

# **Die Rolle von Beobachtererfahrung bei Beurteilungsprozessen im Assessment Center**

## **Dissertationsschrift**

zur Erlangung des akademischen Grades

Doktor der Philosophie

(Dr. phil.)

vorgelegt dem

Fachbereich Psychologie und Sportwissenschaften

der Johann Wolfgang Goethe-Universität

von

Dipl. Psych. Thomas Randhofer

geboren am 21. Juli 1961

in Aachen

Oktober 2004

## Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b> .....	<b>5</b>
<b>2</b>	<b>Verhaltensbeurteilung</b> .....	<b>10</b>
2.1	Beeinträchtigung der Qualität von Verhaltensbeurteilung im Rahmen von Personalauswahlverfahren: Bewerberinterview und AC.....	10
2.2	Stereotype.....	12
2.3	Beobachtungs-/Beurteilungsfehler .....	13
2.4	Strukturelle Maßnahmen zur Verbesserung von konvergenter und diskriminanter Validität.....	15
2.5	Die Qualifizierung von Beobachtern als Möglichkeit der Qualitätssteigerung.....	18
2.5.1	Verbesserung von konvergenter und diskriminanter Validität durch spezielle Beobachtertrainings.....	19
2.5.1.1	Fehlervermeidungs- und Beobachtungsgenauigkeitsorientiertes Training.....	21
2.5.1.2	Frame-of-Reference Training.....	23
2.6	Die Person "Beobachter".....	25
2.6.1	Psychologen, Experten und Novizen als Beobachter.....	27
2.7	Forschungshypothesen.....	30
<b>3</b>	<b>Methode</b> .....	<b>33</b>
3.1	Die Entwicklung des AC-Systems .....	33
3.1.1	Das Beobachtungsinventar und seine Verhaltensmerkmale.....	34
3.1.2	Die Beobachterbögen und die Beurteilungsskalen.....	35
3.2	Die Rahmenbedingungen.....	36
3.2.1	Auftrag und Zielgruppe.....	36
3.2.2	Der Stellenwert des AC im Rahmen der Auswahlprozedur.....	37
3.3	Das AC-Material.....	38
3.3.1	Zielgruppenspezifisches Beobachtungsinventar und Anforderungsprofil für Vertriebsleiter.....	38
3.3.2	Zielgruppenspezifisches Beobachtungsinventar und Anforderungsprofil für Versicherungsverkäufer.....	39
3.3.3	Die Konstruktion der Übungen.....	41
3.3.4	Beobachtermaterial und Teilnehmermaterial zu den Übungen.....	44
3.3.5	Die Beobachterbögen.....	44
3.4	Die Beobachter.....	45
3.4.1	Die unternehmensinternen Beobachter.....	45
3.4.2	Die AC-Assistenz.....	46
3.4.3	Die Aufgaben von AC-Leiter, AC-Assistenz, Experten und Novizen.....	47
3.5	Die AC-Teilnehmer.....	48

3.6	Die Durchführung des AC.....	48
3.6.1	Die Instruktion der Beobachter.....	48
3.6.2	Die Instruktion der Teilnehmer.....	50
3.6.3	Der Ablauf des eigentlichen AC-Tages.....	51
3.6.3.1	Ablauf der Übungen.....	51
3.6.3.2	Die erste Beobachterkonferenz.....	54
3.6.3.3	Die Rückmeldegespräche.....	55
3.6.3.4	Die abschließende Beobachterkonferenz.....	55
3.7	Der Weg der Ergebnisfindung in den einzelnen Phase des AC.....	55
3.7.1	Datenerhebung während der Übungen.....	56
3.7.2.	Ergebnisfindung in der ersten Beobachterkonferenz.....	57
3.7.2.1	Rangfolge 1: Berechnung von Mittelwerten.....	57
3.7.2.2	Rangfolge 2: Die analytische und die emotionale Sichtweise.....	57
3.7.2.3	Vergleich von Rangfolge 1 und Rangfolge 2 zur Ergebnisfindung.....	58
3.7.2.4	Ergebnisfindung in der zweiten Beobachterkonferenz.....	59
3.8	Verrechnung der Daten.....	61
3.9	Statistisches Vorgehen.....	64
<b>4</b>	<b>Ergebnisse</b> .....	<b>66</b>
4.1	Mittlere Korrelationen der Beobachtergruppen untereinander über die Übungen hinweg.....	66
4.1.1	Übungsbezogene mittlere Korrelationen zwischen AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter – Novizen.....	66
4.1.2	Übungsbezogene mittlere Korrelationen zwischen AC-Assistenz - Experten und AC-Assistenz – Novizen.....	70
4.1.3	Übungsbezogene mittlere Korrelationen zwischen Experten und Novizen.....	71
4.2	Mittlere Korrelationen der Beobachtergruppen untereinander über die Verhaltensmerkmale hinweg.....	72
4.2.1	Merkmalsbezogene mittlere Korrelationen zwischen dem AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter – Novizen.....	73
4.2.2	Merkmalsbezogene mittlere Korrelationen zwischen AC-Assistenz – Experten und AC-Assistenz – Novizen.....	75
4.2.3	Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen.....	76
4.3	Korrelation der Beobachtergruppen innerhalb der Übungen.....	78
4.4	Eindrucksbildung über die Teilnehmer auf Grund von Rankings.....	79
4.5	Dieselben Forschungshypothesen - eine andere Stichprobe.....	81
4.5.1	Mittlere Korrelationen der Beobachtergruppen untereinander über die Übungen hinweg.....	82
4.5.1.1	Übungsbezogene mittlere Korrelationen zwischen AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter – Novizen.....	82
4.5.1.2	Übungsbezogene mittlere Korrelationen zwischen AC-Assistenz - Experten und AC-Assistenz – Novizen.....	84
4.5.1.3	Übungsbezogene mittlere Korrelationen zwischen Experten und No- vizen.....	85

4.5.2	Mittlere Korrelationen der Beobachtergruppen untereinander über die Verhaltensmerkmale hinweg.....	86
4.5.2.1	Merkmalsbezogene mittlere Korrelationen zwischen dem AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter – Novizen.....	86
4.5.2.2	Merkmalsbezogene mittlere Korrelation zwischen AC-Assistenz – Experten und AC-Assistenz – Novizen.....	87
4.5.2.3	Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen.....	89
4.5.3	Korrelation der Beobachtergruppen innerhalb der Übungen.....	89
4.5.4	Eindrucksbildung über die Teilnehmer auf Grund von Rankings.....	91
<b>5</b>	<b>Diskussion.....</b>	<b>93</b>

**Zusammenfassung**

**Literatur.....**

**Anhang**

## 1 Einführung

Assessment Center (AC) wurden und werden eingesetzt, um Personalentscheidungen zu treffen (vgl. Jeserich et al., 1989). Sie dienen der Analyse von Verhaltenspotenzialen, die Rückschlüsse auf zukünftige berufliche Leistungsfähigkeiten zulassen sollen. Dabei ist der Einsatz von AC nicht unproblematisch: eine Fülle von Schwierigkeiten kommen auf den Anwender zu, der die Gültigkeit und/oder die Genauigkeit seines "Diagnoseinstruments AC" überprüfen möchte (vgl. Höft & Funke, 2001; Kleinmann et al., 1995; Kleinmann, Exler, Kupsch u. Köller, 1997; Kompa 1989; Maukisch, 1989; Neubauer, 1989; Schuler, 1989). Diese Probleme gilt es, im Vorfeld zu erfassen und, wenn möglich, in der AC-Durchführung zu kontrollieren.

Ungeachtet der mit einem AC verbundenen Komplikationen erfreut sich dieses Verfahren einer großen Beliebtheit, wird von nahezu allen namhaften Unternehmen in der Bundesrepublik Deutschland eingesetzt, und seine Überlegenheit gegenüber anderen Verfahren wird seit langem (vgl. Jeserich, 1981) nachhaltig betont.

Unabhängig vom Zweck des AC (Entwicklung von Mitarbeitern oder Auswahl von internen/externen Mitarbeitern) lassen sich AC als Einzel- oder Gruppenverfahren konzipieren. In beiden Varianten werden Teilnehmer berufsnahen Situationen ausgesetzt, in denen sie berufsspezifisch relevante Verhaltensweisen zeigen sollen. Bei einem Einzel-AC wird nur ein Teilnehmer beobachtet und beurteilt, bei einem Gruppen-AC gibt es mehrere Teilnehmer, deren Verhaltensweisen in unterschiedlichen Situationen beurteilt werden.

Wesentliche Merkmale unterscheidet ein AC von anderen Verfahren: im Gegensatz zu Testverfahren, Interviews oder Biographischen Fragebögen werden im AC die Verhaltensweisen von Teilnehmern durch Beobachter systematisch beobachtet, beurteilt und bewertet. Das Ziel ist dabei, aus den Beobachtungsergebnissen gültige, auf den zukünftigen Berufserfolg bezogen relevante Informationen über die beobachtete Person zu erhalten. Dieser Schluss wird von unterschiedlichen Annahmen gestützt:

- a) Man vermutet, dass ein Zusammenhang zwischen dem beobachteten Verhalten und dem Verhalten der Person außerhalb der Beobachtungssituation besteht, und schließt vom beobachteten Verhalten auf psychische Determinanten,

die diesem Verhalten zugrunde liegen sollen (vgl. Michel, 1964). Pawlik (1976) bezeichnete dieses Vorgehen als Statusdiagnostik, die von einem traditionellen Verständnis des Eignungsbegriffes (vgl. Hoyos, 1974; Jäger, 1970) ausgeht und die Stabilität von Personen-, bzw. Persönlichkeitsvariablen betont.

- b) Man nimmt an, dass das beobachtete Verhalten repräsentativ für das zu erkundende Gesamtverhalten sei, wie bspw. Ghiselli (1966) mit seinen Profilvergleichen nahelegte (vgl. auch Maukisch, 1989).

Diese Grundannahmen werden in der Regel vom AC-Konstrukteur vor dem Hintergrund seines Weltbildes, seiner impliziten Persönlichkeitstheorien sowie seiner sozialen, psychologischen und branchenspezifischen Kompetenzen in ein AC-System umgesetzt. Sie finden ihren Niederschlag in den verbalen Formulierungen zur Kategorisierung der zu beobachtenden Verhaltensweisen der AC-Teilnehmer, in der Konstruktion der Beurteilungssysteme und in den Übungsinhalten. Das bedeutet, dass es nicht das AC schlechthin gibt, sondern immer nur ein AC, das von der besonderen Note seines Konstrukteurs geprägt ist .

In der Literatur finden sich viele Forderungen, die als Standards der AC-Technik an ein AC als diagnostisches Instrument, an seine Bestimmungsgrößen und an die mit diesem Instrument arbeitenden Diagnostiker, die Beobachter, gestellt werden.

Um nur einige zu nennen: Jeserich (1980) weist z.B. darauf hin, ....

- dass vor dem Einsatz des AC Anforderungen zu definieren seien, anhand derer die Teilnehmer überprüft werden (vgl. auch Donat & Moser, 1989);
- dass die Übungen realitätsnah erstellt werden sollten;
- dass Beobachtung und Beurteilung voneinander zu trennen sind;
- dass die Urteile der Beobachter miteinander vergleichbar und von Personen und Situationen unabhängig sein sollen;
- dass die Teilnehmer von möglichst vielen Beobachtern beobachtet werden, um Fehleinschätzungen durch einzelne Beobachter vorzubeugen; nach Byham (1977) ist bspw. das effizienteste Verhältnis von Beobachtern zu Teilnehmern 1 : 2.

Diese und andere Standards, wie sie bspw. der Arbeitskreis Assessment Center als

Qualitätskriterien für die Entwicklung und Durchführung von AC (vgl. Arbeitskreis Assessment Center e.V., 1996, sowie International Task Force on Assessment Center Guidelines, 2000) definierte, scheinen notwendig zu sein; denn der "Wildwuchs" auf dem AC-Sektor hat dazu geführt, dass diese Verfahren vor allem bei den Teilnehmern bezüglich ihrer Aussagegültigkeit auf geringe Akzeptanz stoßen (wenn auch ihr Einsatz hingenommen wird).

Die Einhaltung derartiger Standards ist um so bedeutsamer, wenn man bedenkt, dass das AC im Vergleich zu allen anderen eignungsdiagnostischen Verfahren das zeitlich und finanziell aufwändigste Verfahren ist. Insofern ist dieses Verfahren fortwährenden Rechtfertigungszwängen unterworfen – sowohl den unternehmensinternen Auftraggebern gegenüber hinsichtlich seiner Erfolgsprognose, wie auch unternehmensinternen Beobachtern gegenüber hinsichtlich seiner Akzeptanz.

Folgende Kostenrechnung für ein Personalauswahl-AC mag dies verdeutlichen:

Eine Führungskraft im AT-Bereich, mittleres Management, verdient im Jahr geschätzte 60.000 Euro, das sind pro Arbeitstag bei 220 Arbeitstagen pro Jahr ca. 275 Euro. Für ein AC werden ca. 8 Beobachter benötigt.

Bei einer Durchführung von einem externen Beratungsunternehmen verlangt dieses pro Arbeitstag und Berater ca. 1250 Euro. Normalerweise wird das AC von 2 Personen moderiert/durchgeführt.

Ein Hotel mittlerer Kategorie hat eine Vollpensionspauschale von ca. 75 Euro pro Person, zuzüglich einer Pauschale für Raummiete, Seminarge-tränke und die Nutzung von Seminarmedien (Overheadprojektor, Flip - Chart etc.) von ca. 200 Euro pro Tag.

Für die bundesweit anreisenden Beobachter und Teilnehmer fallen neben einer Spesenpauschale von 23 Euro/Tag noch Reisekosten in Höhe von durchschnittlich 150 Euro/ Person an.

Für ein zweitägiges AC mit acht Beobachtern, acht Teilnehmern und zwei Moderatoren entstehen dabei folgende Kosten (dabei bewegen sich die unten genannten Zahlen eher an der unteren Grenze; hinzu kommen dabei noch die nicht messbaren Kosten, die entstehen, wenn der auf den Arbeitsplatz bezogene Produktivitätsverlust der Beobachter

einbezogen wird (sogenannte Opportunitätskosten) etc..

8 Beobachter á 2 Tage á 275 Euro	-> 4400 Euro
8 * 2 Tage Spesenpauschale (23 Euro)	-> 368 Euro
16 * Reisekosten a 150 Euro	-> 2400 Euro
2 * 2 Manntage des Beratungsunternehmens	-> 5000 Euro
Reisekosten / Spesen des Beratungsunternehmens	-> 500 Euro
Hotelkosten für 18 Personen	-> 1350 Euro
Hotelpauschale (Raummiete etc.)	-> 400 Euro
<b>Gesamtsumme</b>	<b>-&gt; 14418 Euro</b>

Entsprechend hoch sind die Erwartungen, die Unternehmen mit dem Einsatz von AC verbinden. Und es ist offenkundig, dass man angesichts dieser Zahlen größtmögliche Sorgfalt bei der Konstruktion und Durchführung eines AC walten lassen sollte, um möglichst "realistische", "genaue" Beurteilungen der Teilnehmer zu erhalten - und um letztendlich die "richtigen" Teilnehmer zu identifizieren.

Der weitest möglichen Vermeidung von Beobachtungs- und Beurteilungsfehlern kommt daher im AC eine große Bedeutung zu. Es verwundert nicht, dass mehr noch als das AC-System selbst der Prozess der Beobachtung von Verhaltensweisen, die diagnostische Urteilsbildung und die Analyse der Einflüsse, denen ein AC unterworfen ist, in der Vergangenheit Gegenstand vieler Untersuchungen war.

Viele Bemühungen konzentrierten sich darauf, die möglicherweise größte Fehlerquelle in einem AC, den Beobachter, durch spezielle Trainings zu einem Beobachtungs-Experten werden zu lassen, der gegenüber Störeinflüssen "immun" wird, um mit einer Verbesserung des Beobachterverhaltens die Qualität der AC-Ergebnisse zu steigern. Die Durchführung eines Beobachtertrainings wird auch von unterschiedlichen Quellen als eines der Qualitätskriterien eines AC betrachtet (vgl. Arbeitskreis Assessment Center e.V, 1996, sowie die International Task Force on Assessment Center Guidelines, 2000). Das bedeutet, dass zu den o.a. Kosten noch die Kosten für ein Beobachtertraining dazugerechnet werden müssen. Sie belaufen sich auf ca. 4700 Euro je Tag für acht Beobachter und zwei Moderatoren, ohne Reisekosten, Hotelkosten und Spesen.

Doch scheinen die Befunde zum Erfolg von Beobachtertrainings nicht immer eindeutig für ein derartiges Training zu sprechen. Es gibt sowohl Stimmen, die für den positiven Effekt von Beobachtertrainings sprechen, wie auch Stimmen, die gegenteiliger Auffas-

sung sind (vgl. Kap. 2 in dieser Arbeit). Da Kosten- und Zeitaufwand die Akzeptanz der AC-Durchführung vor allem bei den internen Führungskräften, die in der Regel als Beobachter an einem derartigen Verfahren teilnehmen sollen, schmälern, ist es aus ökonomischen Gründen interessant der Frage nachzugehen, ob man nicht möglicherweise auf ein Beobachtertraining verzichten kann, wenn es nicht zu einer gravierenden Qualitätssteigerung der AC-Ergebnisse führt.

Die vorliegende Arbeit möchte diese Frage beantworten, indem sie die Übereinstimmung von unterschiedlich qualifizierten Beobachtern im Rahmen von "Real-Life-AC" untersucht. Dabei handelt es sich um AC, bei denen betriebliche Mitarbeiter eines Versicherungsunternehmens als Beobachter wirkten, und Teilnehmer, die sich der AC-Situation stellten, um einen Arbeitsplatz in diesem Unternehmen zu erhalten. Durch den Vergleich der Urteile von Experten-Beobachtern und Novizen-Beobachtern soll herausgefunden werden, ob...

- Experten-Beobachter und Novizen-Beobachter sich in ihrer Beurteilungsqualität voneinander unterscheiden;
- Novizen-Beobachter (falls der o.a. Punkt zu Gunsten der Experten-Beobachter beantwortet werden muss) ohne großen Aufwand auf das selbe Leistungsniveau gehoben werden können wie Experten-Beobachter
- Experten-Beobachter und Novizen-Beobachter hinsichtlich ihrer Einschätzung der Teilnehmer in einem Personalauswahl-AC zu übereinstimmenden Ergebnissen kommen können, obwohl sie unterschiedlich hohe Personalauswahlkompetenz haben.

Nach einer kurzen Darstellung der Erkenntnisse zu verschiedenen Arten von Beobachtertrainings wird dargestellt, wie für ein Finanzdienstleistungsunternehmen ein AC zur Personalauswahl konzipiert wurde, das sich an den Normen und Werten des Unternehmens orientiert. Im Anschluss daran werden die erarbeiteten Forschungshypothesen an zwei Datensätzen überprüft: einmal an einem Datensatz, bei dem die Beobachter in den Mehrpersonenübungen nicht alle Teilnehmer beobachtet haben, und dann an einem Datensatz, bei dem alle Beobachter alle Teilnehmer in jeder Übung beobachtet haben. Die Ergebnisse werden dargestellt und anschließend diskutiert.

## 2 Verhaltensbeurteilung

### 2.1 Beeinträchtigung der Qualität von Verhaltensbeurteilung im Rahmen von Personalauswahlverfahren: Bewerberinterview und AC

Traditionelle Verfahren der Personalauswahl, wie z.B. das Bewerberinterview, unterliegen in sehr hohem Ausmaß dem Phänomen des "verzerrten Blickwinkels", das man nachweisen kann, wenn man als Validitätskriterium den späteren Berufserfolg betrachtet: Hunter und Hunter (1984) ermittelten z.B. Validitäten von .09 für das nicht standardisierte Interview als Prädiktor für den künftigen Berufserfolg.

Das erscheint nicht weiter verwunderlich, wenn man bedenkt, dass der Interviewer der Problematik der teilnehmenden Beobachtung ausgesetzt ist. Gerade diese Situation ist prädestiniert dafür, Schlussfolgerungen über andere auf der Basis von Teilm Informationen zu treffen, um die situative Komplexität des Interviews (der Interviewer muss fragen, protokollieren, sich neue Fragen ausdenken, Hypothesen entwickeln und prüfen, die Antworten beurteilen etc.) zu reduzieren. Es bleibt dabei nicht aus, dass Beurteilungsprozesse im Rahmen von Interviews eher Resultat globaler Eindrucksbildung sind. Das erklärt, wie diese geringe Validität zustande kommt (vgl. auch Schuler u. Höft, 2001)).

