MFA weisen hinsichtlich Inhalt und Konstruktion starke Ähnlichkeiten zu Übungen im Rahmen von Assessment-Centern auf. Das Assessment-Center wird aber aufgrund seiner Verhaltensnähe und seiner situationsspezifischen Elemente den Ansprüchen der klassischen Testtheorie nicht in vollem Umfang gerecht. Die Vorhersageleistung einzelner Verfahren spielt hierbei eine größere Rolle als die Bestimmung von Itemkennwerten und Reliabilitäten. Es verwundert daher nicht, dass auch im Testmanual der MFA nur sehr spärlich Angaben zu den Gütekriterien zu finden sind. Gravierend ist besonders, dass die Gütekriterien trotz der starken Unterschiede zwischen PuB- und PC-Version nicht für beide Versionen getrennt erhoben wurden.

Nachfolgend gehen wir genauer auf die für die Beurteilung von Testverfahren zentralen Kriterien der Objektivität, Reliabilität und Validität ein.

### 7.1 Objektivität

Durch die standardisierte Vorgabe der Instruktion, die automatische Rohwertermittlung und den identischen Algorithmus der quantitativen Auswertung ist in der PC-Version die zu fordernde Durchführungs- und Auswertungsobjektivität gegeben. Im Hinblick auf die Durchführungsobjektivität ist kritisch anzumerken, dass es zwischen Computer- und PuB-Version doch recht große Unterschiede gibt, die sich auf die Instruktion und das Handling des Testmaterials beziehen.

Die PuB-Version verlangt die Transkription in die Computerversion, was bei den Verfahren MFA-K und MFA-F leicht gelingt. Die Übertragung der freien Antworten der PuB-Version der MFA-S in eine der zwei bis sieben Antwortmöglichkeiten der Computerversion ist hingegen problematisch. Ebenso schwierig scheint die Übertragung beim MFA-Z, da die Rechenfehler aus der PuB-Version nicht übertragen werden können. Diese beiden Einschränkungen lassen an der Auswertungsobjektivität der PuB-Version zweifeln. Im Testhandbuch fehlen Angaben zur Interrater-Übereinstimmung bei diesen beiden Verfahren.

Die automatisierte Übersetzung der Rohwerte in eine qualitative Auswertung ist aus inhaltlicher Sicht problematisch, sichert aber die Interpretationsobjektivität.

## 7.2 Reliabilität

Nur zu drei der vier Verfahren werden die Reliabilitäten in Form von Cronbachs Alpha und Split-half-Reliabilitäten angegeben (siehe Tabelle 1). Die Berechnungen basieren auf einer Stichprobengröße von  $N=181$  Personen.

**Tabelle 1:** Reliabilitäten der einzelnen MFA-Teilverfahren

Verfahren	Cronbachs-Alpha	Split-half
MFA-F	.61	.63
MFA-S	.79	.84
MFA-Z	.73	.69
MFA-K	keine Angabe	keine Angabe

Die Kennwerte bestätigen den Verfahren nur teilweise eine gute Konsistenz. Die niedrigen Werte bei der MFA-F, die eine Auswertung auf Individualbasis eigentlich verbieten (vgl. Lienert, 1969, S. 309), werden im Manual als „zufriedenstellend“ (S. 41) bezeichnet. Da die Testpersonen bei den MFA-K eine unterschiedliche Anzahl von Items bearbeiten können, ist es hier nicht möglich, entsprechende Parameter anzugeben. Hier wäre sicherlich ein Rasch-Skalierter Subtest sinnvoller gewesen. Bedauerlicherweise fehlen auch Angaben zur - gerade aus dem Blickwinkel der Personalentwicklung bedeutsamen - Retestrelia-

## 7.3 Validität

Zur Bestimmung der Validität wurden die MFA zusammen mit anderen Auswahlverfahren im Rahmen von Assessment-Centern durchgeführt. Hier werden durchgängig Stichprobengrößen von  $N=132$  genannt; dabei bleibt offen, ob es sich um dieselbe Stichprobe handelt, die zur Normierung herangezogen wurde. Abbildung 37 der Handanweisung gibt Auskunft darüber, mit welchen Verfahren vergleichbare Dimensionen erhoben wurden:

MFA-Z misst demnach das logisch konzeptionelle Denken und Handeln. Dieses wurde auch gemessen mit einer Kurzpräsentation, einem inszenierten Gespräch mit Mitarbeitenden und dem Postkorb-Verfahren. MFA-K und MFA-S sollen die soziale Kompetenz erheben. Dies wurde auch gefordert in dem Verfahren „Führungsaufgabe“ und im Gespräch mit den Mitarbeitenden. MFA-F misst das Führungsverhalten ebenso wie es die Gruppendiskussion, das Gespräch mit den Mitarbeitenden, das Interview und das Postkorb-Verfahren tun.

Leider versäumen es die Autoren, alle Korrelationen zwischen AC-Verfahren und den MFA sowie die Interkorrelation der MFA-Verfahren anzugeben, so dass keine Aussagen zur diskriminanten Validität möglich sind. Beispielsweise werden in Tabelle 10 des Manuals die Ergebnisse der MFA-F nur mit den Ergebnissen des Interviews korreliert und nicht mit den anderen Verfahren, die ebenfalls das Führungsverhalten messen sollen (nämlich laut Abbildung 37: Gruppendiskussion, „Mitarbeitergespräch“ und Postkorb-Verfahren).

Tabelle 9 des Manuals korreliert als Maß für die Übereinstimmungsvalidität die MFA-Verfahren mit den Anforderungen im AC, ohne zu verdeutlichen, wie die Kriteriumsmaße operationalisiert wurden. Es zeigt sich, dass alle Korrelationen zwischen den MFA-Verfahren und den Anforderungen signifikant sind (sie schwanken zwischen .36 und .58). Allerdings wird weder in Tabelle 9 noch in Tabelle 10 des Manuals die Stichprobengröße angegeben, was die Aussagekraft der Signifikanzen einschränkt.

Die fehlende Signifikanz zwischen MFA-F und Führungsverhalten ( $r=.22$ ) wird damit begründet, „dass im Assessment-Center das Führungsverhalten primär das aktive Steuern von Gruppen umfasst. Die MFA-F misst hingegen zusätzlich den angemessenen Einsatz von Führungstechniken“ (S. 47).

Als weitere Belege für die gute interne Validität wird eine Einwegvarianzanalyse aufgeführt, bei der die Potentialaussage als gestufte unabhängige Variable und die Ergebnisse in den jeweiligen Verfahren als abhängige Variable fungieren. Zur Ver-

gleichbarkeit der vier Verfahren werden die Ergebnisse der MFA auf einer gemeinsamen Skala (0-4) abgebildet. Die graphische Darstellung (Abb. 38) zeigt, dass die Verfahren gut zwischen geeigneten, bedingt geeigneten und nicht geeigneten Personen trennen. Ungeklärt bleibt aber, wieso beim MFA-F die Ergebnisse den Erwartungen widersprechen und welche statistischen Werte (F-Werte) der Varianzanalyse und dem Scheffé-Test zugrunde liegen. Weiterhin hätte man sich in der Abbildung 38 die Angabe von Streuungen gewünscht, um Überlagerungen der Bewertungsverteilung zwischen den Gruppen der geeigneten, bedingt geeigneten oder nicht geeigneten Personen erkennen zu können. Gleiches gilt auch für die Abbildungen 39 und 40 im Manual.