Um zu verhindern, dass die subjektive Filterung von Informationen vor dem Hintergrund individuell gefärbter Erfahrungen geschieht und globale Urteilstendenzen begünstigt, werden z.B. auf den Arbeitsplatz bezogene Anforderungsanalysen mit der Critical Incident Technique (Flanagan, 1954) durchgeführt. Damit wird die Grundlage geschaffen, aus den Anforderungen heraus individuelle Fähigkeiten zu definieren. Standardisierte Interviewleitfäden mit vorgegebenen Beurteilungskategorien und berufsrelevante Managementsimulationen, wie man sie vom AC her kennt, sollen helfen, die Rationalisierung von Beobachtung, Beurteilung und Bewertung zu fördern und eine globale Eindrucksbildung verhindern. Verhaltensweisen sollen quasi "isoliert" erfasst werden - ein unbestreitbarer Vorteil gegenüber der nicht standardisierten Interviewsituation.

AC, die die o.a. Komponenten in sich vereinigen können, leben von dem Versprechen, einen Menschen in berufsrelevanten Situationen genau betrachten, seine Schwächen und Stärken differenziert offenzulegen und seinen Berufserfolg präzise vorhersagen zu können (vgl. Turnage et al., 1984; Thornton et al., 1987). In einem AC sind Beobachter

gezwungen, beurteilungsrelevante Verhaltenseigenschaften von Kandidaten neben einer Fülle von anderen Verhaltensweisen in unterschiedlichen Situationen gezielt herauszufiltern und dann zu beurteilen. Als Diagnostiker in der AC-Situation müssen sie lediglich lernen, auf sie einwirkende und durch sie verursachte Störfaktoren zu kontrollieren. Doch auch in einem derartigen Setting, in dem scheinbar "analytisch" beobachtetes Verhalten beurteilt wird, gibt es Einflüsse, die die Urteilsqualität der Beobachter beeinträchtigen. Sie kommen z.B. dadurch zustande, dass Beobachter den beurteilten Personen dann größeres Interesse entgegenbringen, wenn sie sich diesen Personen gegenüber als ähnlich einschätzen.

Taft (1959) berichtet, dass die höchste Validität von den Beobachtern erzielt wird, die am meisten Ähnlichkeit mit den zu beurteilenden Personen verspüren (vgl. auch Kagan, 1967). McLaughlin (1979) konnte ebenfalls nachweisen, dass Beobachter die Merkmale von Beurteilten exakter wiedergeben konnten, von denen sie meinten, dass diese ihren eigenen Merkmalen ähnlich seien.

Das Problem, das dem Beurteilungsprozess im Rahmen eines AC anhaftet, besteht darin, dass es, wie Comelli (1973) feststellte, schwer fällt, Personenmerkmale und Verhaltensmerkmale isoliert voneinander zu beurteilen. Eine Erklärung dafür liefert Hofsommer (1980), dessen Auffassung nach die AC-Situation dadurch geprägt wird, dass ein Machtgefälle zwischen Beobachtern und Teilnehmern, wie bei jeder Art von Personalbeurteilung, in Form von einer komplizierten sozialen Wechselwirkung zwischen dem Beobachter und dem Beurteilten besteht, das Wahrnehmungsverzerrungen und/oder Fehlinterpretationen begünstigt. Ist sich jedoch der Beobachter dieser Wechselwirkung und seiner persönlichen Art, Personen und Verhaltensmerkmale wahrzunehmen und zu interpretieren, bewusst, werden derartige Einflüsse reduziert.

Managementdiagnostik, die den Anspruch hat, zukünftigen Berufserfolg prognostizieren zu wollen und damit dem Gedanken Pawliks (1976) folgt, Eignungsdiagnostik als Prozessdiagnostik zu betrachten, muss bemüht sein, Verfahren zu entwickeln, die die Kompetenzen oder die Potentiale einer Person unabhängig vom jeweiligen Diagnostiker differenziert erfassen können.

Um dieses Ziel zu erreichen ist es erforderlich, Fehler bei der Beobachtung und Beurteilung von Verhalten, die einen klaren Blick für die Potenziale von Bewerbern verhin-

dern, zu identifizieren und geeignete Gegenmaßnahmen zu entwickeln.

## 2.2 Stereotype

Beobachtungs- und Beurteilungsfehler entstehen z.B. ganz allgemein durch die Art und Weise, wie Personen wahrgenommen werden. Koltuv (1962) beschrieb, dass Menschen die Tendenz haben, Verhaltens- oder Charaktereigenschaften eher im Sinne einer globalen und nicht im Sinne einer differenzierten Eindrucksbildung wahrzunehmen. Wird bspw. eine Person als engagiert erlebt, dann vermutet man auch Eigenschaften wie Durchsetzungsvermögen oder Tatkraft und glaubt, einen „Erfolgstyp“ vor sich zu haben (Halo-Effekt, s.u.).

Dass sich diese Einschätzung mit der Realität nicht decken muss, liegt auf der Hand: im Prinzip werden hier Stereotype der Beobachtenden reproduziert. Zu wissen, "wie ein erfolgreicher Manager zu sein hat", bewirkt die Suche nach entsprechenden Verhaltensankern oder aber nach typischen inhaltlichen Aussagen, die das Normen- und Werteverständnis (= Stereotyp) des Beobachters am besten wiedergeben.

Diese Form der Eindrucksbildung über andere Menschen lernen wir quasi automatisch. Bei bestimmten Reaktionen von anderen Menschen diesen bestimmte Eigenschaften zuzuordnen (vgl. Feldman, 1981) erleichtert alltägliche soziale Wahrnehmungsprozesse und entlastet vor kontinuierlichen Entscheidungs- und/oder Prüfprozeduren, da ein Gesamteindruck aus (teilweise) unvollständigen Informationen erstellt wird (bzgl. der Bildung von Schemata vgl. DeSoto, 1961).

In eine ähnliche Richtung wiesen die früher Ergebnisse von Asch (1946). Demzufolge gibt es abhängig vom jeweiligen Beobachter charakteristische Eigenschaften, die einen Einfluss auf die Eindrucksbildung über andere Menschen ausüben. Welche Eigenschaften dies sein können, hängt von dem Einzelnen, seiner Lernerfahrung und seiner sozialen Wahrnehmung ab. Benedetti und Hill (1960), sowie Jones (1954) fanden in ihren Untersuchungen heraus, dass man sich in der Beurteilung anderer Menschen von den typischen Auffassungen der soziale Gruppe, der man sich zugehörig fühlt, leiten lässt. Sie werden darin durch Uhlenbrock und Vormberger (1990) bestätigt, die bei ihren Auswertungen von Beobachterurteilen unterschiedliche Faktorenstrukturen in Ab-

hängigkeit von der jeweiligen Berufsgruppe der Beobachter nachweisen konnten (vgl. auch Schuler u. Prochaska, 1999).

Auf Unterschiede bezüglich der Urteilsgenauigkeit innerhalb einer bestimmten sozialen Gruppe verweisen z.B. Luft (1950) oder Cattell (1957). Neubauer (1990) identifizierte bspw. im Gegensatz zu McClelland (1951) geschlechtsabhängige Unterschiede zwischen Frauen und Männern in ihrer Art und Weise, Verhalten zu beobachten – möglicherweise ein Effekt geschlechtsspezifischer Unterschiede (vgl. Halpern, 1997).

### **2.3 Beobachtungs- / Beurteilungsfehler**

Andere Fehler in der Beobachtung und Beurteilung von Personen beziehen sich auf den Prozess der Informationsverarbeitung von der Beobachtung über die Beurteilung zur Klassifizierung der wahrgenommenen Verhaltensweisen, die die Basis für die Personalentscheidung bildet. Die Fehlerhaftigkeit dieses Prozesses wurde vielfach untersucht (vgl. Saal et al., 1980), und die Beseitigung von Beobachtungs- und Beurteilungsfehlern als wichtiger Beitrag zur Verbesserung von konvergenter und diskriminanter Validität diskutiert (s.u., vgl. auch Schuler u. Höft, 2001). Diese Fehler sollen im Folgenden kurz dargestellt werden – die Klassifizierung orientiert sich dabei an den Ausführungen von Bortz und Döring (1995).

Der gravierendste Beurteilungsfehler ist nach Thorndike (1920) der Halo-Effekt. Damit ist die Tendenz gemeint, aufgrund eines dominierenden Einzelmerkmals den Gesamteindruck über eine Person zu bilden. Das kann sich z.B. darin äußern, dass sich der Teilnehmer eines AC in einzelnen Verhaltensmerkmalen nicht mehr differenziert darstellen lässt. Zum Halo-Effekt lassen sich verschiedene Interpretationen entwickeln. Einmal könnte es sein, dass Menschen per se eher global als differenziert beobachten; dann könnte man unterstellen, dass sie nicht in der Lage sind, zwischen verschiedenen Verhaltensmerkmalen zu differenzieren und schließlich könnte man vermuten, dass es eine allgemeine Tendenz gibt, Menschen in unterschiedlichen Verhaltensmerkmalen ein gleich hohes Leistungsniveau zuzusprechen. Entscheidend ist, dass bei der Beurteilung von Personen über eine Anzahl verschiedener Dimensionen eher ganzheitlich geurteilt wird als sich auf separate Dimensionen zu beziehen. Willingham und Jones (1958) nannten diesen Effekt „composite halo“.

Weiterhin ist der Effekt des ersten Eindrucks zu nennen, der, wenn er sich stabilisiert, zum Halo-Effekt wird. Der Effekt des ersten Eindrucks wird durch implizite Persönlichkeitstheorien sowie durch Vorurteile begünstigt und kommt dadurch zustande, dass die Eindrücke, die man zu Beginn der Beobachtung gesammelt hat, wegen der Flut der Informationen während des weiteren Beobachtungsprozesses am ehesten haften bleiben. Sympathie und Antipathie spielen hier eine bedeutsame Rolle, denn anscheinend erfolgt nach diesen Kriterien die erste Beurteilung anderer Menschen.

Koltuv (1962) bspw. fand heraus, dass bei einem Fehlen von Sympathie die Korrelationen zwischen den Urteilen über verschiedene Verhaltensmerkmale hinweg stiegen (ein Indikator für eine undifferenzierte Wahrnehmung des Gegenübers). Sie interpretierte, dass ohne den Faktor „Sympathie“ möglicherweise Informationen fehlen, die zu einer Verunsicherung der Beurteiler führen und dadurch eine differenzierte Betrachtung verhindert wird.

Eine andere Fehlerquelle ist der Milde- bzw. Strengeeffekt. Er entsteht dadurch, dass man dem Gegenüber entweder eine zu hohe oder zu niedrige Bewertung zukommen lässt, oder aber von seinem individuellen Beurteilungsstil her durchweg milde oder streng urteilt (vgl. Saal, Downey u. Lahey, 1980).

Als weiterer Beobachtungs-/Beurteilungsfehler ist der Kontrasteffekt zu nennen. Hierbei handelt es sich um das Phänomen, dass die Urteilsbildung durch vorhergegangene Ereignisse nachhaltig beeinflusst wird. Wird beispielsweise in einer mündlichen Prüfung ein Prüfling beurteilt, der „objektiv“ gesehen eine mittelhohe Selbstsicherheit besitzt, und war vor ihm ein selbstunsicherer Prüfling an der Reihe, dann wird dieser Prüfling anders beurteilt, als wenn vor ihm ein Prüfling mit einem hohen Grad an Selbstsicherheit an der Reihe gewesen wäre.

Weitere Beurteilungsfehler sind:

- der Recency-Effekt; die zuletzt erhaltenen Eindrücke beeinflussen das Beobachterurteil;
- die Orientierung am persönlichen Wertesystem; der Beobachter orientiert sich in seinen Beurteilungen an seinem eigenen Anspruch an sich selbst (vgl. Heitmeyer & Thom, 1988)

- der Kontakt-Effekt: der häufige Kontakt mit einem Teilnehmer zieht Urteilsverzerrungen nach sich.

#### **2.4 Strukturelle Massnahmen zur Verbesserung von konvergenter und diskriminanter Validität**

Die in der Literatur als problematisch benannten "Störquellen", die Beobachtungs- und Beurteilungsfehler (vgl. Brandstätter, 1979; Fox, Bissman u. Herrman, 1983; Jeserich, 1981; Murphy, 1982), finden ihren Niederschlag in den unbefriedigenden niedrigen konvergenten und diskriminanten Validitäten (vgl. Campbell u. Fiske, 1958) der AC-Ergebnisse. Unter geringer konvergenter Validität ist zu verstehen, dass die Beobachterurteile bezüglich ein und derselben Beobachtungsdimension über verschiedene Übungen hinweg niedrig miteinander korrelieren. Das bedeutet, dass diese spezielle Beobachtungsdimension nicht eindeutig identifiziert werden kann.

Unter niedriger diskriminanter Validität ist zu verstehen, dass die Korrelationen der Beobachterurteile über die einzelnen Beobachtungsdimensionen hinweg innerhalb einer Übung hoch sind, was bedeutet, dass es bei den Beobachtern letztendlich eine globale Eindrucksbildung über die Teilnehmer gibt. Dadurch werden innerhalb einer Übung die einzelnen Beobachtungsdimensionen nicht mehr eindeutig identifiziert, was jedoch nicht erwünscht ist, wenn man bedenkt, dass die Urteile der Beobachter über berufsrelevante Verhaltensmerkmale ein guter Prädiktor für den späteren Berufserfolg sein sollen.

Ein die globale Eindrucksbildung unterstützender Faktor scheint die Komplexität des zu beobachtenden Verhaltens zu sein. Kazdin (1982) führte aus, dass, je größer die Komplexität des zu beurteilenden Verhaltens ist, je mehr Verhaltensmerkmale beobachtet und beurteilt werden müssen, um so geringer die Übereinstimmung der Beobachter sei. Handelt es sich um globale Urteile, um Eindrücke über andere Menschen, dann werden diese Eindrücke zusätzlich durch sozio-kulturelle Komponenten sowohl seitens der Kandidaten wie auch seitens der Beobachter beeinflusst. Auch "Experten", die sich berufsmäßig mit der Eindrucksbildung und der Einschätzung von anderen Menschen beschäftigen, haben Schwierigkeiten mit diesem Phänomen.

Nun können die sozio-kulturellen Facetten der Teilnehmer und/oder der Beobachter

---

nicht eliminiert werden – aber die Anzahl der zu beobachtenden Verhaltensmerkmale kann reduziert werden. Gaugler und Thornton (1989) konnten nachweisen, dass sich die konvergente Validität dann erhöhte, wenn die Anzahl der zu beobachtenden Dimensionen reduziert wird.

Das heißt für ein AC, dass bei der Konstruktion der Übungen und/oder der Beobachtungsinventare im Vorfeld genau geprüft werden muss, welche Verhaltensmerkmale für die Zielposition relevant sind – und nur die sollten dann auch im AC beobachtet werden. Da die Messgenauigkeit und die Zuverlässigkeit des AC außerdem von dem zur Verfügung stehenden Kategoriensystem, der Art der Beurteilungsskala und deren Handhabung durch die Beobachter beeinflusst werden, sollte die Tätigkeit der Beobachter zusätzlich dadurch erleichtert werden, dass diese Verhaltensmerkmale so eindeutig beschrieben sind, dass Fehlinterpretationen bei der Beurteilung des beobachteten Verhaltens möglichst unterbunden werden können.

Eine leicht handhabbare Beurteilungsskala kann ferner dazu beitragen, die Konzentration der Beobachter auf die Beobachtung beurteilungsrelevanter Verhaltensweisen zu lenken (vgl. Domsch u. Jochum, 1982) und die Beobachter nicht unnötig dadurch zu strapazieren, dass sie vielschichtige Überlegungen vornehmen müssen, um eine Verhaltensbeurteilung durchführen zu können.

Auf die Bedeutung der Eindeutigkeit zu beobachtender Verhaltensmerkmale als Unterstützung bei der Erhöhung der konvergenten Validität verweisen auch Kleinmann, Wechsler, Kuptsch und Köller (1995). Sie führen an, dass sich Beobachter, entsprechend der Hypothesentheorie der sozialen Wahrnehmung von Lilly und Frey (1993) in der ersten AC-Übung einen Eindruck darüber bilden, wie sie die Teilnehmer bezüglich der einzelnen Verhaltensmerkmale einschätzen. Sollten die Verhaltensmerkmale nun in den folgenden Übungen wenig gut beobachtet werden können, dann neigen sie dazu, auf ihre Einschätzungen aus dem Eindruck der ersten Übung zurückzugreifen.

Das bedeutet, dass die Übungen im AC und die zu beobachtenden Verhaltensmerkmale aufeinander abgestimmt sein müssen. Wenn man beispielsweise in verschiedenen Übungen Belastbarkeit beobachten möchte, dann sollten diese Übungen auch so konstruiert sein, dass das Merkmal Belastbarkeit (wie auch immer es definiert und operationalisiert sein mag) gezeigt werden kann. Oder, um es drastischer zu formulieren:

durch den spezifischen Übungstypus eines "Stressinterviews" ist es nicht naheliegend, das Verhaltensmerkmal "Teambildung unterstützen" beobachten zu wollen.

Möglicherweise ist das zu Stande kommen von geringer konvergenter Validität darauf zurückzuführen, dass bei weniger gut konstruierten Übungen genau dieser Sachverhalt nicht genügend beachtet wird. Die Feststellung von Sackett und Dreher (1982), die Beobachtungen in einem AC seien stark situativ geprägt, ist von daher gemäß Kleinmann et al. (1995) das Resultat von schlecht konstruierten Übungen und einer hohen Belastung der Beobachter bei Informationsaufnahme- und Informationsverarbeitungsprozessen durch die Anzahl der zu beobachtenden Verhaltensmerkmale.

Des weiteren gehen Kleinmann et al. (1995) davon aus, dass eine bessere Beobachtbarkeit der Verhaltensmerkmale sowie eine Reduktion derselben in den einzelnen Übungen zugleich die diskriminante Validität erhöht, die von Reilly et al. (1990) als unbefriedigend dargestellt wird. Verantwortlich für das Ergebnis von Reilly et al. (1990) ist jedoch das zeitgleiche Erfassen mehrerer möglicherweise voneinander abhängiger Verhaltensmerkmale. Hennessy, Mabey und Warr (1998) empfehlen den Einsatz von Verhaltenschecklisten. Mit diesen Listen wird den Beobachtern eine Struktur angeboten, an der sie sich orientieren können. Dadurch wird ihre kognitive Belastung reduziert und sie können sich mehr auf den Beobachtungsprozess konzentrieren als bei dem oben beschriebenen Vorgehen.

Wird zudem noch, wie Schleicher, Day, Mayes und Riggio (2002) herausarbeiteten, die Anzahl der zu beobachtenden Verhaltensmerkmale reduziert und werden die Verhaltensmerkmale klar definiert, dann hat das einen deutlichen Effekt auf die Konstruktvalidität und die diskriminante Validität.

Kleinmann et al. (1995) betonen die Bedeutung der Unabhängigkeit der zu beobachtenden Verhaltensmerkmale. Sie führten eine Untersuchung durch, in deren Rahmen sie eine Art des Beobachtertrainings wählten, bei dem mit Hilfe von Videoaufzeichnungen Konsens bei den Beobachtern über das zu beobachtende Verhalten und den Beurteilungsmaßstab sichergestellt wurde. Ferner wurde ein genauigkeits- und fehlervermeidungsorientierter Trainingsansatz gewählt. Ihre Untersuchung zeigte, dass die diskriminante Validität sich erhöht, wenn die Verhaltensmerkmale unabhängig voneinander sind und die Beobachter nur wenige Verhaltensmerkmale beobachten müssen.

## 2.5 Die Qualifizierung von Beobachtern als Möglichkeit der Qualitätssteigerung

Beobachtertrainings zielen in der Regel darauf ab, sich der eigenen Stereotype und der möglichen Beobachtungs- und Beurteilungsfehler bewußt zu werden. Zur Unterstützung dieses Bewusstwerdungsprozesses, so fordert Hofsommer (1980), muss der Beobachter psychologisch geschult werden. Derartig qualifiziert ist er in der Lage, eine kritische und differenzierte Wahrnehmung von seinem eigenen Verhalten zu gewinnen. Der Beobachter brauche ein breites psychologisches Wissen über menschliches Verhalten und benötige ein intensives Training unter kontrollierten Bedingungen. Letztendlich müsse das Training dazu führen können, dass der Beobachter seine sozialen Kompetenzen entwickeln und dann in der Beurteilung anderer Personen eine gewisse Beständigkeit vorweisen kann. Diese Fähigkeit ist nach Gage (1952), sowie Bender und Hastorf (1953) eine wichtige Voraussetzung, um zutreffende Beurteilungen über andere Menschen abgeben zu können. Hofsommer (1980) spricht hier allgemein von sozialen Kompetenzen und deutet dadurch an, dass die Fähigkeit zur Beobachtung und Beurteilung im AC nicht durch eine spezielle Berufsausbildung bedingt ist, sondern durch Schulungsprogramme und möglicherweise auch einen häufigen Einsatz als Beobachter im AC trainiert werden kann (vgl. Domsch u. Jochum, 1982). Zur Steigerung der Qualität von AC-Ergebnissen vertreten Domsch und Jochum (1982) die Auffassung, dass das Training sozialer Kompetenzen, die Schulung der sozialen Wahrnehmung oder die Sensibilität gegenüber dem eigenen Beurteilungsverhalten möglicherweise die Beobachtergenauigkeit und die Stabilität des Beobachterverhaltens zusätzlich erhöhen, und dadurch zu einer Steigerung der Qualität der AC-Aussagen führen.