Zur Bestimmung der *prognostischen Validität* wurde die weitere Karriereentwicklung ehemaliger Assessment-Center-Teilnehmerinnen und -Teilnehmer mit den erzielten Leistungen in den MFA-Verfahren verglichen. „Der berufliche Erfolg wurde als Umstufung von einer Hierarchiestufe in die nächsthöhere bzw. nächsthöheren operationalisiert“ (S. 48). Die Auswertung zeigt, dass Personen, die höher eingestuft wurden, in allen vier Verfahren statistisch bedeutsam bessere Ergebnisse erzielten. Es wird jedoch nichts über den kontrollierten Zeitraum mitgeteilt. Da unklar ist, welche Auswirkung ein MFA-Ergebnis für die Mitarbeiterin oder den Mitarbeiter hat - denkbar wäre eine aus der MFA selbst abgeleitete Höherplatzierung im Unternehmen -, ist die Aussagekraft dieser Analyse erheblich reduziert.

Erwähnenswert ist, dass die Entwicklung der MFA organisationspezifisch vorgenommen wurde. Die Mitarbeiterinnen und Mitarbeiter des kooperierenden Unternehmens bestimmten, welche Verhaltensweisen beispielsweise gutes Führungsverhalten und soziale Kompetenz ausmachen. Hieraus wurden die entsprechenden Items abgeleitet. Wenn zur Bestätigung der prognostischen Validität die Höherstufung im selben Unternehmen angeführt wird, spricht das zunächst nur für die Brauchbarkeit des Verfahrens für das jeweilige Unternehmen. Ob die dabei erhobenen Facetten auch in anderen Unternehmen und damit in anderen Organisationskulturen bedeutsam sind, ist fraglich. Es muss auch die Frage erlaubt sein, inwieweit die drei Dimensionen des MFA überhaupt erschöpfend die bedeutsamen Führungseigenschaften zukünftiger Vorgesetzter erfassen können.

Anzumerken ist weiterhin, dass die Kriteriumsvaliditäten (prognostische und Übereinstimmungsvalidität) nicht zwingend dafür sprechen, dass die angenommenen Konstrukte valide erfasst wurden. Die hohe Kriteriumsvalidität könnte auch durch einen Reaktivitätseffekt bedingt sein, d.h. beide Methoden messen eher die Fähigkeit der Personen, die Anforderungsdimensionen zu erkennen. Der Einfluss eines solchen Faktors der „sozialen Urteilskompetenz“ bei intransparenten Verfahren (und dazu zählen sicherlich auch die MFA) ist mehrfach bestätigt

worden (Kleinmann, 1993, 1997; Sackett & Dreher, 1982).

#### 7.4 Normierung

Über die zugrunde liegende Normierungsstichprobe wird keine Angabe gemacht. Für die Bestimmung der Mittelwerte, Standardabweichungen, Cronbachs Alpha und der Split-half-Reliabilität wird eine Stichprobengröße von  $N=181$  genannt. Bei den Angaben zur internen und prognostischen Validität wird eine Stichprobengröße von  $N=132$  genannt. Es fehlen Angaben zum Alter, zur Hierarchiestufe und zu anderen für die Personalentwicklung bedeutsamen Variablen der Normstichprobe.

Wie schon mehrfach beschrieben, kann nicht von einer Gleichheit der beiden Testversionen (Computer und PuB) ausgegangen werden. Umso erstaunlicher ist, dass im Testmanual nicht explizit mitgeteilt wird, mit welcher Version die Normierung vorgenommen wurde. Bei der Schilderung des Zeitlimits der PC-Version heißt es: „Alle Normierungen der MFA wurden mit diesen Zeitvorgaben vorgenommen“ (S. 16). Die dabei genannten Zeiten entsprechen den Zeitlimits der Computerversion, was nahe legt, dass die Normierung mit der Computerversion vorgenommen wurde. Andererseits werden drei Seiten vorher die Zeiten der PuB-Version geschildert. Dort heißt es dann auch: „Die Normen der MFA-Verfahren sind auf diese Bearbeitungszeiten abgestimmt“ (S.13). Wegen der großen Unterschiede zwischen beiden Versionen wäre es sicherlich sinnvoll gewesen, eigenständige Normierungen durchzuführen.

Ungeklärt ist auch die Frage, auf welche Bearbeitungssequenz der vier Teilaufgaben sich die Normierung bezieht. Zwischen den Instruktionen bestehen große Differenzen, was die Transparenz der abverlangten Leistungen betrifft. Dies könnte dazu führen, dass die in einem Verfahren angeregten impliziten Annahmen über die Kriterien auch für die nachfolgenden Verfahren übernommen werden. Die Reihenfolge der Bearbeitung wird also vermutlich das Antwortverhalten der Teilverfahren beeinflussen.

## 8 KRITISCHE ANMERKUNGEN

Unsere abschließenden kritischen Anmerkungen beschäftigen sich mit der Anwendungsfreundlichkeit, der Fairness und der Ökonomie des Verfahrens. Daran schließt sich eine zusammenfassende Bewertung an.

## 8.1 Anwendungsfreundlichkeit

Zunächst zur Benutzerfreundlichkeit für den TI: Die Hinweise zur Installation und zur Bedienung des Testprogramms für den Testleiter sind gelungen. Selbst nicht-computergeübten Personen sollte es mit der Handanweisung gelingen, das Programm zum Laufen zu bringen. Das Testmanual begibt sich inhaltlich auf das Niveau von psychologisch nicht vorgebildeten Personen. Es wird beispielweise die Berechnung eines Stanine-Wertes erläutert oder in Abbildung 32 des Manuals gezeigt, wie eine Normalverteilung aussieht. Das Vorgehen, mit dem man zur Auswertung gelangt, wird für jedes der vier Verfahren - teilweise redundant (Abbildung 5 und Abbildung 13 sind beispielsweise identisch) - umfangreich erläutert. Instruktionen zum Handling des Computerprogramms füllen den größten Teil der Handanweisung.

Die im Handbuch als einfach beschriebene Aufladung des MFA-Decoders (Dongels) durch Eingabe eines neuen Abrechnungs-codes funktionierte bei unseren Testläufen allerdings nicht problemlos. Es war ein neuer Dongel notwendig, um weitere Tests durchführen zu können, was bei laufendem Assessment-Center sehr hinderlich gewesen wäre. Ein Export der Rohdaten oder der Auswertungsdaten in andere Programme (z.B. als ASCII-File) ist nicht möglich. Es besteht allerdings eine eingeschränkte Exportmöglichkeit, „falls Sie gleichzeitig das ebenfalls im Hogrefe-Verlag erschienene Assessment-Center-Verwaltungsprogramm AC-TOOLS nutzen“ (S. 13).

Gegenüber den ausführlichen Erklärungen für den TI sind die Erläuterungen des Computer-Handlings für die Tp vernachlässigt worden. Hierzu einige Beispiele:

- Die Funktion des Scrollens zum Lesen der Instruktionen bei der MFA-F oder der MFA-Z wird ebenso wenig erläutert wie das Anklicken verschiedener Antwortalternativen. Die Anordnung der Klickflächen am unteren Rand des Fensters ist etwas umständlich. Vielleicht wäre es hierbei günstiger gewesen, die Antwortalternativen selbst anklickbar zu machen oder die Klickflächen direkt hinter den Antwortalternativen zu platzieren.
- In der MFA-S und der MFA-F besteht die Möglichkeit, die einmal ausgewählten Alternativen durch einen erneuten Klick wieder rückgängig zu machen und zu blättern. Diese Möglichkeiten werden in der Instruktion aber nicht erläutert.
- Besonders kritisch ist die unzureichende Instruktion bei der Zeitaufgabe: Hier wird nicht auf die hintereinander geschalteten Flächen (Stadtplan, Wegplanung) hingewiesen. Die Prinzipien zum Auswählen des nächsten Ziels, d.h. die Überführung von der „to do“-Liste in die Liste der erledigten Ter-

mine wird der Tp ebensowenig erklärt wie die Option, eine einmal gewählte Reihenfolge zu ändern.