Es ist jedoch zu fragen, ob breites psychologisches Wissen vom Menschen und die Entwicklung sozialer Kompetenzen allein die Gewähr für eine qualifizierte Tätigkeit als Beobachter sind. Abgesehen davon, dass diese Begriffe im Bedarfsfall genauer beleuchtet werden müßten, erscheint es sinnvoller, Wert auf eine differenzierte Tätigkeits- und Anforderungsanalyse zu legen, aus der heraus Anforderungen an das Zielverhalten der AC-Teilnehmer gelegt werden können. Je genauer eine Tätigkeit beschrieben werden kann, je genauer Anforderungen an erfolgreiche Stelleninhaber, die diese Tätigkeit ausführen, formuliert werden können, desto genauer können Beziehungen zwischen Anforderungen und erfolgsrelevanten Fähigkeiten hergestellt werden.

Vielleicht sind gerade die Zustimmung der Beobachtergruppe zu den Anforderungen,

die augenscheinliche Abbildung des Berufslebens in praxis- und realitätsnahen Übungen und die Zustimmung zu den zu messenden Fähigkeiten wichtigere Voraussetzungen für die Qualität der Beobachterleistung als die mehr oder weniger unspezifische Entwicklung sozialer Kompetenzen oder ein breites psychologisches Wissen vom Menschen. Da das umfassendste Wissen um die Anforderungen, die an ein bestimmtes Berufsbild gestellt werden, am ehesten bei den mit diesem Berufsfeld vertrauten Personen vorhanden ist, können möglicherweise diese Personen am ehesten geeignete, potenzielle Bewerber "richtiger" einschätzen als "Berufsfeld-Laien" mit einer durch Trainings erworbenen hohen sozialen Kompetenz. Die "Qualifizierung" der Beobachter wäre in diesem Falle indirekt dadurch erfolgt, dass die Tätigkeit in einem Unternehmen ein nicht in einem AC abbildbares Expertenwissen aufbaut, das zu "guten" Personalentscheidungen befähigt – unbefriedigend für eine differenzierte an berufsrelevanten-Verhaltensmerkmalen orientierte Eignungsdiagnostik.

Unter dem Blickwinkel der Identifikation der Beobachter mit dem Verfahren und seiner Akzeptanz (vgl. die frühen Ausführungen von Franke, 1959) ist es allerdings legitim und vielleicht auch erforderlich, die Sichtweise des Unternehmens durch dessen Mitarbeiter (= Beobachter) im AC zu repräsentieren. Die Kunst besteht darin, dieses Wissen im Rahmen eines ressourcenorientierten Ansatzes transparent zu machen und AC-gerecht anzupassen. Nur dann kann die Aussage von Hinrichs (1976), der meint, das AC "provides uniform standards for judgement by trained observers" bestätigt werden. Da nach Franke (1959) Beobachtung und Beurteilung wesentlich von der Einstellung des Beobachters zu der Beobachtungssituation, der Person des Beobachteten und dem Zweck der Beobachtung beeinflusst werden, muss darauf geachtet werden, dass zusätzlich die dem AC zu Grunde liegenden Rahmenbedingungen den Beobachter vermittelt und von diesen verstanden werden.

### **2.5.1 Verbesserung von konvergenter und diskriminanter Validität durch spezielle Beobachtertrainings**

Spezielle Beobachtertrainings zielen darauf ab, Beobachter dazu anzuhalten die einzelnen Anforderungsdimensionen genauer zu beobachten und durch die Vermeidung von Beobachtungs- und Beurteilungsfehler zwischen den unterschiedlichen Anforderungsdimensionen differenziert beobachten zu können (vgl. dazu bspw. Levine u. Butler, 1952; Latham, 1975; Gerpott, 1985; Landy, 1985; Höft u. Funke, 2001). Es ist

schwierig, eine klare Trennung zwischen beiden Fehlerarten einzuhalten, da sie eng miteinander verflochten sind. Aus diesem Grunde wird im Folgenden auf eine derartige Trennung verzichtet.

Wie Lievens (1998) in seinem Übersichtsartikel darstellt, können Beobachtertrainings zwischen 90 Minuten und 3 Wochen dauern. Mehrere Arten von Beobachtertrainings sind zu unterscheiden:

#### Fehlervermeidungs- und Beobachtungsgenauigkeitsorientiertes Training

Hier lernen die Beobachter, typische Beobachtungsfehler wie den Halo-Effekt, den Milde-Effekt und andere zu vermeiden (Fehlervermeidungstraining) oder die einzelnen Anforderungsdimensionen genau zu identifizieren (Beobachtergenauigkeitsorientiertes Training). Dabei kommt den Operationalisierungen der Anforderungsdimensionen eine große Bedeutung zu. Das Fehlervermeidungstraining ist weit verbreitet, jedoch führt es, wie Bernardine und Buckley (1984), sowie Woehr und Huffcutt (1994) nachweisen konnten, nicht zu einer höheren Beurteilerübereinstimmung; das Beobachtergenauigkeitstraining führt, wie Woehr (1992) zeigen konnte, zu einer Erhöhung der diskriminanten Validität.

#### Frame-of-reference-Training

Bei dieser Art von Beobachtertraining geht es darum, die Beobachter dazu zu verleiten, zu beobachtende Verhaltensweisen in einer ähnlichen Art und Weise zu beurteilen, wie es durch eine Referenzbeurteilung vorgegeben wird. Nach Woehr und Huffcutt (1994) bewirkt diese Art von Training die höchste Beobachtungsgenauigkeit. Auch Stamoulis und Hauenstein (1993) gehen davon aus, dass konvergente und diskriminante Validität durch diese Art von Beobachtertraining erhöht werden.

#### Training beurteilungsrelevanter sozialer Prozesse

Hier geht es darum, den Interaktionsprozeß zwischen dem Beurteiler, der das Ergebnis seiner Beurteilung an den Kandidaten rückmeldet und dem Beurteilten zu trainieren. Die Beurteiler werden dabei für ihre Empfindungen und Eindrücke, die sie während der Ergebnissrückmeldung haben, sensibilisiert und lernen, diese Eindrücke in den Zusammenhang zu anderen beurteilungsrelevanten Eindrücken zu stellen. Gerpott (1985) ist der Meinung, dass vor allem für derartige Trainings solche Trainingseffekte nachgewiesen werden können, die die soziale Qualität der Leistungs-

beurteilung verbessern.

Diese zuletzt genannte Trainingsform soll im Zusammenhang mit der vorliegenden Arbeit nicht weiter beleuchtet werden, da der Beobachtungs- und Beurteilungsprozess im AC selber und nicht das Rückmeldegespräch nach dem AC Gegenstand der vorliegenden Untersuchung ist.

### **2.5.1.1 Fehlervermeidungs- und Beobachtungsgenauigkeitsorientiertes Training**

Nach Cattell (1957) ist die wohl wirkungsvollste Methode zur Vermeidung von systematischen Fehlern und zur Verbesserung der Beurteilungsfähigkeit die Schulung von Beobachtern. Beobachter, die sich über mögliche Fehler bei Beobachtung und Beurteilung klar sind, können versuchen, diese zu vermeiden.

Schon die frühen Ergebnisse von Newcomb (1931) deuten in diese Richtung. Er fand heraus, dass die Interraterkorrelation von Ratern, die Jungen eines Sommercamps hinsichtlich 26 Eigenschaften einschätzen sollten, .49 betrug. Sollten diese Jungen jedoch aufgrund ihres gezeigten Verhaltens in den selben 26 Eigenschaften eingeschätzt werden, betrug die Interraterkorrelation nur noch .14. Er schloß daraus, dass... *“The close correlation between traits which is evident in the rating may, therefore, be presumed to spring from logical presuppositions in the minds of the raters, rather than from actual behavior”* (S. 288). D.h., dass die Beobachterübereinstimmung in dem Moment sinkt, in dem das aktuelle Verhalten gezielt beurteilt werden soll. Diese Befunde deuten an, dass in dem Moment, in dem eine differenzierte Verhaltensbeurteilung stattfinden soll, die Beobachter unterschiedliche Dinge beobachten und von daher auch zu unterschiedlichen Beurteilungen gelangen (vgl. auch die Untersuchung zur Mitarbeiterbeurteilung von Ivancevich (1979); danach zeigten Vorgesetzte, die nach einem beurteilerfehlerorientierten Training trainiert worden waren, geringere Korrelations- und Mittelwertstendenzen als untrainierte Beurteiler, die die selben Mitarbeiter beurteilten).

Schon früh zeigten Bobbitt, Gordon und Jensen (1966), dass trainierte Beobachter in der Lage sind, bestimmte Urteilstendenzen zu reduzieren und die Urteilsgenauigkeit zu erhöhen. Die durch ein derartiges Training erreichte Übereinstimmung zwischen diesen

Beobachtern ist verhältnismäßig stabil, d.h., alle Beobachter beobachten dieselben Verhaltensweisen (vgl. auch Bernardin u. Pence, 1980; Pulakos, 1984). Franke, Hoffmann und Kamphaus (1974) vermuten, dass dieser Effekt darauf beruht, dass Beobachter durch das Training lernen, vorgegebene Kategorien trennscharf wahrzunehmen. Ausschlaggebend ist dabei vermutlich die Auseinandersetzung der Beobachter mit der gezielten Information über personenspezifische Überstrahlungseffekte.

Latham, Wexley und Pursell (1975) bestätigen Levine und Butler (1952) wonach hinsichtlich der Vermeidung von Beobachtungsfehlern trainierte Beobachter bei der Beurteilung von hypothetischen Bewerbern, dargeboten per Videofilm, keine Beurteilungsfehler hinsichtlich Ähnlichkeitsfehler, Kontrast- oder Halo-Effekt machten. In ihrer Untersuchung verglichen die Autoren verschiedene Formen von Beobachtertrainings und kamen dabei zu dem Schluss, dass allein das Wissen um Beurteilungsfehler die Qualität der Beobachter nicht verbessert. Die Beobachtungsqualität konnte erst durch ein Training verändert werden, bei der ein Experte als Moderator mit den Beurteilern in der Gruppe über Beurteilungsfehler diskutierte.

Die Schlussfolgerung von Latham et al. (1975, S. 554) war: *“Specifically, these results show, that training program that carefully apply basic principles of learning in accordance with the task being taught can eliminate more than one rating error committed by subjects from a different population”*.

Die Bedeutung der Reflexion über das eigene Beobachterverhalten wiesen auch Wexley et al. (1972) nach. Beobachter erbrachten nach einem ausführlichen Beobachtertraining, das zum Ziel hatte, Kontrasteffekte zu vermeiden, signifikant bessere Beobachtungsleistungen als ohne Training. Diesen Effekt führten sie darauf zurück, dass den Beobachtern die Möglichkeit gegeben wurde, ausführlich an standardisierten über Video dargebotenen Situationen die Beurteilung zu üben und gleichzeitig von einem Moderator eine Rückmeldung bezüglich der Genauigkeit der Beurteilung zu erhalten.

Auch die Möglichkeit, über den Umgang mit impliziten Persönlichkeitstheorien zu reflektieren, führt nach Lammers (1992) zu einer Verbesserung der Beobachtergenauigkeit. Eine reine Fokussierung auf die Verarbeitung von Informationen über Beurteilungsfehler und Beurteilergenauigkeit zieht hingegen keine deutliche Steigerung der Beobachterqualität nach sich (vgl. Gerpott, 1985).

### 2.5.1.2 Frame-of-Reference-Training

Die oben angesprochene Reflexionsmöglichkeit wird mit dem Frame-of-Reference-Ansatz realisiert. Goodstone und Lopez (2001) entwickelten für ein auf diesen Ansatz ausgerichtetes Training zunächst Definitionen für jede Anforderungsdimension und definierten dann die einzelnen qualitativen Leistungsausprägungen für jede einzelne Dimension. Diesen einzelnen Leistungsausprägungen ordneten sie dann spezifische Verhaltensweisen zu (vgl. Tabelle1). Im Verlauf des Beobachtertrainings gingen sie so vor, dass die Beobachter Verhaltensbeispiele von AC-Teilnehmern beobachten mussten. Dann wurde Ihnen das Beobachtungsinventar vorgelegt und sie mussten die von Ihnen beobachteten Verhaltensweisen anhand dieses Inventars einschätzen.

Item	1 (unacceptable)	2 (marginal)	3 (acceptable)	4 (superior)
<b>Customer Orientation *</b>				
1 Faulty equipment	Does not handle; makes no contact to effort Williams	Has equipment checked out when convenient	Has equipment checked out immediately	Has equipment checked out immediately. Contact Williams and customer to keep them both informed
2 Sanchez and Smith	No rating allowed			
3 Magnolia Boulevard	Does not handle; makes attempt to ensure that job is serviced properly.	Makes a note to review plans but takes no action.	Takes action to ensure that the job is scheduled and serviced in accordance with the procedure.	Takes immediately action to ensure that job is scheduled and serviced properly. Informs requesting department and customer of actions taken.
* Focuses on external and internal customer satisfaction. Understands and responds to customer needs by considering the impact of efforts on customers.				

**Tab.1: Beispiel für die Gestaltung des Beobachtungsinventars beim frame-of-reference Ansatz (vgl. Goodstone und Lopez (2001), S. 104)**

In einem dritten Schritt wurden die Einschätzungen der Beobachter mit den Einschätzungen verglichen, die zuvor Experten abgegeben hatten. Über die Unterschiede in den Beobachtungen wurde dann mit den Beobachtern einzeln gearbeitet, sodass sie ein den Beobachtungsexperten ähnliches stereotypes Beobachtungs- und Beurtei-

ungsverhalten entwickeln konnten (die Experten waren in dem Fall Psychologen). Nach dieser individuellen Trainingsphase wurde den Beobachtern nochmals eine Art Prüfungsmaterial dargeboten. In Abhängigkeit von den von ihnen gezeigten Beobachtungsleistungen wurden sie entweder individuell weiter trainiert, oder aber überhaupt nicht mehr für eine weitere Tätigkeit als Beobachter in einem AC vorgesehen.

Goodstone und Lopez (2001) äußerten sich vorsichtig zum Erfolg einer derartigen Vorgehensweise. In ihrer Untersuchung wurden letztendlich vom Unternehmen die besten 21 Teilnehmer von insgesamt 78 Teilnehmern für weiterführende Positionen vorgeschlagen. "Although the internal consistency reliability coefficients for the measures are excellent (.85 to .97) we do not know to what extent the FOR may have contributed to this outcome" (Goodstone & Lopez, 2001, S. 106).

Einen Vorteil für den Frame-of-Reference-Trainingsansatz gegenüber anderen Formen des Beobachtertrainings konnten Schleicher, Day, Mayes und Riggio (2002) in einem Vergleich nachweisen (vgl. Tabelle 2):

	<b>Trained assessors</b>	<b>Control group</b>
Communication skills	.83*	.76*
Decision making skills	.89*	.69*
Leadership Skills	.75*	.65*
Average reliability across the tree dimensions	.82*	.70*

\* Cronbach's alpha

**Tab.2: Vergleich der Interrater Reliabilitäten beim Frame-of-Reference-Training (vgl. Schleicher, Day, Mayes und Riggio (2002), S. 740)**

In diesem Vergleich erzielte die Frame-of-Reference-Trainingsgruppe gegenüber der Kontrollgruppe in punkto Beobachtungsgenauigkeit signifikant bessere Übereinstimmungen als die Kontrollgruppe. Diese Ergebnisse stimmen optimistisch, es darf jedoch nicht vergessen werden, dass, wenn man ein Frame-of-Reference-Training durchführen möchte, das Beobachtungsinventar wie schon weiter oben beschrieben gestaltet sein muss.

Daraus ergeben sich Forderungen für die Konstruktion der AC-Übungen und die Gestaltung der Instruktion für die einzelnen Übungen: die Übungen müssen nämlich so gestaltet sein, dass das gewünschte im Beobachtungsinventar beschriebene Zielverhalten mit einer großen Wahrscheinlichkeit gezeigt werden muss, um die Übung zu bewältigen. Die Instruktionen für die Teilnehmer müssen so gestaltet sein, dass diese angeregt werden, von sich aus das gewünschte Zielverhalten zu entwickeln - und zwar in einer Art und Weise, die zu den in den im Beobachtungsinventar niedergeschriebenen Referenzbeispielen für die Verhaltensausprägung passt.

## **2.6 Die Person „Beobachter“**

Trotz aller Bemühungen, durch die Gestaltung von Beobachtertrainings die konvergente und diskriminante Validität zu erhöhen, sind die Ergebnisse unbefriedigend. Ob diese Anstrengungen wirklich Wirkung zeigen, bezweifeln auch Spychalski et al. (1997), die darauf hinweisen, dass Beurteiler eines AC selten bezüglich ihrer Beurteilungen evaluiert werden und von daher auch keine Verbesserung der Beobachterqualität erwartet werden kann. Daher sollte ein Augenmerk auf die Person des Beobachters selbst gelegt werden, der während eines AC einer Fülle von Informationen ausgesetzt ist. Da Beobachter direkt und indirekt kontinuierlich beobachten, bekommen sie viele Informationen mit, die nicht beurteilungsrelevant sind, die aber davon abhängig sind, wie sich alle Teilnehmer verhalten.

Die Sinnhaftigkeit eines Beobachtertraining wird von verschiedenen Autoren in Frage gestellt. So argumentiert Franke (1974), dass auch ohne Training die Beobachter für ihre Urteilsbildung relevante Facetten der zu beurteilenden Person gemäß ihrer beobachtertypischen Sichtweise identifizieren können. Diese Facetten geraten durch ein Beobachtertraining in den Hintergrund; gleichwohl würden aber ihre besondere Beachtung zu einer genaueren Beurteilung der Kandidaten führen.

Bormann (1977) fand in seiner Untersuchung, dass Urteilsgenauigkeitsmaße, die an standardisierten Situationen und hypothetischen Personen, dargeboten in Form eines Films, erhoben wurden, nicht signifikant mit psychometrischen Indizes für Beurteilungsfehler korrelierten. In einer anderen Untersuchung konnten Bormann (1979) wie auch Bernardin und Pence (1980) nachweisen, dass ein fehlerorientiertes Beobachtertrai-

ning sogar zu einer Verschlechterung der Urteilsgenauigkeit führte. Sie widersprechen damit Ivancevich (1979), indem sie argumentieren, Korrelations- und Mittelwertstendenzen seien keine Belege für „Fehler“ im Verhalten der Beurteiler, sondern spiegeln vielmehr die Leistung der Beurteilten genau wider. Daher müssen die Beobachter in einem AC dazu angeleitet werden, den Beobachtungsprozess in die Elemente "Beobachten", "Beurteilen" und "Bewerten" zu trennen.

Diese Trennung vorzunehmen war in den Anfängen der AC-Methode ausgesprochen schwierig, wenn man berücksichtigt, wie die Beobachter traditionell zu ihren Beurteilungen im AC gekommen sind: damals wurde das Verhalten der Kandidaten während der einzelnen Übungen von den Beobachtern protokolliert und dann wurden in einer Beobachterkonferenz erst am Ende des AC die Beurteilungen bezüglich der beobachteten und protokollierten Verhaltensweisen festgelegt. Bei diesem Vorgehen verlieren jedoch die Beobachter durch das Niederschreiben möglicherweise wichtige Informationen, da sie durch das Protokollieren der Verhaltensweisen und das gleichzeitige Beobachten des Verhaltens selber in ihrer Aufmerksamkeit beeinträchtigt werden. Daher fordert Hasemann (1983) auch, in einem Beobachtertraining die Wahrnehmungsschärfe zu schulen, Teilaspekte aus dem Beobachtungsspektrum herausfiltern zu lernen und sprachliche Sicherheit der Beschreibungen zu trainieren. Da Menschen ihre soziale Umwelt in Schemata einordnen mit deren Hilfe sie ihren Wahrnehmungsprozess vereinfachen, führt möglicherweise diese Art der Urteilsfindung zu Fehlurteilen (siehe oben). Andererseits führte der Diskussionsprozess in den Beobachterkonferenzen zu Einflüssen auf die Urteilsfindung z.B. durch in der Hierarchie des Unternehmens höher stehende Beobachter, oder durch Beobachter, die wortgewandt "ihren" Kandidaten durch das AC bringen wollten. Letzteres erklärt z.B. die Forderung von Jeserich (1981), keine direkten Vorgesetzten als Beobachter in das AC zu integrieren.

Auf einen interessanten Aspekt weisen Exler und Kleinmann (1997) hin. Sie vermuten, dass beobachtende Experten und Novizen gemeinsame impliziete Kriterien nutzen, die die Eignungsurteile beeinflussen. Klimonski und Stricklund (1977) brachten bereits mit ihrem Begriff der "indirekten Kriterienkontamination" zum Ausdruck, dass der Einfluss implizierter Kriterien der Beobachter für das zu Stande kommen der prädiktiven Validität verantwortlich ist. Diese impliziten Kriterien werden nicht artikuliert, sie sind aber den Mitgliedern einer Organisation bewusst, wenn sie auch möglicherweise tabuisiert sind. Das führt dazu, dass im Beobachtungsprozess nicht berufsrelevante, aus gezielter Be-

obachtung abgeleitete Verhaltensleistungen der eingestellten Person, sondern eben besagte implizite Kriterien berücksichtigt werden. Exler und Kleinmann (1997) vermuten, dass bei einer über die Hierarchien zusammengesetzten Beobachtergruppe die Beobachter der unteren Hierarchien sich an den vermuteten Vorstellungen (= impliziten Kriterien) ihrer jeweiligen Vorgesetzten orientieren. Wenn dem so wäre, dann hätten hierarchieübergreifend sowohl Vorgesetzte wie auch Mitarbeiter relativ stabile, dauerhaft vorhandene gemeinsame Wahrnehmungsstrukturen in ihrem kognitiven Netzwerk abgebildet und würden dadurch zu ähnlichen Ergebnissen kommen. Möglicherweise können diese Ergebnisse auch erzielt werden, wenn beide ähnliche implizite Persönlichkeitstheorien nutzen (vgl. Lammers, 1992).

### **2.6.1 Psychologen, Experten und Novizen als Beobachter**

Verbesserungen der Urteilsqualität sollen auch dadurch erzielt werden können, dass zusätzlich zum Beobachtertraining bestimmte Berufsgruppen als Beobachter tätig sein sollen. So erhöhen bspw. Psychologen die Qualität der Beobachtung im AC (vgl. Jeserich, 1981). Sie haben aufgrund ihrer Berufsausbildung am leichtesten Zugang zur Materie des AC und kommen dank ihrer spezifischen Kompetenzen eher zu den „richtigen“ Beobachtungen, Beurteilungen und Schlußfolgerungen über die beobachteten Personen. Auch Thornton et al. (1987) führten in ihrer Meta-Analyse den Nachweis, dass Psychologen als Beobachter die Qualität der AC-Ergebnisse verbessern.

Diese Meinung wird von Domsch und Jochum (1982) nicht geteilt (vgl. auch Lievens, 2001). Bereits Murray (1938) vertrat die Ansicht, dass die analytische Sichtweise des Psychologen, bei der das Emotionale weitgehend zurückgedrängt wird, zu einer Verschlechterung der Beurteilungsfähigkeit führt. Die Untersuchungen von Buzby (1924), sowie Kelly und Fiske (1951) scheinen diese Meinung zu bestätigen, und Kremers (1962) vermutete, dass die Verschlechterung der Beurteilungsfähigkeit von Psychologen darauf zurückgeführt werden kann, dass diese Berufsgruppe in der Beurteilung anderer Menschen aufgrund ihres Wissens eher verunsichert ist.

Alternativ zu der Berufsgruppe der Psychologen sollen auch gut vorbereitete und trainierte Beobachter die erforderliche Leistung erbringen können; sie sind dann der Leistung von Spezialisten ebenbürtig (Jeserich, 1981; Sagie u. Magnezy, 1997; Lievens,

2001)). Allerdings fehlt hier die Erklärung, was einen "Spezialisten" auszeichnet.

Thornton et al. (1987) konnten keinen systematischen Zusammenhang zwischen der Validität des AC und der Dauer des Beobachtertrainings, dem Zeitpunkt der Kriterien-erhebung und der Zeit, die zur Integration der Daten bereitgestellt wird, finden. Sie schlossen daraus, dass Beobachter, die betriebliche Normen und Werte bezüglich Beförderung intuitiv kennen, einen Blick für die Personen haben, die im Unternehmen vorankommen werden, und diese besser beurteilen können als "außenstehende" Beobachter.

Greenwood und McNamara (1967) kamen zu ähnlichen Ergebnissen. Sie untersuchten die Beobachterübereinstimmung in einem situativen Testverfahren mit AC-typischen Übungen (drei Gruppenübungen unterschiedlichen Charakters). Beobachter mussten dazu unabhängig voneinander sechs Teilnehmer auf einer fünfstufigen verbal und numerisch operationalisierten Skala bezüglich des Kriteriums "Effektivität" einschätzen. Dieses Kriterium wurde zusätzlich durch Verhaltensanker erläutert. Ausserdem mussten sie die Teilnehmer nach jeder Übung in eine Rangfolge bringen. Beobachter waren in dieser Untersuchung Psychologen/Soziologen (= Experten, d.Verf.) und „*high-level experienced manager but with only minimum training in assessment center technique*“ (= Novizen, d.Verf.). Teilnehmer an dem Testverfahren waren Mitarbeiter einer Elektrofirma aus dem unteren bis mittleren Management (die Werte für die Reliabilitäten der ratings und rankings finden sich im Anhang 1, S. 123).

Abgesehen davon, dass für einen bestimmten Typ von Gruppenübung (manufacturing) die Reliabilitäten höher liegen als bei den anderen beiden Gruppenübungen, wurde festgestellt, dass das ranking und das rating über alle Beobachter hinweg sich nur unwesentlich voneinander unterscheiden (vgl. Anhang 1, S. 123). Interessant ist auch der Vergleich der Reliabilitäten zwischen den Experten und den Novizen. Die Autoren interpretieren ihre Ergebnisse in dem Sinne, dass es für den von ihnen untersuchten Typ von AC keinen signifikanten Unterschied zwischen Experten und Novizen gäbe. Ihrer Meinung nach sind diese Ergebnisse darauf zurückzuführen, dass die Beobachter genaue Informationen darüber bekommen hatten, welches Verhalten von Ihnen beobachtet, und wie das Verhalten beurteilt werden soll.

Während Sagie und Magnezy (1997) betonen, dass der Beobachertyp (in ihrer Untersuchung Manager/Psychologen) einen ausgesprochen starken Einfluss auf die diskri-

minante Validität des AC hat (Psychologen nutzen für ihre Urteilsbildung fünf Verhaltensmerkmale, Manager zwei), kann dieses Ergebnis von anderen Autoren nicht bestätigt werden. Lewry (1993) fand in seiner Untersuchung heraus, dass die Erfahrung im Umgang mit AC (man könnte diesen Faktor als Beobachtertraining on the job gelten lassen) entgegen der Auffassung von Mash (1974), wonach die Beobachtergenauigkeit stark von situativen Komponenten abhängt, die in einem AC wirken, und auf die die Beobachter vorbereitet werden sollten, keinen signifikanten Einfluß auf die AC-Ergebnisse hatte. Auch Lorenzo (1984) zeigte, dass Beobachter mit einem Beobachtertraining im Vergleich zu Novizen weniger „dimensional accuracy“ aufweisen.

Auf einen anderen Aspekt, die Bedeutung von Erfahrungswissen wies erstmalig Taft (1959) hin. Seine Forderung, Personalverantwortliche sollten als Beurteiler eingesetzt werden, gründete er auf der Erkenntnis, dass sich die höchsten Validitäten bei Beurteilern ergaben, die am meisten mit den einzuschätzenden Teilnehmern vertraut waren. Das bedeutet, dass beispielsweise Linienmanager als Beobachter gewonnen werden müssten.

Dimension	Control training	Data-driven training	Schema-driven training
Problem analysis and solving			
Students	.62	.72	.80
Managers	.73	.76	.88
Interpersonal Sensitivity			
Students	.79	.85	.88
Managers	.85	.86	.91
Planning and Organization			
Students	.64	.77	.83
Managers	.70	.87	.85
Note: These generalizability coefficients were obtained by separate generalizability studies within each dimension. Assessors were the facets. Candidates were the object of measurement.			

**Tab.3: Lievens (2001, S. 262): Generalizability Coefficients per Dimension for various Samples and Training Groups**

Lievens (2001) betont, dass Manager auf jeden Fall als Beobachter einbezogen werden sollten, wenn es darum geht, Personalauswahl zu betreiben. Lievens (2001) streitet nicht die Werthaftigkeit eines Beobachtertrainings ab (zum Frame-of-Reference-Training vgl. die Ausführungen weiter oben).

Er konnte aber zusätzlich herausarbeiten, dass es Sinn macht, in Anlehnung an Schleicher und Day (1998) die Normen und Werte einer Organisation zu identifizieren, transparent zu machen und dann in ein Leistungsmesssystem zu übertragen. Dieses System muss, so seine Forderung, den unternehmensinternen Beobachtern präsentiert werden. Dann ist zu erwarten, dass diese Beobachter im AC ihr mentales Netz benutzen, um das beobachtete Verhalten nach für sie relevanten Aspekten abzuklopfen. Er bestätigt damit Huck (1973) der formulierte: "*Other relevant dimensions of job performance must be identified and defined, and assessment center techniques designed to measure them*" (S. 209).

## 2.7 Forschungshypothesen

Beobachterübereinstimmung und Beobachtergenauigkeit sind Maße, mit denen man die Qualität der AC-Ergebnisse überprüfen kann. Die Fähigkeit, Menschen zu beurteilen, ist trainierbar, und schlägt sich möglicherweise in den beiden o.a. Maßen nieder. Allerdings deuten die o.a. Ausführungen an, dass es widersprüchliche Ansichten zum qualitätsverbessernden Effekt von Beobachtertrainings gibt. In den Anfängen des AC wurden die Beobachter dadurch belastet, dass sie das Verhalten der Kandidaten protokollieren und dann interpretieren mussten (es gab keine systematische Beobachterführung). Was protokolliert wurde, blieb jedem Beobachter selber überlassen – die Beobachtungsergebnisse wurden in einer Beobachterkonferenz zusammengetragen und dann kollektiv interpretiert. Die mit einem derartigen Vorgehen verbundene Unschärfe in der Beobachterübereinstimmung führte zu einer systematischen Untersuchung typischer Beobachtungsfehler, zur Entwicklung von fehlervermeidungs- und beobachtungsgenauigkeitsorientierten Beobachtertrainings und schließlich zum Frame-of-Reference-Training. Dabei wurden handhabbare Beurteilungsskalen und ein eindeutiges System der Zuordnung von beobachteten Verhaltensweisen zu den zu beurteilenden Verhaltenskategorien genutzt, um Fehlinterpretationen der Beobachter nach Möglichkeit zu vermeiden.

Parallel dazu führten die Erkenntnisse aus der Untersuchung des Beurteilungsprozesses im AC zur Entwicklung von Verhaltenschecklisten zur gezielten Beobachtung von beurteilungsrelevanten Verhaltensweisen, abgeleitet aus Tätigkeitsanalysen und Stellenbeschreibungen. Die dadurch erzielten Verbesserungen in punkto Beobachterge-

nauigkeit und Beobachterübereinstimmung konnten noch gesteigert werden, als Erfahrungswissen in die Konstruktion von Beobachtungsinventaren integriert wurde. In der Personalauswahl erfahrene Manager sind daher ein weiterer Gewinn für die Qualität von AC-Ergebnissen, unabhängig davon, ob sie Experten und er Novizen sind, weil sie beide möglicherweise auf ihnen gemeinsame implizite Kriterien im Personalbeurteilungsprozess zurückgreifen. Zusätzlich können Psychologen qualitätssteigernd wirken, weil sie differenzierter zu ihrer Urteilsbildung kommen als erfahrene Manager.

Wenn nun in einem Beobachtungsinventar Erfahrungswissen abgebildet wird, und zusätzlich Verhaltenschecklisten eingesetzt werden, mit denen Beobachter gesteuert werden, die dieses Erfahrungswissen teilen, dann muss es möglich sein, dass erfahrene und unerfahrene Beobachter zu ähnlichen Schlussfolgerungen über AC-Teilnehmer kommen. Durch das Erfahrungswissen werden die Normen und Werte der Organisation abgebildet, die den Organisationsmitgliedern geläufig sind. Eine Fokussierung auf diese Normen und Werte im Beobachtungsprozess während des AC mit Checklisten schützt die Beobachter vor Informationsüberladung und Fehlinterpretationen. Für die Konstruktion eines AC könnten die o.a. Ergebnisse bedeuten, dass es Sinn macht, die Normen und Werthaltungen des sozialen Umfeldes, in dem das AC abgehalten wird im Beobachtungsinventar abzubilden. Das kann z.B. dadurch geschehen, dass das Beobachtungsinventar und die AC-Übungen von den Mitarbeitern des Unternehmens entwickelt werden. Damit ist die Erwartung verbunden, dass Unterschiede zwischen den Beobachtern schon im Vorfeld reduziert werden können, weil die Beobachter ihre Gedanken- und Erlebnisswelt im Verfahren repräsentiert fühlen. Vermutlich kann man sogar, wenn die o.a. Aspekte berücksichtigt werden, auf ein Beobachtertraining verzichten.

Folgende Forschungshypothesen sollen diese Vermutung zu überprüfen:

1. Wenn spezifische trainierbare Beobachterkompetenzen für die Qualität von AC-Ergebnissen ausschlaggebend sein sollen, dann müssen Unterschiede in der Beobachterübereinstimmung in Abhängigkeit vom Trainiertheitsgrad der Beobachter bestehen. Unternehmensinterne Führungskräfte und/oder Mitarbeiter, die wenigstens einmal bei einem AC als Beobachter dabei waren (= trainierte Experten) korrelieren dann mit einem vielfach trainierten

AC-Leiter und einer gut trainierten AC-Assistenz höher als unternehmensinterne Mitarbeiter, die noch nie bei einem AC als Beobachter dabei waren (= untrainierte Novizen).

2. Die Korrelationen innerhalb der einzelnen AC-Übungen sind für die Experten geringer als für die Novizen, weil die Experten besser zwischen den Verhaltensmerkmalen differenzieren können als die Novizen (für die Novizen lässt sich ein Halo-Effekt nachweisen).
3. Hat das Beobachtertraining einen Einfluss auf die Beobachterübereinstimmung, dann bilden die rankings trainierter Beobachter über die Teilnehmer genauer die Reihenfolge der Teilnehmer im AC ab als die rankings untrainierter Beobachter.

### 3 Methode

#### 3.1 Die Entwicklung des AC-Systems

Die vorliegende Untersuchung stützt sich auf ein AC, bei dessen Konstruktion folgende Aspekte und Forderungen berücksichtigt wurden:

##### Beobachtungsinventar

- man braucht ein eindeutiges Beobachtungsinventar, das den Beobachtern die Arbeit erleichtert und hilft, Fehlinterpretationen zu vermeiden (vgl. Reid, 1982; Kleinman et al., 1995);
- die einzelnen Verhaltensmerkmale müssen benannt und mit Verhaltensankern versehen werden, die Verhaltensweisen möglichst prägnant beschreiben (vgl. Taylor u. Hastmann, 1956; Rojahn, 1984; Kleinmann et. al., 1995);
- die Beurteilungsskala muss leicht handhabbar sein (vgl. Domsch und Jochum, 1982);
- die Übungen müssen realitätsnah sein (vgl. Jeserich, 1981; Arbeitskreis Assessment Center e.V., 1996);

##### Beobachtertraining

- man muss Bewusstwerdungsprozesse schulen (vgl. Hofsommer, 1980); einen ähnlichen Effekt hat möglicherweise auch der häufige Einsatz als Beobachter in einem AC (vgl. Domsch u. Jochum, 1982);
- Expertenwissen (vgl. Lievens, 2001) ist wichtig, um zu guten Beobachtungsergebnissen zu kommen;

##### Datengewinnung

- Beobachtungsfehlern kann man begegnen, indem man Mittelwerte über die Urteile mehrerer Beobachter bildet (vgl. Cooper, 1981);
- die analytische Sichtweise und die emotionale Sichtweise haben beide ihre Berechtigung (Murray, 1938; Meehl, 1954, Sawyer, 1966; Feltham, 1988);
- um Beurteilungsfehler zu vermeiden, sollte man zuerst alle Teilnehmer auf einer Verhaltensdimension beurteilen, bevor man zur nächsten Verhaltensdimension übergeht (vgl. Landy, 1985).

### 3.1.1 Das Beobachtungsinventar und seine Verhaltensmerkmale

Ein allgemeines Beobachtungsinventar bildet die Grundlage für das im Rahmen dieser Arbeit relevante spezifische Beobachtungsinventar. Die Konstruktion dieses "Basis-Beobachtungsinventars" vollzog sich in drei Schritten: im ersten Schritt wurde durch die Analyse von verfügbarer Literatur und die Kenntnis von AC-Systemen verschiedener Unternehmen ein Pool von 20 Verhaltensmerkmalen zusammengestellt (siehe Anhang 3).

Im zweiten Schritt wurden zu diesen Verhaltensmerkmalen konkrete Definitionen gesucht. Um diese Definitionen zu den Verhaltensmerkmalen entwickeln zu können, wurde mehreren Pädagogen, Psychologen, Juristen, Mediziner, Ingenieuren und Betriebswirten (n= 14) in Interviews die Frage gestellt:

*"Was verstehen Sie unter dem Begriff ..."*

Aus der Auswertung der Interviewdaten heraus ergaben sich die konkreten Definitionen für diese 20 Verhaltensmerkmale. Da nach Rojahn (1984) operational konkret formulierte Verhaltensmerkmale generell zu besseren Interrater-Übereinstimmungen führen, wurde zu den Verhaltensmerkmalen und deren Definitionen von zwei Psychologen ein Pool von 15 positiv formulierten Operationalisierungen pro Verhaltensmerkmal erstellt. Diese sollten so konkret wie möglich beobachtbare Verhaltensweisen beschreiben und den Beobachtern wenig Raum zu Interpretationen geben.

Im dritten Schritt wurde das so komplettierte Beobachtungsinventar noch einmal dem oben beschriebenen Personenkreis zur Begutachtung vorgelegt. Hier lautete die Frage:

*"Für wie wahrscheinlich halten Sie es, dass die Operationalisierung ... zu dem Verhaltensmerkmal..... mit der Definition..... passt?"*

Anhand einer Ratingskala von 1 (= unwahrscheinlich) bis 4 (= sehr wahrscheinlich) wurden alle Operationalisierungen bezogen auf die Verhaltensmerkmale und deren

Definitionen eingeschätzt. Letztendlich wurden für jedes Verhaltensmerkmal die 10 Operationalisierungen ausgewählt, die unter den Beurteilern die größten Übereinstimmungen erzielten. Das so erstellte Beobachtungsinventar ist damit eindeutig bezogen auf seine Verhaltensmerkmale, und genügend konkret, was deren Operationalisierungen angeht.

**Abb. 1: Entwicklung des Basis-Beobachtungsinventars**

<b>1.</b>	<b>Analyse der verfügbaren Literatur</b>	Verhaltensmerkmale auswählen
<b>2.</b>	<b>Beobachtungsinventar ausarbeiten</b>	Definitionen erarbeiten Operationalisierungen erarbeiten
<b>3.</b>	<b>Beobachtungsinventar prüfen</b>	Modifizieren
<b>Endgültige Fassung</b>		

### 3.1.2 Die Beobachterbögen und die Beurteilungsskala

Die Beobachterbögen sehen folgendermaßen aus: sie bestehen aus einer Beurteilungsskala mit den Werten von 1 - 4, die quantitativ und qualitativ in grafischer Form abgebildet ist. Eine vierstufige Beurteilungsskala wurde benutzt, um typische Urteilstendenzen (z.B. die Tendenz, mittlere Werte anzukreuzen) zu vermeiden. Um den Beobachtern die Vergabe ihrer Urteile zu erleichtern, wurde für die Darstellung der Beurteilungsskala die grafische Form gewählt, ein Vorgehen, das sich an die Ergebnisse von Wells (1960) und Landy & Farr (1980) anlehnt.

**Abb. 2: Beobachterbogen**

**Beobachterbogen „Gruppenübung ....“**

Beurteilungsskala

1 -----2 -----3 -----4

nicht beobachtet    manchmal beobachtet    oft beobachtet    sehr oft beobachtet

**Aufgeschlossenheit**

- .....

\_\_\_\_\_

\_\_\_\_\_

Lammers (1992) führt an, dass bspw. Zeitdruck und die Beurteilung der Teilnehmer entweder nach dem AC oder nach Beendigung der Aufgabe als Beobachtungsfehler den "Fehler des ersten Eindrucks" begünstigen würden, der in der Praxis wenig kontrollierbar sei. Doch hier läßt sich leicht Abhilfe schaffen: man fordert die Beobachter auf, die Teilnehmer nicht bezogen auf ein komplettes Verhaltensmerkmal, sondern bezogen auf eine begrenzte Anzahl von festgelegten, zu beobachtenden Verhaltensankern (z.B. zwei Verhaltensanker) zu beurteilen.