## 8.2 Fairness

Ein fundamentaler Kritikpunkt ist, dass die Tp in den meisten Verfahren darüber im Unklaren gelassen werden, auf was es in den Tests ankommt. Die Transparenz der Bewertungsprozesse und der gestellten Anforderungen ist nicht gegeben. Die Tp kann die geforderten Leistungen nur erraten. Hier ist also einer der wesentlichen Faktoren der sozialen Validität (Schuler & Stehle, 1983) verletzt worden, was auf die Akzeptanz der Verfahren einen negativen Einfluss haben dürfte.

Aus der Sicht des TI ist es - selbst nach der Lektüre des Testmanuals - schwierig, die folgenden Fragen zu beantworten: Was sind die Zielgruppen des Verfahrens? Was ist der Zweck der Messung (Selektion oder Beurteilung)? Was wird überhaupt gemessen?

Bei genauerer Betrachtung der Instruktion, der Ansprüche des Testmanuals und der qualitativen Auswertung fällt auf, dass es deutliche Inkongruenzen und sprachliche Unschärfen gibt. Die geringe theoretische Einbindung verstärkt diesen Eindruck. Die Analyse der MFA-S macht dies nochmals deutlich:

Laut Handanweisung steht die „Sensibilität des Teilnehmers, zwischenmenschliche Probleme wahrzunehmen, und seine Effektivität in der Konfliktbewältigung“ (S.8) im Mittelpunkt der MFA-S. An anderer Stelle ist von der Messung der „Sozialen Kompetenz“ (S. 43) die Rede. Es werden ähnliche Situationen erfasst wie bei der MFA-K, allerdings werden die Dimensionen weiter gefasst: a) „Bedürfnisse, Interessen erkennen und berücksichtigen“ wird ergänzt durch „Informationen offen austauschen“, b) „Unterstützung und Hilfe anbieten“ wird ergänzt durch „Konflikte lösen“. So werden aus zwei plötzlich vier relevante Dimensionen. Die Instruktion der PuB-Version scheint durch den Satz „Bitte versetzen Sie sich in die Lage der Mitarbeiter und Führungskräfte und beschreiben Sie, wie Sie sich verhalten würden“ eher darauf hinzudeuten, dass es neben der Zieldurchsetzung durch die Führungskraft auch um die Beachtung der Bedürfnisse der Mitarbeitenden geht. Eine solche Instruktion fehlt in der Computerversion. Die Auswertung erfolgt für die vier Dimensionen getrennt, obwohl sie in der Tabelle 6 zu zwei Dimensionen zusammengefasst wurden. Die Berechnung des Staninewertes beruht nur auf der Summe aller positiven Antworten. Hier werden die zwei bzw. vier Dimensionen also auf eine Dimension reduziert. Auf die im Handbuch angesprochene Bearbeitungstiefe und den Variantenreichtum der Antworten (S. 43) wird weder in der quantitativen noch der qualitativen Auswertung eingegangen. Der

Erstautor warnt selbst davor, „eine Liste von Anforderungshülsen und redundanten Verhaltenskategorien, die zwar alle die allseits bekannten klangvollen Namen tragen, aber mitnichten trennscharf sind“ (Fennekels, 1996, S. 248) zu verwenden. Dieser selbst gesetzte Anspruch wurde aber nicht eingelöst.

### 8.3 Ökonomie

Zur Klären bleibt, ob das Verfahren in Anbetracht der Anschaffungskosten von 3450,- DM und den Folgekosten in Höhe von 100,- DM pro Person als kostengünstig zu bewerten ist. Die Frage muss natürlich vor dem Hintergrund der Folgen möglicher Fehlentscheidungen, der Entwicklungskosten eigener unternehmensspezifischer Verfahren und dem Aufwand bei der Auswertung eigener AC-Verfahren beurteilt werden. In Anbetracht der hohen Korrelationen zwischen den MFA und anderen AC-Verfahren (Tabelle 10 des Manuals) ist für den Praktiker vor Ort nicht unbedingt einleuchtend, warum die bisherigen Verfahren durch dieses Diagnostikum ersetzt oder ergänzt werden sollten. Abbildung 39 des Manuals zeigt zudem, dass das Postkorb-Verfahren wesentlich deutlicher mit der Karriereentwicklung zusammenhängt. In Anbetracht einer Korrelation zwischen MFA-Z und dem Postkorb-Verfahren von  $r = .65$  bleibt die Frage, was der „diagnostische Mehrwert“ der MFA-Verfahren ist. Wo ist außerdem der theoretische und praktische Fortschritt dieses Verfahrens gegenüber den „hausgemachten“ klassischen AC-Übungen?

### 8.4 Abschließende Anmerkungen

In weiten Teilen des Testmanuals wird deutlich, dass sich die Entwicklung des Verfahrens stark an der Entwicklung von Übungen im Assessment-Center orientiert. Dort werden geringere testtheoretische Ansprüche gestellt, da ein organisationsexterner Einsatz nicht vorgesehen ist. Mit den „Management-Fallstudien“ wurde offensichtlich versucht, derartige Übungen zu einem computergestützten Testverfahren weiterzuentwickeln, das sich an den Kriterien der klassischen Testtheorie orientiert. Der Titel des Verfahrens vermeidet zwar den Ausdruck „Test“, dennoch wird durch die Aufnahme von Normen deutlich, dass versucht wird, die Teilverfahren als allgemeingültiges Diagnostikum zu etablieren. Leider wurden bei der Entwicklung oder Weiterentwicklung die dazu notwendigen Vorarbeiten nicht geleistet (oder sie wurden geleistet und nicht im Testmanual abgedruckt).

Hierzu gehört beispielsweise eine Auseinandersetzung mit der Frage, ob das Verfahren zur Auswahl, Platzierung oder zur Potentialerfassung der Mitarbeiterinnen

und Mitarbeiter verwendet werden soll, darf oder kann. Offen bleibt auch die Frage, warum Persönlichkeitstests (z.B. MFA-K) und Leistungstests (z.B. MFA-Z) ohne gemeinsamen theoretischen Hintergrund in einem Gesamtverfahren integriert werden. Die Auswahl gerade dieser Verfahren scheint willkürlich. Eine theoretische Auseinandersetzung mit den Fragen, welche Dimensionen für zukünftige Führungskräfte bedeutsam sind und wie diese gemessen werden sollen, fehlt. So bleibt auch offen, ob die vier Verfahren den Anspruch haben, alle relevanten Dimensionen zu messen, oder ob sie nur einige Facetten des Führungsverhaltens abdecken und durch weitere Tests ergänzt werden müssen.

Wenn versucht wird, die Übungen als Test zu etablieren, dann muss das Verfahren weiterhin auch den testtheoretischen Ansprüchen genügen. Leider steht es diesbezüglich nicht zum Besten.

Der Anspruch von Fennekels und D'Zouza (1999), ein Verfahren zu entwickeln, das für den Praktiker vor Ort leicht anwendbar ist, ist sicherlich eingelöst worden. Die Auswertung erfolgt automatisch und ermöglicht so eine schnelle Anwendung. Umso problematischer sind die vielen Schwächen, die durch das einfache Handling überdeckt werden. Die eklatanten Mängel hinsichtlich des Auswertungsalgorithmus, die nur spärlichen Angaben zu den Gütekriterien und die theoriearme Zuordnung der Items zu den vermeintlich relevanten Dimensionen rechtfertigen den Preis des Verfahrens nicht. Man kann nur vermuten, dass mit dem Verfahren auch Auswahlentscheidungen getroffen werden sollen. Gerade unter diesem Aspekt wären strenge Kriterien bei der Entwicklung des Verfahrens und der anschließenden Qualitätsprüfung wichtig gewesen. Diese Forderung wurde offensichtlich nicht erfüllt.