Diese Beurteilungen werden während der laufenden Übung und mit Bleistift vorgenommen, damit einmal erstellte Beurteilungen einem sich ändernden Verhalten angepasst werden können. Die Beobachter werden instruiert, ihre Beurteilungen spontan und zeitnah zum beobachteten Verhaltensanker vorzunehmen und zu korrigieren, falls erforderlich. Diese Möglichkeit ist durch die Gestaltung der Beobachterbögen gegeben. Sie entlasten die Beobachter, indem sie deren Aufmerksamkeit steuern und eine numerisch und verbal dargebotene Beurteilungsskala vorgeben.

## **3.2 Die Rahmenbedingungen**

### **3.2.1 Auftrag und Zielgruppe**

Die vorliegende Arbeit untersucht ein AC-System, das zur Auswahl von branchenerfahrenen und branchenfremden Mitarbeitern für den Versicherungsaußendienst konstruiert wurde. Dabei wurden in 3 Jahren insgesamt 267 potenzielle Mitarbeiter in 37 AC bezüglich ihrer Eignung als Außendienstmitarbeiter überprüft. Der vorliegenden Untersuchung kommt eine besondere Bedeutung zu, da es sich hier um Daten aus "real-life-AC" handelt: die Teilnehmer waren allesamt daran interessiert, eine Anstellung im Versicherungsunternehmen zu bekommen. Die Beobachter (bis auf den AC-Leiter und die AC-Assistenz) waren Angehörige des Versicherungsunternehmens und das AC selber war Bestandteil einer mehrstufigen Auswahlprozedur. Insofern unterscheidet sie sich von Untersuchungen, die mit studentischen Beobachtern und Teilnehmern durchgeführt wurden (vgl. Lammers, 1992; Guldin, 1996; Kleinmann et al., 1995; Kleinmann, 1997, Lammers u. Holling, 2000; Kolk, 2001).

Das AC sollte, so die Anforderung des Auftraggebers, nicht länger als einen Tag dau-

ern, und die Teilnehmer sollten unmittelbar im Anschluß an den AC-Tag eine Rückmeldung über ihr Abschneiden bekommen. Als Teil der Rückmeldung sollten auch persönliche Entwicklungsmöglichkeiten und Entwicklungsnotwendigkeiten erörtert werden: das AC musste einen starken Personalentwicklungscharakter haben.

Bei dem Unternehmen handelt es sich um ein Versicherungsunternehmen, das bundesweit dezentral in Form von Geschäftsstellen organisiert ist. Innerhalb der Geschäftsstellen gibt es den Geschäftsstellenleiter, den Innendienst mit Bürovorsteher und Sachbearbeitern, sowie den Versicherungsaußendienst, bestehend aus Regionsleitern, Vertriebsleitern und Versicherungsverkäufern.

Die Außendienstmitarbeiter wurden für zwei unterschiedliche Positionen ausgewählt: einmal für die Position des Vertriebsleiters, zum anderen für die Position des Versicherungsverkäufers. Vertriebsleiter sind Führungskräfte. Sie schulen ihre Mitarbeiter und gewinnen neue Mitarbeiter dazu. Ihre Mitarbeiter sind Versicherungsverkäufer, die entweder nach § 59 HGB als fest angestellte Mitarbeiter oder nach § 84 HGB als freie Handelsvertreter für das Unternehmen tätig sind. Die hier rekrutierten Vertriebsleiter hatten allesamt Branchenerfahrung.

Versicherungsverkäufer ermitteln im direkten Kundenkontakt den Versicherungsbedarf und verkaufen Versicherungsdienstleistungen. Ihre Hauptaufgabe besteht darin, Kunden zu gewinnen. Die hier rekrutierten Versicherungsverkäufer hatten keinerlei Branchenerfahrung.

### **3.2.2 Der Stellenwert des AC im Rahmen der Auswahlprozedur**

Von einem typischen Auswahl-AC unterschied sich dieses AC dadurch, dass hier nicht der beste oder die beiden besten Teilnehmer gesucht wurden, sondern die Bewerber mit der für das Unternehmen günstigsten Erfolgsprognose: jeder Teilnehmer eines AC hätte theoretisch eine Anstellung erhalten können. Dadurch, dass Teilnehmer von verschiedenen Geschäftsstellen zum AC entsandt wurden, sollte vermieden werden, dass sie sich in einen unmittelbaren Konkurrenzkampf um nur eine einzige Position verwickelt fühlten.

Für alle potentiellen Mitarbeiter war zwingend notwendig, ein derartiges Verfahren zu absolvieren. Da mit den Bewerbern in der Regel vor dem AC mehrere Vorgespräche geführt wurden, in deren Rahmen ein Biografischer Fragebogen zum Einsatz kam, war das AC nur ein Mosaikstein des Auswahlprozesses unter mehreren anderen. Gemäß der Anforderung des Auftraggebers hatte das AC einen empfehlenden Charakter.

Für die Personalverantwortlichen der zukünftigen Mitarbeiter sollte keine Verpflichtung bestehen, Teilnehmer, die im AC schlecht abschnitten, abzulehnen und umgekehrt. Sie sollten sich vielmehr mit den AC-Ergebnissen konstruktiv auseinandersetzen und auf deren Basis individuelle Entwicklungspläne für die zukünftigen Mitarbeiter ausarbeiten.

Die Ergebnisse des AC-Tages wurden bei einer Anstellung im Unternehmen Bestandteil der Personalakte, andernfalls waren sie zu vernichten. Eine weitere Voraussetzung für den erfolgreichen Einsatz des AC-Systems war seine Genehmigung durch den Betriebsrat des Unternehmens, der dem Einsatz dieses Systems wegen seiner Transparenz und Realitätsnähe zustimmte.

### **3.3 Das AC-Material**

#### **3.3.1 Zielgruppenspezifisches Beobachtungsinventar und Anforderungsprofil für Vertriebsleiter**

Die Vertriebsleiter wurden in der Vergangenheit bereits mit einem AC ausgewählt. Allerdings hatte dieses System (nicht die AC-Methode selber) im Unternehmen wegen seiner gravierenden (psychologischen, ethischen und methodischen, d. Verf.) Mängel eine geringe Akzeptanz. Es galt also ein AC-System zu konzipieren, das bei den Personalverantwortlichen für Akzeptanz bezüglich der AC-Ergebnisse sorgte und deren personalentwicklerische Arbeit förderte.

Aus Fairness gegenüber den bisher mittels AC ausgewählten Vertriebsleitern und aus Gründen der Vergleichbarkeit mit den bisherigen AC-Ergebnissen musste für diese Zielgruppe auf das schon bestehende Anforderungsprofil zurückgegriffen werden. Dazu war es erforderlich, zu den im alten System überprüften Verhaltensmerkmalen vergleichbare Verhaltensmerkmale zu entwickeln. Für die Zielgruppe der Vertriebsleiter wurde daher anhand der schon bestehenden Verhaltensbegriffe aus dem alten AC ei-

ne Parallele zu den Verhaltensmerkmalen des Beobachtungsinventars im oben beschriebenen Basis-Beobachtungsinventar mit seinen 20 Verhaltensmerkmalen hergestellt.

Zusammen mit sieben Geschäftsstellenleitern, drei Personalreferenten und vier Vertriebsleitern wurde zuerst in einem Workshop darüber befunden, ob Verhaltensmerkmale des Basis-Beobachtungsinventars benutzt werden könnten. Von den 20 Verhaltensmerkmalen wurden letztendlich 13 Verhaltensmerkmale übernommen. Dann wurden die Definitionen und die Operationalisierungen der Verhaltensmerkmale auf ihre Verständlichkeit hin überprüft; missverständliche Formulierungen wurden diskutiert und verändert.

Schließlich wurde das schon existierende Anforderungsprofil auf seine Gültigkeit hin untersucht; trotz widersprüchlicher Meinungen über dessen Stichhaltigkeit entschlossen sich die Verantwortlichen des Unternehmens aus politischen Gründen, dieses Anforderungsprofil beizubehalten.

### **3.3.2 Zielgruppenspezifisches Beobachtungsinventar und Anforderungsprofil für Versicherungsverkäufer**

Für die Versicherungsverkäufer konnte ein neues Anforderungsprofil konzipiert werden. Bezüglich der Verhaltensmerkmale wurde eine Anlehnung an das AC-System der Vertriebsleiter vorgenommen. Das Unternehmen versprach sich von diesem Vorgehen, zukünftige Vertriebsleiterpotenziale schon im Vorfeld leichter identifizieren zu können. Zudem war man im Unternehmen der Meinung, dass für eine erfolgreiche berufliche Tätigkeit dieselben Verhaltensmerkmale erforderlich seien, wenn auch in anderer Ausprägung.

Auf einem Workshop wurde daher vier Geschäftsstellenleitern, drei Vertriebsleitern und drei Verantwortlichen des Unternehmens das Beobachtungsinventar für die Vertriebsleiter vorgelegt und auf seine Stichhaltigkeit hinsichtlich der Zielgruppe "Versicherungsverkäufer" überprüft. Von den 13 Verhaltensmerkmalen wurden 12 Verhaltensmerkmale übernommen, lediglich das Verhaltensmerkmal "Zielorientierung" wurde weggelassen: per definitionem ist mit "Zielorientierung" das an Zielen orientierte Führen von Mitarbeitern gemeint, und das ist eine Aufgabe der Vertriebsleiter. Dieses Ver-

haltensmerkmal wurde ausschließlich in den AC gemessen, in denen Vertriebsleiter ausgewählt wurden.

Ferner wurde in diesem Workshop für die Zielgruppe der Versicherungsverkäufer ein Anforderungsprofil entwickelt. Dazu wurde mit den Anwesenden über die Aufgaben von Versicherungsverkäufern und die daraus resultierenden Anforderungen diskutiert, und die Stellenbeschreibung wurde zu Rate gezogen. Auf der Basis des dadurch gewonnenen gemeinsamen Verständnisses wurde dann den Workshopteilnehmern die Frage gestellt:

*"Wie stark muss ein Versicherungsverkäufer das Verhalten..... zeigen? Bitte schätzen Sie jedes Verhaltensmerkmal auf folgender Skala ein:..."*

<b>Beurteilungsskala</b>						
<b>1</b>	-----	<b>2</b>	-----	<b>3</b>	-----	<b>4</b>
nicht gefordert		manchmal gefordert		oft gefordert		sehr oft gefordert

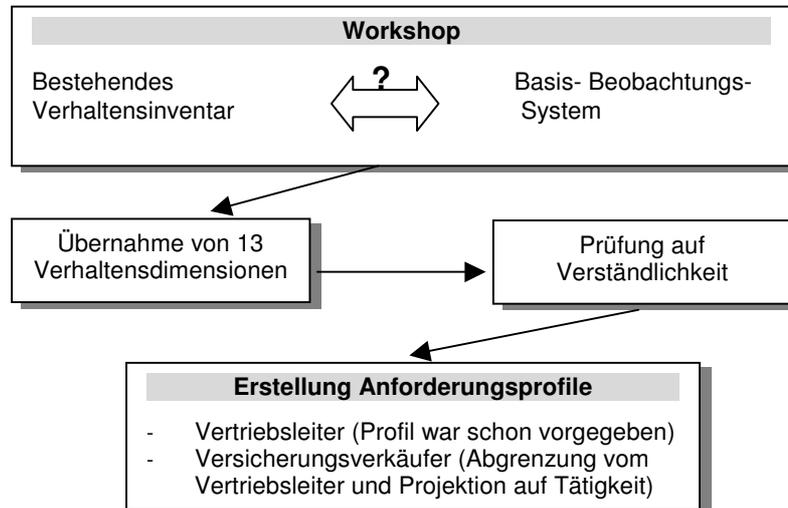
Jeder Workshopteilnehmer musste diese Frage einzeln für sich beantworten, ohne mit seinen Kollegen darüber zu diskutieren. Die Workshopteilnehmer benutzten bei ihrer Einschätzung die o.a. Skala. Sie konnten dabei auch Zwischenwerte vergeben, allerdings nur in Schritten von 0,5 Punkten (z.B.: 1,5), so dass rein numerisch eine siebenstufige Skala zur Erstellung des Anforderungsprofils zur Verfügung stand.

Dann wurden die individuell vergebenen Einschätzungen zusammengefasst. Das Gruppenergebnis wurde auf seine Schlüssigkeit hin überprüft, indem ein Zusammenhang mit der Stellenbeschreibung hergestellt wurde. Außerdem erfolgte ein Abgleich des Anforderungsprofils für Versicherungsverkäufer mit dem Anforderungsprofil der Vertriebsleiter. Das letztlich entwickelte Anforderungsprofil wurde dann von den Anwesenden einstimmig verabschiedet (siehe Anhang 3, graue Felder im Ergebnisprofil).

Nach seiner Genehmigung durch den Betriebsrat wurden außerdem die Geschäftsstellenleiter mit einem Rundschreiben über dieses Anforderungsprofil in Kenntnis ge-

setzt.

**Abb. 3: Erstellung von Beobachtungsinventar und Anforderungsprofil für Vertriebsleiter und Versicherungsverkäufer**



Wenn auch Schuler und Funke (1989) darauf hinweisen, dass ein Verzicht auf psychologische Konstrukte zu Gunsten einer unmittelbaren Zuordnung von Tätigkeits- und Aufgabenelementen die Gefahr in sich birgt, tatsächlich relevante Eignungsdimensionen nicht zu treffen, so scheint das oben skizzierte Vorgehen probat zu sein, um das Erfahrungswissen der Beobachter (und damit für den Berufserfolg der Teilnehmer relevante Aspekte) abzubilden und die Akzeptanz bei Beobachtern und Teilnehmern gegenüber dem AC-System zu fördern.

### 3.3.3 Die Konstruktion der Übungen

Die Übungen des AC lehnten sich an die Anforderungen der jeweiligen Zielposition, wie sie in den Stellenbeschreibungen festgehalten sind, und an die Anforderungen des Berufsalltags an. Dazu wurden Informationen aus dem Berufsleben von Vertriebsleitern und Versicherungsverkäufern, für die das AC konstruiert werden soll, gesammelt, und zusammen mit den Verantwortlichen des Unternehmens erstellt. Auf berufsfremde Thematiken, wie z.B. die bekannten Übungen der Encountergruppentradition (Wüstenübung etc.) wurde vollständig verzichtet, um so die soziale Akzeptanz des Verfahrens bei Beobachtern und Teilnehmern nicht zu gefährden.

Sowohl Vertriebsleiter wie auch Versicherungsverkäufer müssen sich vor Kunden selbst darstellen können, sie müssen im Team arbeiten, um gemeinsam neue Verkaufsstrategien zu entwickeln, sie müssen sich und ihre Meinung zur Diskussion stellen können und sie müssen im Kontakt mit Kolleginnen oder Kunden überzeugen. Somit konnte auf die Struktur von in der Literatur beschriebenen "klassischen" Übungen, wie Selbstpräsentation, Gruppenübung, Zweiergespräch und Stressinterview zurückgegriffen werden.

Die Inhalte der Übungen wurden auf Themen aus der zukünftigen spezifischen Berufsrealität abgestellt. Zu diesem Zweck wurden Expertenbefragungen mit Geschäftsstellenleitern, Vertriebsleitern und Versicherungsverkäufern mittels der Critical Incident Technique (vgl. Flanagan, 1954) durchgeführt. Die Interviews wurden ausgewertet, die Stellenbeschreibungen wurden hinzugezogen und darauf aufbauend wurden unter Einbezug von Mitarbeitern des Unternehmens Übungen konstruiert.

Für die zukünftigen Vertriebsleiter wurden Übungen mit fachspezifischem Hintergrund entwickelt, weil für diese Position nur die Bewerber in Frage kommen sollten, die aus der Versicherungsbranche stammten.

Für die zukünftigen Versicherungsverkäufer wurden Inhalte gewählt, die mit der "Berufswelt Versicherung" zu tun hatten, aber kein fachspezifisches Wissen erforderten, da die Bewerber für diese Position aus branchenfremden Berufen stammten. Für den Fall, dass trotzdem ein Branchenkenner unter den Bewerbern sein sollte, wurden für die Übungen "Zweiergespräch" und "Stressinterview" Alternativübungen mit fachspezifischen Thematiken verarbeitet.

Des Weiteren wurde anhand der Struktur der Übungen eine Zuordnung der Verhaltensmerkmale zu den einzelnen Übungen vorgenommen. Was bedeutet das? Es ist z.B. wenig sinnvoll, in einer Selbstpräsentation den Punkt "Teambildung unterstützen" zu beobachten, sehr wohl aber in der Gruppenübung. Dort wird bspw. von dem zukünftigen Vertriebsleiter gemäß seiner Stellenbeschreibung erwartet, dass er alle Gruppenmitglieder in den Gruppenprozess einbezieht. Ein anderes Beispiel: das Verhaltensmerkmal "Umgang mit belastenden Situationen" ist sehr wichtig. Es wird deshalb in Übungen beobachtet, von denen zu erwarten ist, dass sie in besonderem Maße Belastbarkeit erfordern, nämlich in der Selbstpräsentation und im Stressinterview, und

nicht in Übungen, in denen dieses Verhaltensmerkmal weniger bedeutsam ist (die Übersicht im Anhang 3 verdeutlicht die Verteilung der Verhaltensmerkmale auf die Übungen).

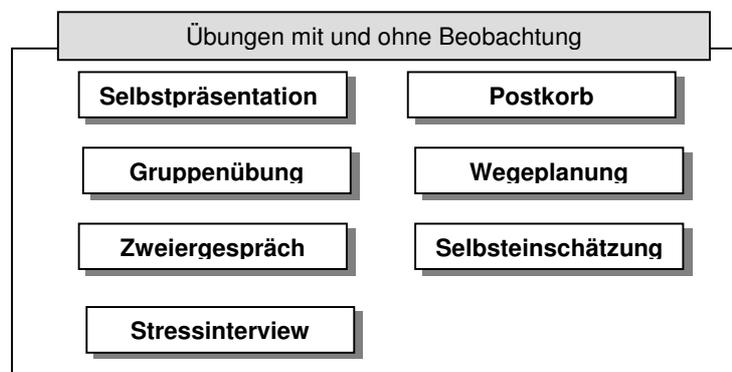
Ebenfalls wurde der Bedeutsamkeit der Verhaltensmerkmale gemäß Anforderungsprofil Rechnung getragen. Verhaltensmerkmale mit einer größeren Wichtigkeit sollten auch öfter beobachtet werden als weniger bedeutsame Verhaltensmerkmale. Zusätzlich wurden die Übungen mit Instruktionen für die Teilnehmer versehen (zur Gestaltung von Übungsmaterialien vgl. die Ausführungen von Sackett, 1987).

Neben den Übungen, in denen Verhalten zu beobachten war, wurden noch drei Übungen eingesetzt, in denen Verhaltensbeobachtung nicht möglich war: ein in Anlehnung an Jeserich (1981) entwickelter Postkorb, eine Wegeplanung, sowie eine Selbsteinschätzung der Teilnehmer bezüglich ihres am AC-Tag gezeigten Verhaltens.

Postkorb und Wegeplanung wurden aus Gründen des AC-Ablaufs eingesetzt. Zum Stressinterview wurden die Teilnehmer nämlich einzeln gerufen, zum Zweiergespräch zu zweit. Das bedeutet für die anderen Teilnehmer, Leerlauf zu haben. Um das zu vermeiden, bearbeiteten sie den Postkorb.

Die Wegeplanung und die Selbsteinschätzung fanden als letzte Übungen statt, wenn sich die Beobachter zur Beobachterkonferenz zurückzogen (s.u.).

**Abb. 4: Die Übungen im AC (vgl. Anhang 2)**



### **3.3.4 Beobachtermaterial und Teilnehmermaterial zu den Übungen**

Für die Beobachter und die Teilnehmer wurden unterschiedliche Begleitmaterialien zu den Übungen erstellt. Die Beobachter erhielten neben den Übungsinhalten auch noch eine Information über die zu beobachtenden Verhaltensmerkmale und deren Definitionen, sowie eine Aufklärung über den Übungstyp und die Übungsdauer (vgl. Anhang 2).

Die Teilnehmer erhielten lediglich das Material, das sie benötigten, um die Übung bewältigen zu können. Dazu gehören neben den Übungsinhalten auch Instruktionen über die genaue Aufgabenstellung sowie das zur Verfügung stehende Zeitbudget (vgl. Anhang 2).

### **3.3.5 Die Beobachterbögen**

Bei den Beobachterbögen wurde darauf geachtet, dass nicht mehr als sieben unterschiedliche Verhaltensmerkmale beobachtet wurden, um die Beobachter nicht zu stark zu belasten.