Es war mühsam, durch aufwendige Testläufe einen kleinen Einblick in die Hintergründe des Auswertungsmodus zu bekommen. An einigen Stellen merkt man, dass das Verfahren die Anwendenden bewusst darüber im Unklaren lässt, wie die Zuordnung der Items zu den Dimensionen und wie die Reduktion der Rohwerte in die quantitative und qualitative Auswertung zustande kommen. Im Anhang des Manuals werden 8 Beispiels-Gutachten aufgeführt. Zunächst kann man sich über den Nutzen dieser „Gutachten“ Gedanken machen, aber man kann sich auch fragen, warum dort nicht die richtigen und falschen Antworten explizit aufgeführt wurden. Sollte so vielleicht vermieden werden, dass man die vier PuB-Verfahren auch ohne Computer und damit ohne Kosten in Höhe von 100,- DM pro Person selbst auswerten kann? Somit bleibt unsere Kritik im Trend einer weiteren Rezension eines Fennekels'schen Computerverfahrens (PC-Office), bei dem ebenfalls die unzureichende Transparenz bemängelt wurde (Riediger & Rolfs, 1998).

Die Idee einer computergestützten Testung wesentlicher Leistungs- und Persönlichkeitsmerkmale von Mitarbeiterinnen und Mitarbeitern ist zunächst lobenswert. Insgesamt scheinen das Computerprogramm und auch die Handanweisung aber unausgereift. An vielen Stellen hat man den Eindruck, dass das Verfahren in Eile fertiggestellt wurde und ohne ausreichende Prüfung auf den Markt gelangte. Nicht zuletzt im Interesse einer fachlich soliden Diagnostik gerade im Personalbereich, in dem testtheoretisch fundierte Verfahren zunehmend Einzug halten, erscheint uns die baldige Nachbesserung dieses fehlerhaften Produkts unumgänglich. Die Qualitätskontrolle des Verlags sollte möglicherweise durch ein wissenschaftliches Gremium ergänzt werden, um solche Fälle zukünftig bereits vor der Publikation zu identifizieren und im gemeinsamen Interesse von Testautoren und Testanwendern in bessere Bahnen zu lenken.

## 9 LITERATUR

- Fennekels, G.P. (1996). *Evaluationsstudie: 17 Jahre AC in einem Großunternehmen*. In Arbeitskreis Assessment Center e.V. (Hrsg.), *Assessment Center als Instrumente der Personalentwicklung*. Hamburg: Windmühle.
- Fennekels, G.P. & D'Zouza, S. (1999). *Management-Fallstudien (MFA). Handanweisung*. Göttingen: Hogrefe.
- Flanagan, J.C. (1954). *The critical incident technique*. *Psychological Bulletin*, 51, 327-358.
- Kleinmann, M. (1993). *Reaktivität von Assessment-Centern*. In W. Hacker & U. Wetter (Hrsg.), *Arbeits- und Organisationspsychologie 19991 in Dresden* (S. 344-349). Bonn: Deutscher Psychologen Verlag.
- Kleinmann, M. (1997). *Transparenz der Anforderungsdimensionen: Ein Moderator der Konstrukt- und Kriteriumsvalidität des Assessment-Centers*. *Zeitschrift für Arbeits- und Organisationspsychologie*, 41, 171-181.
- Lienert, G.A. (1969). *Testaufbau und Testanalyse*. Dritte, durch einen Anhang über Faktorenanalyse ergänzte Auflage. Weinheim: Beltz.
- Riediger, M. & Rolfs, H. (1998). *Computergestützte Postkorb-Verfahren: Mailbox '90, PC-Office und PC-Postkorb „Seeblick“*. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 43-50.
- Sackett, P.R. & Dreher, G.F. (1982). *Constructs and assessment center dimensions: Some troubling empirical findings*. *Journal of Applied Psychology*, 4, 401-410.
- Schuler, H. & Stehle W. (1983). *Neue Entwicklungen des Assessment-Center-Ansatzes beurteilt unter dem Aspekt der sozialen Validität*. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.