Aus dem Pool der Operationalisierungen der Verhaltensmerkmale wurden daher für jedes Verhaltensmerkmal die zwei Operationalisierungen ausgewählt, die auf den Übungsinhalt und das zu erwartende Verhalten möglichst gut passten. Auf den Beobachterbögen waren die Oberbegriffe und die Operationalisierungen, die beurteilt werden mussten, zu sehen (vgl. Anhang 2).

Die Anzahl der zu beobachtenden vorgegebenen Operationalisierungen je Verhaltensmerkmal wurde auf zwei begrenzt, weil bspw. bei maximal sieben zu beobachtenden Verhaltensmerkmalen in der Gruppenübung des Vertriebsleiter-AC immerhin noch 14 Beurteilungen abgegeben werden mussten. Geht man davon aus, dass jede Beurteilung mindestens einmal korrigiert wird, hat man wenigstens 28 Beurteilungsprozesse – ohne die Entscheidungsprozesse zu berücksichtigen, die zu jeder einzelnen Beurteilung führen. Übungen und Beobachterbögen wurden nach ihrer Erstellung den Verantwortlichen des Unternehmens und dem Betriebsrat präsentiert, modifiziert und schließlich nach deren Zustimmung eingesetzt.

### **3.4 Die Beobachter**

#### **3.4.1 Die unternehmensinternen Beobachter**

Beobachter für die Vertriebsleiter-AC waren ausschließlich Geschäftsstellenleiter, und zwar in erster Linie die Geschäftsstellenleiter, die einen Bewerber zum AC entsandt hatten.

Die Beteiligung der direkten Vorgesetzten am Auswahlverfahren widerspricht dem Vorschlag von Jeserich (1981), entspricht aber der Beratungsphilosophie des AC-Konstrukteurs. Zum einen gewinnt der Vorgesetzte zusätzliche Eindrücke zu den Informationen, die er über seinen Bewerber schon gesammelt hat. Diese Eindrücke bildet er sich selber. Er erfährt zum anderen aber auch die Einschätzung seiner Kollegen, was zur Abrundung seines Bildes vom Bewerber wesentlich beiträgt. Beobachterkollegen und AC-Konstrukteur können zudem im Sinne eines Vorgesetztencoaching wertvolle Hinweise auf die zukünftige Führungsarbeit für den Bewerber geben.

Andere Geschäftsstellenleiter, die als Beobachter fungierten, wollten die Beobachtererfahrung im AC nutzen, um sich für den Personalauswahlprozess mehr zu sensibilisieren. Ihre Teilnahme war freiwillig – sie wurden eingeladen oder fragten selber an, ob sie Beobachter bei einem der nächsten AC sein könnten.

Bei den Verkäufer-AC waren immer ein bis zwei Geschäftsstellenleiter und ansonsten Vertriebsleiter (auch hier vorzugsweise die potenziellen Vorgesetzten der Bewerber) anwesend. Geschäftsstellenleiter und Vertriebsleiter waren allesamt Experten bezüglich der Auswahl von Mitarbeitern, da die Personalauswahl aufgrund der branchenüblich hohen Fluktuation ein wesentlicher Bestandteil ihrer Tätigkeit ist. Wegen der längeren Berufstätigkeit waren die Geschäftsstellenleiter jedoch darin erfahrener als die Vertriebsleiter. Bei ca. der Hälfte der AC ( $n=18$ ) fungierten auch Mitarbeiterinnen (in neun Fällen jeweils zwei Mitarbeiterinnen, in neun Fällen eine Mitarbeiterin) der Zentrale, die den Außendienst personalwirtschaftlich betreuten, als Beobachter. Sie waren neben der AC-Assistenz die einzigen Frauen unter den Beobachtern.

Bezüglich der Beobachter gibt es keine demografischen Daten. Es kann nur berichtet werden, dass Geschäftsstellenleiter allesamt erfolgreiche Vertriebsmitarbeiter des Unternehmens waren, ehe ihnen die Leitung einer Geschäftsstelle übertragen wurde. Die

beobachtenden Vertriebsleiter hatten Erfahrung in der Führung und Ausbildung von Aussendienstmitarbeitern.

Die Integration der direkten Vorgesetzten hat nicht dazu geführt, dass Beobachter "ihren Bewerber" unbedingt "durchbringen" wollten. Dieses Phänomen ist überwiegend die Folge von wenig transparenten (z.B. bezüglich des Wegs der Ergebnisfindung) und sozial nicht akzeptierten AC-Systemen. Die Reaktion der Beobachter, ihren Teilnehmer "schützen" (also durchbringen zu wollen) wäre somit als Reaktanz gegenüber den AC-Ergebnissen zu interpretieren.

Andererseits können auch interpersonale Konflikte dieses Phänomen erklären. Beobachter konkurrieren untereinander, wollen gut dastehen etc. Möglicherweise resultiert ein derartiges Verhalten auch aus einer Protesthaltung gegenüber dem AC-Leiter, wenn dieser den Weg der Ergebnisfindung zu sehr "verpsychologisiert", sich nicht auf seine Rolle als Moderator des AC beschränkte und dadurch bei den Beobachtern Widerstand erzeugte.

Bei allen für diese Untersuchung relevanten AC konnten derartige Vorkommnisse nicht beobachtet werden. Im Gegenteil: die Geschäftsstellenleiter und die Vertriebsleiter waren dankbar, ihren Bewerber einmal in anderen Zusammenhängen erleben zu dürfen und waren sehr offen für die Anregungen ihrer Kollegen.

Damit decken sich die Beobachtungen mit den Ergebnissen von Sackett und Wilson (1982), wonach derartige Einflussnahmen bei Beobachtern aus dem mittleren Management kaum vorkamen.

Sieben Geschäftsstellenleiter und vier Vertriebsleiter waren an der Entwicklung des AC-Systems beteiligt.

### **3.4.2 Die AC-Assistenz**

AC-Assistenz in 25 AC waren eine Psychologiestudentin (24 Jahre) und in 11 AC eine Betriebswirtin (25 Jahre). Beide hatten vor ihrem Studium bereits eine Berufsausbildung (Industriekauffrau und Bankkauffrau) abgeschlossen und sie hatten schon vor den für diese Untersuchung relevanten AC mit dem AC-Leiter zusammen gearbeitet,

kannten seine AC-Philosophie und sind von ihm ausgebildet worden.

### **3.4.3 Die Aufgaben von AC-Leiter, AC-Assistenz, Experten und Novizen**

Die Beobachter beobachteten die Teilnehmer während der Übungen und führten die Beobachterbögen. In der Selbstpräsentation und im Stressinterview beobachteten sie jeden einzelnen Teilnehmer, in den Gruppenübungen nur die vier Teilnehmer, die sie am besten sehen konnten. Dieses Vorgehen wurde gewählt, um die Aufmerksamkeit der Beobachter durch eine zu große Zahl an zu beurteilenden Teilnehmern nicht unnötig zu beanspruchen. Im Zweiergespräch beobachteten sie beide Gesprächspartner gleichzeitig.

Während des AC hatten die Beobachter nur eine Aufgabe: das Verhalten der Teilnehmer still zu beobachten. Von diesem Grundsatz wich das Verhalten der Beobachter im Stressinterview ab. Im Stressinterview beobachteten die Beobachter nicht nur das Verhalten der Teilnehmer, sie diskutierten auch mit ihnen, mussten also teilnehmend beobachten.

Die Beobachter gaben den Teilnehmern zwischen den Übungen keine Rückmeldung über das von ihnen beobachtete Verhalten. In den Pausen mischten sie sich als Ansprechpartner unter die Teilnehmer, sprachen mit ihnen aber nicht über das in den Übungen beobachtete Verhalten.

Sofern es ihnen möglich war, hielten die Beobachter in allen Übungen auch formfreie Bemerkungen fest, die später in den Beobachterkonferenzen (s.u.), falls notwendig, zu Rate gezogen werden konnten.

Die AC-Assistenz betreute die Beobachter, indem sie während des AC Fragen klärte und die Beobachter mit Unterlagen versorgte. Sie war auch für die Übertragung der Daten aus den Beobachterbögen und deren Verrechnung mittels EDV verantwortlich (zu diesem Zweck kam ein eigens erstelltes Datenverarbeitungstool zum Einsatz). Von ihrem Status her galt sie als normaler Beobachter und wurde in den Beurteilungsprozess vollkommen integriert.

Der AC-Leiter war Ansprechpartner für die Teilnehmer. Er beantwortete alle während des AC aufkommenden Fragen und war bemüht, unnötige Anspannung zu vermeiden. Das geschah z.B. dadurch, dass nach jeder einzelnen Übung den Teilnehmern die Möglichkeit gegeben wurde, sich zum Inhalt und zum Sinn der Übung zu äußern. Er moderierte die einzelnen Übungen an, teilte den Teilnehmern das Übungsmaterial aus, sorgte für die Einhaltung des Zeitplans und leitete die Beobachterkonferenzen.

### **3.5 Die AC-Teilnehmer**

Die AC-Teilnehmer für die Vertriebsleiterposition kamen allesamt aus der Versicherungsbranche. Sie wurden entweder über den persönlichen Kontakt oder über Anzeigen in Zeitungen auf das Unternehmen aufmerksam gemacht. Ihre bisherigen Tätigkeiten deckten sich überwiegend mit ihren zukünftigen Tätigkeiten. Alle Bewerber hatten schon Personalverantwortung gehabt. Ihr Alter schwankte zwischen 25 und 55 Jahren. Unter den Bewerbern für die Vertriebsleiterposition (n = 87) waren lediglich zwei Frauen.

Die potenziellen Versicherungsverkäufer (n = 180, davon 41 Frauen) wurden ebenfalls wie die Vertriebsleiter rekrutiert. Sie stammten bis auf wenige Ausnahmen aus branchenfremden Berufen. Ihr Alter schwankte zwischen 19 und 39 Jahren.

Weitere demografische Daten können über die Teilnehmer nicht berichtet werden, da die Bewerbungsunterlagen aus Gründen des Datenschutzes im Unternehmen verblieben und dem das AC durchführenden Dienstleister nicht zugänglich gemacht wurden.

### **3.6 Die Durchführung des AC**

#### **3.6.1 Die Instruktion der Beobachter**

Der vorliegende AC-Typ verzichtet gänzlich auf ein Beobachtertraining im herkömmlichen Sinne: es gibt lediglich eine Instruktion für die Beobachter. Die Beobachter wurden dabei in Form eines halbtägigen Workshops (zum Ablauf des AC vgl. Anhang 3) mit folgendem vertraut gemacht:

- der Vorgeschichte, die zum zu Stande kommen des Auftrags führte
  - Sinn und Zweck des AC in ihrem Unternehmen
  - dem Verfahren "AC" und seinen Besonderheiten
  - der AC-Philosophie des durchführenden Dienstleisters
  - den Übungen des AC
  - ihrer Aufgabe als Beobachter
- sowie dem Ablauf des AC-Tages als solches.

Danach wurde mit den Beobachtern zusammen ein "naives" zielgruppenspezifisches Profil von Verhaltensmerkmalen entwickelt, indem mit ihnen die Frage diskutiert wurde wurde:

*"Was glauben sie, muss ein Vertriebsleiter / Versicherungsverkäufer für Verhaltensweisen besitzen, um im Beruf erfolgreich sein zu können?"*

Diese Verhaltensmerkmale wurden den Verhaltensmerkmalen des offiziellen Anforderungsprofils gegenübergestellt, Gemeinsamkeiten und Unterschiede wurden diskutiert. Unterschiede gab es dabei in der Regel lediglich in der Formulierung der Verhaltensbegriffe, also im Sprachgebrauch, nie dagegen im Verständnis in bezug auf die Verhaltensmerkmale.

Des weiteren wurden den Beobachtern das Beurteilungssystem, der Gebrauch der Beobachterbögen und der Weg der Ergebnisfindung erläutert, die Übungen wurden durchgesprochen und die Aufgaben der Beobachter, der AC-Assistenz und des AC-Leiters dargestellt.

Außerdem hatten die Beobachter noch die Möglichkeit, sich untereinander über ihre Erfahrungen mit dem "Instrument AC" und anderen Verfahren der Personalauswahl auszutauschen, und deren Vor- und Nachteile zu diskutieren. Der Vorteil bei diesem Vorgehen besteht darin, seitens der Beobachter schon im Vorfeld mögliche Widerstände abzubauen und deren Identifikation mit dem Verfahren positiv beeinflussen zu können. Um Teilnehmern gegenüber keine Vorurteile aufkommen zu lassen, wurden deren Bewerbungsunterlagen von den Beobachtern (ausgenommen natürlich der potenziellen Vorgesetzten), der AC-Assistenz und dem AC-Leiter nicht eingesehen. Erst in der Be-

obachterkonferenz, in der über Annahme oder Ablehnung der Teilnehmer befunden wurde, erfolgten selten Rückgriffe auf die Bewerbungsunterlagen, um sich von den Bewerbern ein noch genaueres Bild machen zu können.

### **3.6.2 Die Instruktion der Teilnehmer**

Die Teilnehmer reisten am Vorabend des AC an und wurden nach der Begrüßung und einer kurzen Präsentation des Unternehmens ebenfalls mit dem Ablauf des AC vertraut gemacht. Im Prinzip ähnelte die Teilnehmerinstruktion dem Vorgehen bei der Beobachterinstruktion. Die Teilnehmer wurden jedoch nicht über die Inhalte der Übungen, sondern nur über deren Struktur informiert. Um Ängste zu nehmen, wurden Ihnen die Beobachterbögen aller Übungen zur Begutachtung überlassen.

Die Verhaltensmerkmale des Anforderungsprofils sowie das Anforderungsprofil selber wurden ihnen vorgestellt und mit ihnen diskutiert. Dieses Vorgehen orientiert sich an der Transparenzforderung von Kleinmann (1997), der nachwies, dass AC-Teilnehmer in intransparenten AC-Situationen dann besonders gut abschnitten, wenn sie die relevanten Verhaltensmerkmale erkannten (Verbesserung der Konstruktvalidität). Tendenziell wurden diese Befunde von Kolk (2001) bestätigt. Während Kleinmann (1997) jedoch problematisiert, dass bei einem Offenlegen der Anforderungsdimensionen die Kriteriumsvalidität sinkt, und daher ein derartiges Vorgehen vom Anwendungszweck des AC (Selektion vs. Personalentwicklung) abhängig gemacht werden sollte, legen andere Befunde nahe, aus Gründen der Sozialen Akzeptanz des AC (Schuler und Stehle (1983) sprechen von Sozialer Validität) seitens der Teilnehmer und der Beobachter eine Offenlegung der Anforderungsdimensionen zu favorisieren.

Nicht zuletzt für die in einem Personalauswahl-AC "durchgefallenen" Teilnehmer scheint dieses Vorgehen besonders wichtig zu sein, um einen positiven Eindruck vom Auswahlprozess zu bekommen (vgl. Randhofer, 1995).

Die Teilnehmerinstruktion war zeitlich nicht limitiert, dauerte in der Regel aber nicht länger als zwei bis drei Stunden. Alle Teilnehmer hatten schon im Vorfeld ein vom AC-Leiter erstelltes schriftliches Informationsmaterial als Vorbereitung auf den AC -Tag durch ihre potentiellen Vorgesetzten bekommen.

### **3.6.3 Der Ablauf des eigentlichen AC-Tages**

#### **3.6.3.1 Ablauf der Übungen**

Der AC-Tag begann mit der Selbstpräsentation der Kandidaten. Sie wussten von der Information am Vorabend, dass dies die erste Übung sein würde. Die Selbstpräsentation wurde als erste Übung gewählt, weil die Teilnehmer sich zum einen besser kennenlernen konnten, zum anderen aber auch die Scheu vor dem Setting verlieren sollten: sie saßen in der Mitte des Raumes an einem rechteckigen Tisch, die Beobachter saßen paarweise an Einzeltischen auf die Ecken des Raumes verteilt, so dass sie die Teilnehmergruppe gut beobachten konnten. Im AC gab es maximal neun Teilnehmer und acht Beobachter.

Um sich auf diese Übung vorzubereiten hatte die Teilnehmer 10 Minuten lang Zeit und durften sich dann fünf Minuten lang selbst präsentieren.

Die zweite Übung war eine Gruppenübung mit Konkurrenzcharakter. Hier ging es darum, die eigene Meinung gegenüber den anderen Teilnehmern zu vertreten und sie für diese Meinung zu gewinnen. Diese Gruppenübung wurde gewählt, um Gruppenbildungsprozesse zu fördern und um die Differenzierung einzelner Rollen in der Gruppe zu ermöglichen. Der Konkurrenzcharakter der Übung sollte zusätzlich dazu beitragen. Diese Übung dauerte insgesamt 60 Minuten.

Nach der zweiten Übung verliessen die Teilnehmer den Raum zu einer kleinen Pause. In dieser Zeit schilderten zuerst die Beobachter, dann die AC-Assistenz und schließlich der AC-Leiter (in der Reihenfolge) in Form eines kurzen Blitzlichts ihre Eindrücke aus den ersten beiden Übungen, ohne dabei einzelne Personen zu benennen. Ihre Schilderungen bezogen sich nur auf den Gruppenprozess; persönliche Wertungen und/oder Einschätzungen zu einzelnen Teilnehmern unterblieben.

In keiner Phase des AC bekamen Beobachter von der AC-Assistenz oder dem AC-Leiter eine Rückmeldung zu ihren Beobachtungen, oder bekamen Gelegenheit, über ihre Beobachtungen untereinander zu sprechen. Die Beobachter sollten nicht in ihrer ihnen typischen Wahrnehmung normiert werden, sondern sollten mit ihrem individuellen Wertemaßstab anhand des eindeutig interpretierbaren Beobachtungsinventars die Teilnehmer beobachten (zur Vermeidung von Inter-Assessor-Einflüssen vgl. Lewry,

(1993). Exler et al. (1997), wiesen nach, dass bei einer Zusammensetzung der Beobachtergruppe aus Vorgesetzten und Mitarbeitern die Mitarbeiter vor allem dann den Vorgesetzten ähnlichere Beurteilungen abgaben, wenn sie in einem Fragebogen zur Messung der "Sozialen Erwünschtheit" hohe Werte hatten und wenn ihnen im Vorfeld ein Feedback zu ihren Beurteilungen über ihre Vorgesetzten angekündigt wurde).

In der vorliegenden Untersuchung waren bei einem Teil der AC zur Auswahl von Versicherungsmitarbeitern Geschäftsstellenleiter und deren Mitarbeiter gleichermaßen als Beobachter tätig; in einem anderen Teil der AC waren ausschließlich Geschäftsstellenleiter die Beobachter. In beiden Fällen jedoch gab es starke Bemühungen, sich sozial wünschenswert zu zeigen: einmal motiviert durch den Gedanken, ein guter Mitarbeiter sein zu wollen, zum anderen getrieben von dem Bemühen, sich vor dem Kollegen keine Blöße zu geben. Das Verhindern von Feedback oder von Gesprächen über die Teilnehmer sollten den von Exler et al. (1997) beschriebenen Effekt vermeiden.

Die Teilnehmer hatten währenddessen eine Pause von ca. 20 Minuten. Vor der Pause wurden sie darüber informiert, dass die nächste Übung das Stressinterview sein würde. Zum Stressinterview wurde die Teilnehmergruppe aus zeitökonomischen Gründen in zwei Gruppen geteilt (s.u.). Die Gruppen wurden von der AC-Assistenz oder dem AC-Leiter moderiert. Die Reihenfolge, in der die Teilnehmer zum Stressinterview gerufen wurden, wurde von der jeweiligen Beobachtergruppe festgelegt, wobei darauf geachtet wurde, dass die Teilnehmer, die in den beiden ersten Übungen einen weniger gut belastbaren Eindruck gemacht hatten, zuerst befragt wurden.

In dieser Übung wurde den Teilnehmern ein konflikthaltiger Sachverhalt in schriftlicher Form vorgelegt, den sie sich durchlesen sollten, um dann nach ihrer Meinung zu dem Sachverhalt durch die anwesenden Beobachtern befragt zu werden.

Das Stressinterview wurde zu diesem Zeitpunkt gewählt, weil die Tageszeit in etwa mit der Spitze der physiologischen Leistungskurve zusammenfällt, und diese Übung im allgemeinen als besonders belastend erlebt wird. Die Dauer des Stressinterviews betrug inklusive der Vorbereitungszeit des Teilnehmer ca. 25 Minuten.

Das Stressinterview wurde immer von der AC-Assistenz, bzw. dem AC-Leiter eröffnet und auch beendet. Sollten während des Stressinterviews von der AC-Assistenz, bzw. dem AC-Leiter Belastungsgrenzen bei den Teilnehmern erkannt, änderten diese ihren konfrontativen Befragungsstil in Richtung auf einen unterstützender wohlwollenden Be-

fragungsstil. Die Beobachter waren angehalten, sich diesem Verhalten anzupassen; im Falle einer Mißachtung schritten AC-Assistenz oder AC-Leiter unmittelbar gegen den betroffenen Beobachter ein und pufferten so ungewollte Härten ab.

Nach dem Stressinterview folgte eine Pause in Form des Mittagessens, und danach fanden Zweiergespräche statt. In dieser Übung mussten sich zwei Teilnehmer zu versicherungstypischen Aufgabenstellungen miteinander unterhalten, und zwar aus der Rolle des Kunden oder aus der Rolle des Versicherungsmitarbeiters heraus. Die Gesprächspaarungen und die Rollenverteilungen wurden vom AC-Leiter und von der AC-Assistenz festgelegt. Dabei wurde darauf geachtet, möglichst Paarungen mit gleich starken Gesprächspartnern zusammenzustellen, um weitestgehend Fairness für die Teilnehmer anzustreben. Für die Rollenzuweisung gab es keine Zuordnungskriterien; hier entschied das Los.

Die Gespräche dauerten ca. 10 Minuten, ohne die Zeit, die die Gesprächspartner benötigten, um sich auf ihre Gesprächsrolle vorzubereiten: den Teilnehmern wurden nämlich kurze Rollenvorgaben und Instruktionen gegeben, mit der Aufforderung, sich in die geschilderte Rolle hineinzusetzen und aus dieser Position heraus das später stattfindende Gespräch zu führen. Den Teilnehmer, mit dem sie das Gespräch führen sollten, kannten sie bis zu dem Zeitpunkt, an dem beide zum Gespräch gerufen wurden, nicht. Die Rollenvorgabe ihres Gesprächspartners kannten sie ebenfalls nicht. Erst nach dem Gespräch wurden zuerst die Gesprächspartner zu ihrer Einschätzung des Gesprächsverlaufs befragt; dabei wurden auch die jeweiligen Rollenvorgaben vorgelesen, um die Gesprächseinschätzung für die Gesprächsteilnehmer zu erleichtern.

Eine selbstkritische Anmerkung sei an dieser Stelle gestattet: gemäß den Qualitätskriterien für AC, herausgegeben vom Arbeitskreis Assessment Center e.V. (1996), soll vermieden werden, dass Teilnehmer in Rollenspielsituationen aufeinandertreffen. Nach der damaligen Einschätzung war das direkte Aufeinandertreffen von Teilnehmern jedoch vertretbar, da es bei den AC ja darum ging, möglichst viele Teilnehmer als Potenzialträger zu identifizieren und nicht nur einen. Insofern wäre es naheliegend, kein Konkurrenzverhalten unter den Teilnehmern zu erwarten.

Die Frage ist allerdings, ob die Teilnehmer vor und während des AC dieser Aussage Glauben schenkten. Es zeigte sich, daß sich die Teilnehmer untereinander eher kom-

petitiv verhielten - möglicherweise, um die vermuteten Erwartungen des ebenfalls anwesenden potentiellen Vorgesetzten zu erfüllen. Es ist daher nur schwer abzuschätzen, inwiefern angesichts der o.a. Punkte "weitgehend normales Verhalten" überhaupt beobachtet und beurteilt werden konnte. Zudem ist die Vergleichbarkeit der einzelnen Gesprächsverläufe wegen der großen Varianz, hervorgerufen durch die unterschiedlichen Gesprächspaarung, nicht gegeben.

Es wurde zwar darauf Wert gelegt, dass ähnlich starke Gesprächspartner aufeinandertreffen sollten, diese Ähnlichkeit wurde aber nicht weiter objektiv erhoben. Außerdem wurden auch unterschiedliche Gesprächsinhalte in den einzelnen Zweiergesprächen bearbeitet, so dass auch hier die Vergleichbarkeit eingeschränkt war. Vor diesem Hintergrund würde ein Gesprächspartner mit einem stereotypen Gesprächsverhalten zu einem einheitlich für alle Teilnehmer geltenden Gesprächsthema, z.B. ein Beobachter oder ein Schauspieler, helfen, dieses Problem zu beseitigen.

Im Anschluß an die Zweiergespräche fand nochmals eine Gruppenübung statt. Diesmal handelte es sich um eine Übung ohne Konkurrenzcharakter, um den AC- Tag harmonisch ausklingen zu lassen. Auch diese Gruppenübung dauerte, wie die erste Gruppenübung, ca. 60 Minuten.

Danach mussten sich die Teilnehmer bezüglich ihres am AC-Tag gezeigten Verhaltens selbst anhand des kompletten Beobachtungsinventars einschätzen. Der Haupteffekt dieser Übung bestand darin, durch die Einsicht in das Beobachtungsinventar bei den Teilnehmern zusätzlich Akzeptanz gegenüber dem Verfahren zu fördern und sich auf das kommende Rückmeldegespräch vorzubereiten.

### **3.6.3.2 Die erste Beobachterkonferenz**

Nach dieser Übung zogen sich die Beobachter zurück, um über die Akzeptanz oder die Ablehnung der Teilnehmer zu beraten. Insgesamt 1 1/4 Stunden waren für diesen Prozess vorgesehen. Währenddessen bearbeiteten die Teilnehmer, um nicht unnötigen Leerlauf zu haben, eine Wegeplanung. Auch die Wegeplanung wurde, wie schon der Postkorb, nicht weiter berücksichtigt. Auf besondern Wunsch der Teilnehmer bekamen diese jedoch eine Rückmeldung über den richtigen Lösungsweg.

### **3.6.3.3 Die Rückmeldegespräche**

Nach der Beobachterkonferenz und vor den Rückmeldegesprächen hatten die Teilnehmer noch einmal in der Gruppe die Möglichkeit, sich zum Verlauf des Tages in Form eines Blitzlichtes zu äußern; danach wurden sie einzeln zu den Rückmeldegesprächen gebeten. Im Rückmeldegespräch wurden sie zu ihrer Einschätzung bezüglich des AC-Tages befragt, dann wurde ihnen die Einschätzung der Beobachterkonferenz (siehe Anhang 3) eröffnet.

Die Gespräche mit den abgelehnten Teilnehmern wurden prinzipiell vom AC-Leiter und der AC-Assistenz, zusammen mit einem Unternehmensangehörigen geführt, um den subjektiv erlebten Mißerfolg psychologisch auffangen zu können. Alle anderen Gespräche wurden immer von zwei Beobachtern gemeinsam durchgeführt, um sich anschließende Fragen der Teilnehmer über den weiteren Gang der Dinge aus Unternehmenssicht beantworten zu können.

Der Charakter der Rückmeldegespräche war vom Gedanken geprägt, in der Rückmeldung auch Entwicklungsmöglichkeiten oder Entwicklungsbedarf aufzuzeigen (s.o.). Sie dauerten ca. 15-20 Minuten.

### **3.6.3.4 Die abschließende Beobachterkonferenz**

Nach den Rückmeldegesprächen fanden sich die Beobachter zur abschließenden Beobachterkonferenz zusammen. Hier wurde für die Teilnehmer, die eine positive Empfehlung bekommen hatten, durch die Diskussion unter den Beobachtern ein Ist-Profil über alle Verhaltensmerkmale erstellt und mit dem Anforderungsprofil verglichen.

## **3.7 Der Weg der Ergebnisfindung in den einzelnen Phasen des AC**

Die Ergebnisfindung ist eng verbunden mit der Philosophie des jeweiligen AC-Konstrukteurs. Es können Entscheidungen über die Teilnehmer durch die rein numerische Auswertung der Beobachterbögen getroffen werden. Dann gibt es AC, bei denen keine Beobachterbögen vorliegen – dort müssen am Ende des AC-Tages die Ergebnisse in einer zeitaufwändigen Beobachterkonferenz erarbeitet werden.

### 3.7.1 Datenerhebung während der Übungen

Mit jeder einzelnen Übung des AC erhielten die Beobachter Beobachterbögen, die sie während des gesamten AC behalten durften. Auf den Beobachterbögen beurteilten sie für jede Operationalisierung eines Verhaltensmerkmals, wie ausgeprägt der Teilnehmer das gewünschte Verhalten gezeigt hat. Die Beurteilungen nahmen sie während der laufenden Übung vor.

Abgeleitet wurde dieses Vorgehen aus der Untersuchung von Murphy et al. (1982), die nachweisen konnten, dass unter den Bedingungen "sofortige Beurteilung" vs. "zeitlich verzögerte Beurteilung" bei der Bedingung "zeitlich verzögerte Beurteilung" die Höhe der Korrelationen der Beobachterurteile ansteigt. Die Autoren deutet das als Hinweis auf Inferenzen, die den Urteilsprozeß beeinflussen – und dadurch möglicherweise einen Halo-Effekt fördern.

Die Beobachter benutzten die bekannte Skala von 1-4 (s.o.), durften aber nur ganzzahlige Werte vergeben. Dieses Vorgehen wurde gewählt, um die Beobachter durch einen zu starken Differenzierungszwang, der durch eine breiter gefächerte Beurteilungsskala ausgelöst wird (es müsste z.B. entschieden werden, was den Unterschied zwischen 2 und 2,5 ausmacht etc.) nicht unnötig zu belasten.

Um einmal gefällte Beurteilungen korrigieren zu können nutzten die Beobachter für ihre Aufzeichnungen Radiergummi und Bleistift.

Ferner berechneten die Beobachter für jeden Beobachterbogen den Gesamtdurchschnitt. Alle Gesamtdurchschnitte für jeden Beobachterbogen über alle Beobachter hinweg wurden in ein eigens für die Verwertung dieser Daten erstelltes Computerprogramm eingegeben; auf diese Lösung wurde zurückgegriffen, weil die Dateneingabe über Tastatur und die automatisierte Auswertung unter Zeitdruck weniger Fehlern unterworfen ist als die Datenübertragung und die Datenverrechnung von Hand.

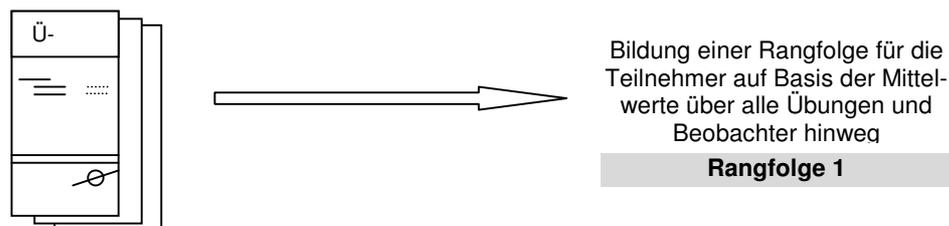
Somit lagen zu Beginn der Beobachterkonferenz die Mittelwerte für jeden Teilnehmer über alle Verhaltensmerkmale und über alle Beobachter hinweg vor. Die Bildung von Mittelwerten geht auf Cooper (1981) zurück, der vorschlug, die Urteile mehrere Beobachter in Form von Mittelwerten zusammenzufassen, um Wahrnehmungstendenzen auszumitteln.

### 3.7.2 Ergebnisfindung in der ersten Beobachterkonferenz

#### 3.7.2.1 Rangfolge 1: Berechnung von Mittelwerten

Zu Beginn der ersten Beobachterkonferenz wurde anhand dieser Mittelwerte eine Rangfolge über die Teilnehmer erstellt. Sie wurde den Beobachtern zunächst nicht mitgeteilt. Grundlage für diese Rangfolge waren die auf die Verhaltensmerkmale bezogenen auf den Beobachterbögen festgehaltenen Urteile der Beobachter über die jeweiligen von ihnen beobachteten Teilnehmer in den einzelnen Übungen. Die AC-Assistenz sammelte nach den Übungen die Beobachterbögen ein und verarbeitet die Beobachterurteile.

**Abb. 5: Rangfolge über die Teilnehmer anhand von Mittelwerten**



#### 3.7.2.2 Rangfolge 2: Die analytische und die emotionale Sichtweise

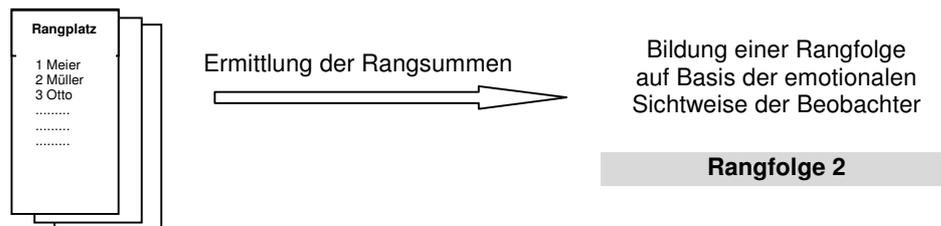
Neben der oben beschriebenen mechanischen Datenverarbeitung wurde auch die emotionale Sichtweise der Beobachter in bezug auf die Teilnehmer berücksichtigt.

Die Beobachter mussten am Beginn der ersten Beobachterkonferenz schriftlich und einzeln eine Rangfolge über die Teilnehmer erstellen, ohne ihre Beobachterbögen als Hilfestellung zu nutzen und ohne die Rangfolge über die Teilnehmer zu kennen. Es handelt sich also um eine Rangfolge, erstellt durch den Gesamteindruck von den Teilnehmern, mehr oder weniger emotional geprägt, und ohne Kenntnis von der ersten Rangfolge zu haben.

Diese Rangfolgen wurde dann einzeln verlesen und zusammengefasst. Pro Teilneh-

mer wurden dann die von den Beobachtern vergebenen Rangplätze addiert; der Teilnehmer mit der geringsten Rangsumme erhielt Rangplatz 1 etc.

**Abb. 6: Rangfolge über die Teilnehmer anhand des Gesamteindrucks**



Voraussetzung für den Erfolg eines derartigen Vorgehen ist allerdings, dass die Beobachter nach den einzelnen Übungen keinerlei Rückmeldung über die von ihnen gemachten Beobachtungen erhalten, und auch untereinander keine Diskussionen über die Teilnehmer führen.

### 3.7.2.3 Vergleich von Rangfolge 1 und Rangfolge 2 zur Ergebnisfindung

Beide Rangfolgen wurden miteinander verglichen, Übereinstimmungen und Unterschiede wurden diskutiert. Für die Personalentscheidung beim vorliegenden AC-Typ kann in dieser Phase wichtig sein, wie nah der durch die Verhaltensbeobachtung zu Stande gekommene Gesamtmittelwert des einzelnen Teilnehmers bei dem Gesamtmittelwert nach dem Anforderungsprofil liegt; das ist jedoch nicht das ausschlaggebende Kriterium. Entscheidend ist vielmehr, wie die unterschiedlichen Rangplätze der Teilnehmer zu erklären sind. Beispielsweise kann ein Teilnehmer den Rangplatz eins bei der mechanischen Datenverarbeitung und den Rangplatz acht bei der emotionalen Betrachtung haben. In dem Fall verhält sich dieser Teilnehmer technisch gesehen optimal, produziert aber bei den Beobachtern Widerstände, die zu einer entsprechenden Herabstufung führen (die Beobachter stützen sich dann auf Verhaltensmerkmale, die quasi unabhängig vom Beobachtungsinventar registriert werden).

Vor allem die Beobachter, die in ihrer individuellen Rangfolge eine sehr unterschiedliche Sichtweise von einzelnen Teilnehmern hatten (im Extremfall kann ein Beobachter einen Teilnehmer als Nummer eins einschätzen, während ein anderer Beobachter die-

sen Teilnehmer als Nummer neun einschätzt) wurden gebeten, ihre Einschätzungen zu begründen und miteinander zu diskutieren.

Auf der Grundlage dieser Diskussion wurden dann die Empfehlungen für die Teilnehmer ausgesprochen. Empfehlungen bedeutet im vorliegenden Fall: Teilnehmer ist ohne Einschränkungen, mit Einschränkungen oder nicht für die Position geeignet. Alle Empfehlungen wurden begründet, die Begründungen wurden schriftlich festgehalten und dienten in den Rückmeldegesprächen als Hilfestellung für die Gesprächsführer.

In dieser Beobachterkonferenz wurde ferner festgelegt, welches persönliche Feedback den Teilnehmern unabhängig von ihrem Abschneiden über das von ihnen gezeigte Verhalten am AC-Tag in den Rückmeldegesprächen gegeben werden sollte. Außerdem wurde bestimmt, welche Beobachter mit welchen Teilnehmern die Rückmeldegespräche führen sollten.

#### **3.7.2.4 Ergebnisfindung in der zweiten Beobachterkonferenz**

Nach den Rückmeldegesprächen fand eine zweite Beobachterkonferenz statt. Hier wurde differenziert über die Teilnehmern diskutiert, die wenigstens eine eingeschränkte Empfehlung für die zu besetzende Position bekommen hatten. In dieser Beobachterkonferenz ging es auch darum, die Ausprägung dieser Teilnehmer auf den einzelnen Verhaltensmerkmalen zu definieren und dem Anforderungsprofil gegenüberzustellen.

Dabei wurde ein Vorgehen gewählt, das schon von Stevens und Wonderlic (1934) propagiert wurde, um der mangelnden Trennung zwischen den verschiedenen Verhaltensmerkmalen durch die Beobachter entgegenzuwirken. Sie sind der Meinung, dass man diesem Effekt begegnen kann, indem man die Beobachter alle Teilnehmer zunächst bezüglich nur eines Verhaltensmerkmals beurteilen läßt, bevor man das nächste Verhaltensmerkmal beurteilt (vgl. Landy, 1985). Lammers (1992) bezeichnet dieses Vorgehen als nicht durchführbar für ein "Live-AC". Dennoch ist dieser Vorschlag recht einfach umzusetzen: anhand der Beobachterbögen und des kompletten Beobachtungsinventars werden alle Teilnehmer bezüglich der ersten Beobachtungsdimension eingeschätzt, bevor es zur zweiten Beobachtungsdimension geht etc.

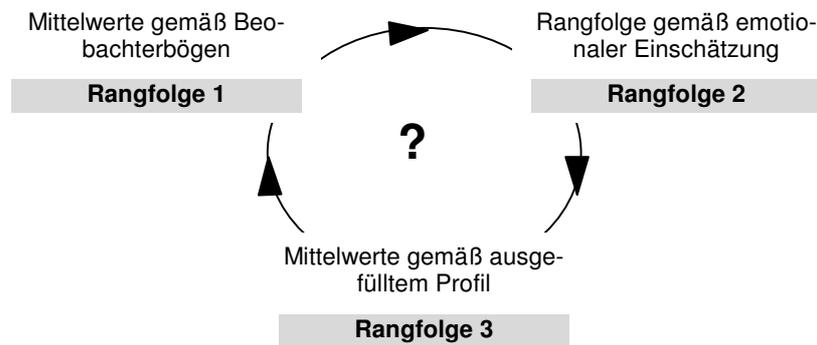
Taylor & Hastman (1956) meinten, dass dieses Vorgehen nicht zu dem erwünschten Ergebnis führt, weil es vielmehr zu einer Angleichung der verschiedenen Verhaltensmerkmale bezogen auf den einzelnen Teilnehmer kommt. Dieses Phänomen konnte in der Praxis bei den durchgeführten AC nicht beobachtet werden. Allerdings versuchten die Beobachter verschiedentlich, in der Diskussion über die Ausprägung eines Verhaltensmerkmals zwischen den Teilnehmern eine Rangfolge bezogen auf dieses Verhaltensmerkmale zu bilden und sie hinsichtlich ihrer Ähnlichkeiten und/oder Unterschiede zu kontrastieren. Es ging ihnen darum, gegenüber den Teilnehmern Urteilsgerechtigkeit im Diskussionsprozess walten zu lassen. Diesem Effekt konnte durch das Hinziehen der schriftlichen Aufzeichnungen aus den Beobachterbögen begegnet werden.

Es wurde darauf Wert gelegt, dass die Einschätzungen der Beobachter nach Möglichkeit einstimmig getroffen wurden. Wegen der Möglichkeit der differenzierten Betrachtung des Teilnehmergehaltens wurde in dieser Beobachterkonferenz die bekannte Beurteilungsskala von 1- 4 um die Werte 1,5, 2,5 und 3,5 erweitert.

Am Ende der Beobachterkonferenz wurde erneut eine Rangfolge über die Teilnehmer anhand der nun auf die Verhaltensmerkmale bezogenen fertig erstellten Profile gebildet. Es wurde untersucht, ob sich Rangplätze gegenüber den ersten beiden gebildeten Rangfolgen geändert haben. Die Ergebnisse wurden interpretiert und als Hilfestellung für die spätere Einarbeitung schriftlich festgehalten.

Die Beurteilungen der Beobachterkonferenz wurden zusammen mit der Empfehlung aus der ersten Beobachterkonferenz, in der es um Akzeptanz oder Ablehnung der einzelnen Teilnehmer ging, den potenziellen Vorgesetzten, sofern sie nicht Beobachter beim AC waren, zugefaxt. Zudem wurden Beobachter benannt, die diesen Vorgesetzten telefonisch ihre Eindrücke über deren Teilnehmer schildern konnten. Auch der AC-Leiter stand für Rückfragen zur Verfügung.

**Abb. 7: Vergleichende Betrachtung der verschiedenen Rangfolgen**



### 3.8 Die Verrechnung der Daten

Die für diese Untersuchung vorliegenden Daten wurden mit dem Statistikprogramm SPSS verrechnet. Als Maß für die Übereinstimmung der Beobachterurteile wurde der Produkt-Moment-Korrelationskoeffizient gewählt; die Auswertungen beziehen sich auf Daten, die sowohl mechanisch wie auch "klinisch" kombiniert werden (vgl. Feltham, 1988). Insgesamt gibt es zwei unterschiedliche Datensätze:

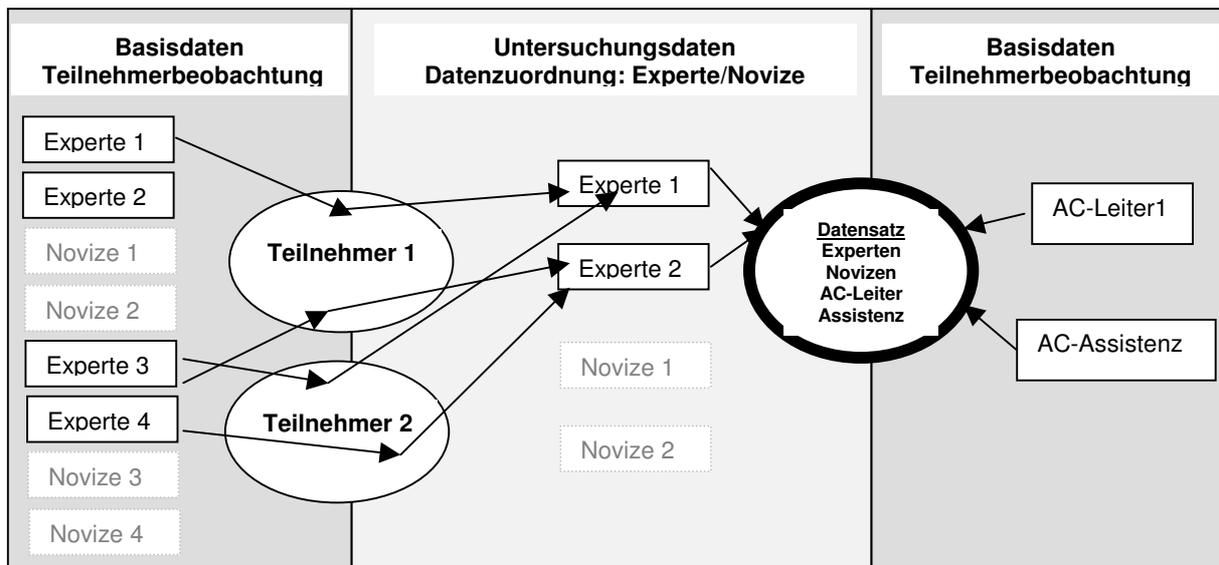
- Der erste Datensatz bezieht sich auf die Beobachtungen von 74 Experten, 73 Novizen, zwei AC-Assistenzen und einem AC-Leiter und 242 Teilnehmern. Zwischen den o.a. unterschiedlichen Zielpositionen (Vertriebsleiter/Versicherungsverkäufer) wurde nicht weiter differenziert. Zur Überprüfung der Forschungshypothesen wurden entsprechend dem üblichen in der Literatur beschriebenen Vorgehen (vgl. Bray u. Grant, 1966; Guldin, 1996; Kleinmann, 1997) mittlere Korrelationen zwischen den einzelnen Beobachtergruppen sowohl bezogen auf die einzelnen Übungen, wie auch bezogen auf die einzelnen Verhaltensmerkmale berechnet.

Des Weiteren erfolgte die Berechnung der übungsspezifischen Korrelationen für jede Beobachtergruppe.

Schließlich wurde geprüft, ob es einen Unterschied in der Leistungseinschätzung der Teilnehmer zwischen den Beobachtergruppen gibt. Dazu wurde die Korrelation zwischen der von den Beobachterbögen unabhängig erstellten "emotional" gefärbten Gesamtrangfolge über die Teilnehmer mit den beobachtergruppenspezifischen rankings berechnet.

Das Problem bei diesem Datensatz besteht darin, dass nicht alle Beobachter alle Teilnehmer in jeder Übung gesehen haben (s.o.) und von daher die Zellbesetzungen für die einzelnen Berechnungen unterschiedliche Fallzahlen aufweisen.

Abb. 8: Aufbau von Datensatz 1 am Beispiel von Übung 2 für die Gruppe der Experten



Was bedeutet das? Es geht in der vorliegenden Untersuchung darum herauszufinden ob Experten und Novizen sich in ihrem Beurteilungsverhalten voneinander unterscheiden. Das heisst, dass es darum geht, wie AC-Teilnehmer von diesen Personengruppen im Vergleich zu AC-Assistenz und AC-Leiter beurteilt werden. Die Basisdaten, bspw. von Herrn Meier, wurden gemäß seinem Expertenstatus (er war schon einmal bei einem AC als Beobachter dabei gewesen) als Experte 1 abgelegt. In der Übung 1, der Selbstpräsentation, beurteilte er alle Teilnehmer. In der Übung 2, der Gruppenübung, beurteilte er jedoch nur 4 Teilnehmer – darunter Teilnehmer 1, 3, 5, 8.

Der Experte 2, bspw. Herr Müller, beobachtete Teilnehmer 1 und Teilnehmer 2 in dieser Übung nicht.

Im Datensatz sollen aber die Beurteilungen von wenigstens 2 Experten für alle Teilnehmer in Übung 2 stehen. Aus diesem Grunde wurde dann geschaut, ob vielleicht Experte 3 den Teilnehmer 2 beobachtet hat. Wenn ja, wurden dessen Basisdaten den Untersuchungsdaten als Beurteilungsdaten dem Experten 1 für

Teilnehmer 2 zugeordnet.

Da Experte 3 auch Teilnehmer 1 beobachtet hat, wurden diese Daten den fehlenden Untersuchungsdaten von Experten 2 für Teilnehmer 1 zugeordnet. Die Untersuchungsdaten zum Teilnehmer 2 für den Experten 2 steuerte in diesem Falle Experte 4 mit seinen Basisdaten bei.

Für die Gruppe der Novizen wurde ebenso verfahren. Es wurde also immer in den Übungen, in denen die Beobachter nicht alle Teilnehmer beobachtet haben (beide Gruppenübungen, Zweiergespräch und Stressinterview), geprüft, welcher andere Beobachter mit gleichem Beobachterstatus (Experte/Novize) seine Basisdaten den Untersuchungsdaten beisteuern könnte. Im Extremfall könnten die Untersuchungsdaten von Experte 1 bei 8 Teilnehmern für 4 Teilnehmer aus den Basisdaten von Experten 1 und für die anderen 4 Teilnehmer aus den Basisdaten von 3 anderen Experten stammen.

In einem anderen AC könnten wiederum Herr Müller und Herr Meier als Beobachter tätig gewesen sein – allerdings könnte Herr Müller als Experte 4 und Herr Meier als Experte 1 die Basisdaten geliefert haben: die Zuordnung der Beobachter zu "Experte 1, Experte 2, Experte 3 oder Experte 4" war rein zufällig.

Wichtig ist festzuhalten, dass die Untersuchungsdaten von Experten und Novizen sich aus den Beurteilungsdaten vieler Personen zusammensetzen, während die Beurteilungsdaten von AC-Assistenz und AC-Leiter direkt in den zu untersuchenden Datensatz einfließen.

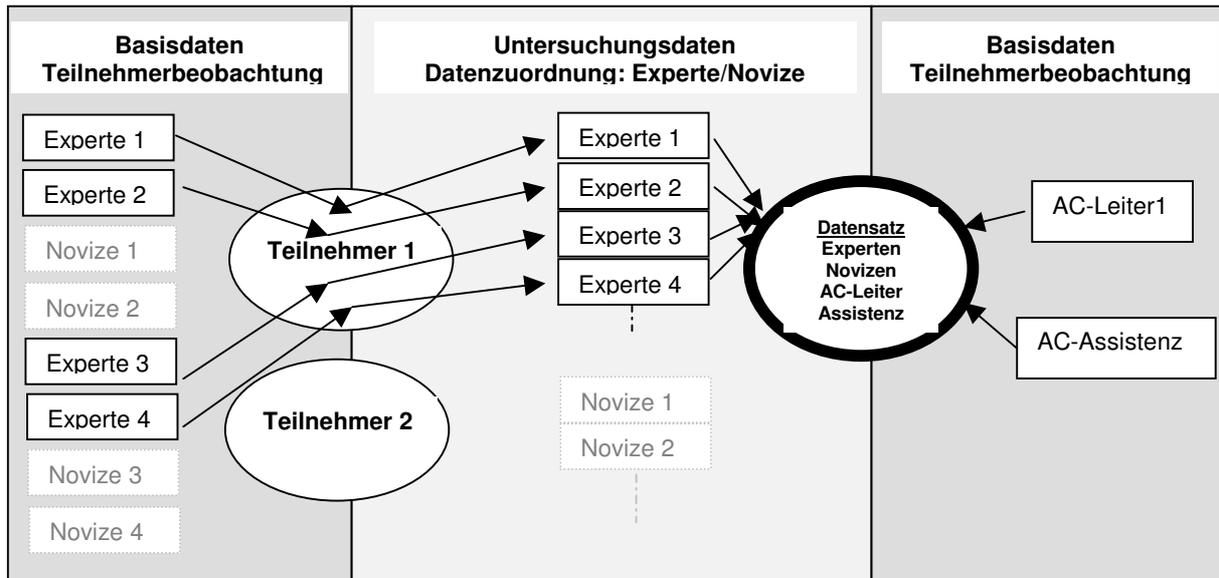
Datensatz 1 wurde so aufgebaut, dass Beobachtungsgegenstand immer der Teilnehmer ist.

- Der zweite Datensatz bezieht sich auf die Beobachtungen von neun Experten, 15 Novizen, einer AC-Assistenz, einem AC-Leiter und 25 Teilnehmern. Ein ähnlicher Aufbau wie für Datensatz 1 wurde auch für Datensatz 2 gewählt. Die Besonderheit bei diesem Datensatz besteht darin, dass die Anzahl der Teilnehmer maximal fünf betrug, so dass alle Beobachtergruppen alle Teilnehmer in jeder einzelnen Übung beobachtet haben. Daher sind hier die Basisdaten gleich den Untersuchungsdaten, weil wegen der geringen Anzahl an Teilnehmern immer alle Beobachter alle Teil-

nehmer in jeder Übung beobachtet haben. Mit diesem Datensatz wurden dieselben Berechnungen vorgenommen, wie mit dem ersten Datensatz.

Zusätzlich wurden in den Fällen, wo es möglich war, die vorliegenden Korrelationen auf Signifikanz überprüft.

Abb. 9: Aufbau von Datensatz 2



### 3.9 Statistisches Vorgehen

Die Auswertungen des vorliegenden Datensatzes (Datensatz 1) beziehen sich auf die Beobachtungen von insgesamt 81 Experten und 96 Novizen, zwei AC-Assistenzen (s.o.) und einem AC-Leiter. Die Daten wurden in 37 AC an 267 Teilnehmern erhoben. Eine Signifikanzprüfung der Daten ist ausgesprochen schwierig. Wie bereits weiter oben bemerkt (vgl. Kap. 3.8), setzen sich die Daten aus unabhängigen wie auch aus abhängigen Stichproben zusammen. Die Experten haben allesamt einmal als Novizen in einem AC beobachtet, insofern fließen in die Berechnung Werte ein und derselben Person ein, einmal als Novize und einmal als Experte (abhängige Stichprobe); andererseits gibt es Novizen, die nicht weiter als Experten in Erscheinung traten (unabhängige Stichprobe). Es ist kein statistisches Verfahren bekannt, das diesem Sachverhalt Rechnung tragen könnte – zumindest bietet SPSS eine derartige Lösung nicht an. Des Weiteren sind die Fallzahlen für die einzelnen Berechnungen unterschiedlich.

Es wäre jedoch wünschenswert herauszufinden, ob sich, bezogen auf den vorliegenden Datensatz, Experten und Novizen in ihrer Beobachterübereinstimmung mit dem AC-Leiter oder der AC-Assistenz voneinander unterscheiden. Daher wurde folgende Annahme getroffen:

Die Daten werden, wo es möglich ist, so miteinander verrechnet, als ob es sich um Daten aus unabhängigen Stichproben handeln würde. Will man die Korrelationen für zwei unabhängige Stichproben mit unterschiedlichen Fallzahlen auf Signifikanz überprüfen, muss man die Korrelationen in Fisher's Z-Werte transformieren und dann die z-Werte (z ist dabei die Einheit der Standardnormalverteilung) berechnen.

Nach Bortz (1979, S.263) kommt dabei folgende Formel zur Anwendung:

$$z = \frac{Z_1 - Z_2}{\sigma_{(Z_1 - Z_2)}}$$

wobei

$$\sigma_{(Z_1 - Z_2)} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

Dabei wird überprüft, ob die beiden Stichproben aus derselben Grundgesamtheit stammen. Diese Annahme muss verworfen werden, wenn der gefundene Wert größer ist als der nach vorgegebenen Signifikanzniveau zu erwartende Wert. In der vorliegenden Untersuchung wurde diese Vorgehensweise gewählt, um - wenn möglich - mittlere Korrelationen zwischen Beobachtergruppen auf Signifikanz zu prüfen.

## 4 Ergebnisse

### 4.1 Mittlere Korrelationen der Beobachtergruppen untereinander über die Übungen hinweg

Forschungshypothese 1 lautet.

Wenn spezifische trainierbare Beobachterkompetenzen für die Qualität von AC-Ergebnissen ausschlaggebend sein sollen, dann müssen Unterschiede in der Beobachterübereinstimmung in Abhängigkeit vom Trainiertheitsgrad der Beobachter bestehen. Unternehmensinterne Führungskräfte und/oder Mitarbeiter, die wenigstens einmal bei einem AC als Beobachter dabei waren (= trainierte Experten) korrelieren dann mit einem vielfach trainierten AC-Leiter und einer gut trainierten AC-Assistenz höher als unternehmensinterne Mitarbeiter, die noch nie bei einem AC als Beobachter dabei waren (= untrainierte Novizen).

Zur Prüfung dieser Hypothese wurden die übungsbezogenen mittleren Korrelationen zwischen den einzelnen Beobachtergruppen überprüft. Es wurden dazu Produkt-Moment-Korrelationen berechnet, die dann mittels Fisher's Z transformiert wurden. Im Falle der Experten und der Novizen, wo jeweils 2 Datensätze (Experte 1 und Experte 2, vgl. Kapitel 3.8) zu einem Datensatz der Beobachtergruppe "Experten" oder "Novizen" zusammengefasst wurden, wurde aus den beiden Korrelationen das arithmetische Mittel gebildet, das dann wieder rückgerechnet wurde (vgl. Bortz, 1979, S. 261).

#### 4.1.1 Übungsbezogene mittlere Korrelationen zwischen AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter - Novizen

Die mittleren Korrelationen zwischen dem AC-Leiter und der AC-Assistenz, dem AC-Leiter und den Experten sowie dem AC-Leiter und den Novizen sind in Tabelle 4 aufgeführt. Die Annahme ist folgende: der AC-Leiter hat das AC konstruiert und war über lange Jahre als Beobachter und AC-Leiter tätig. Er hatte bereits mehrere Beobachtertrainings erhalten und kann von daher quasi als "Oberexperte" gelten. Die AC-Assistenz hat den nächst höheren Expertenstatus, da sie vom AC-Leiter ausgebildet wurde und schon vor den für diese Untersuchung relevanten AC als Beobachter tätig war. Die dritthöchste Expertise weisen die unternehmensinternen "Experten" auf. Sie waren alle mindestens einmal schon als Beobachter bei einem AC dabei, und hatten

alle Erfahrung in der Auswahl von Mitarbeitern.

In Tabelle 4 sind unterschiedlich hohen Fallzahlen aufgeführt. Diese kommen folgendermaßen zu Stande:

In der Übung "Selbstpräsentation" werden immer alle Teilnehmer von allen Beobachtern beobachtet. In den Gruppenübungen beobachten die AC-Assistenz, die Experten und die Novizen jeweils nur die vier Teilnehmer, die sie am besten sehen können. Der AC-Leiter beobachtet aber auch in dieser Übung konsequent alle Teilnehmer.

In der Übung "Stressinterview" gibt es eine Aufteilung der Teilnehmergruppe in zwei Gruppen: eine Gruppe wird vom AC-Leiter moderiert, die andere Gruppe von der AC-Assistenz. Aus diesem Grunde liegen hier keine Korrelationen zwischen AC-Leiter und AC-Assistenz vor.

In der Übung "Zweiergespräch" werden ebenfalls zwei Gruppen gebildet, jedoch gibt es eine bestimmte Anzahl von AC, bei denen wegen der geringen Teilnehmerzahl auf die Zweiteilung verzichtet wurde. Dies ist folgendermaßen zu erklären: das "Stressinterview" dauert in der Regel 15 Minuten je Teilnehmer. Bei sechs Teilnehmern bedeutet dies, dass eine Beobachtergruppe mit je drei Teilnehmer 60 Minuten lang aktiv ist. In diesem Fall kann im Zweiergespräch auf die Einteilung in zwei Gruppen verzichtet werden, da ein Zweiergespräch 10 Minuten je Teilnehmer dauert, und bei sechs Teilnehmern  $3 \times 10 \text{ Minuten} = 30 \text{ Minuten}$  Zeitpolster benötigt werden (es sprechen ja immer zwei Teilnehmer miteinander). Die Zeit für diese zwei Übungen, also die 60 Minuten für das "Stressinterview" und die 30 Minuten für das "Zweiergespräch", ist genau die Zeit, die die Teilnehmer brauchen, um den "Postkorb" und die "Wegeplanungsaufgabe" zu bearbeiten. Insofern liegen nur für einen Teil der AC gemeinsame Beobachtungen zwischen AC-Assistenz und AC-Leiter in der Übung "Zweiergespräch" vor.

Bei den AC mit ungerader Teilnehmerzahl übernahm der AC-Leiter standardmäßig mehr Teilnehmer in der Übung "Stressinterview", während die AC-Assistenz dann in der Übung "Zweiergespräch" mehr Teilnehmer hatte (oder umgekehrt). Zudem gab es eine Reihe von AC, in denen nur maximal drei Experten/Novizen beobachteten. Entsprechend dem in Kapitel 3.8. beschriebenen Vorgehen konnten die in diesem Fall fehlenden Daten nicht (z.B. bei 0 - 3 Experten oder Novizen) oder nur unzureichend aus den Daten weiterer Experten/Novizen aufgefüllt werden. Daher erklären sich die unterschiedlichen Fallzahlen.

In der Betrachtung der mittleren Korrelationen von Datensatz 1 fällt auf, dass unabhängig davon, ob es sich um die AC-Assistenz, die Expertengruppe oder die Novizengruppe handelt, die Korrelationen mit dem AC-Leiter ähnlich hoch sind. Zudem sind sie insgesamt deutlich geringer (halb so hoch) als in der Untersuchung von Greenwood und McNamara (1967, vgl. Anhang 1, S. 123). Während die AC-Assistenz und die Novizen in der Übung "Zweiergespräch" am höchsten mit dem AC-Leiter korrelieren, korrelieren die Experten mit diesem in der Übung "Stressinterview" am höchsten.

Da zudem angenommen wird, dass der AC-Leiter die höchste Beobachtungskonstanz hat und sich am stärksten bei seinen Beurteilungen an die zu beurteilenden Verhaltensoperationalisierungen der Beobachterbögen hält, deuten hohe Korrelationen mit dem AC-Leiter auf eine hohe Expertise der entsprechenden Beobachtergruppe hin, da eine Übereinstimmung mit dem Blickwinkel des AC-Leiters besteht. Diese Übereinstimmung ist nach den vorliegenden Ergebnissen für die untersuchten Beobachtergruppen übungsabhängig. Die AC-Assistenz beobachtet am genauesten das Miteinander (Zweiergespräch), das einen Indikator für partnerschaftliches Verhalten darstellt, während die Experten ihren Schwerpunkt mehr beim Stressinterview haben, wo es darum geht, sich durchzusetzen und für seine Meinung zu kämpfen.

Die Novizen, die noch keinerlei Erfahrung mit dem AC haben, weisen in den zwei ersten Übungen niedrigere Korrelationen mit dem AC-Leiter auf als in den drei letzten Übungen. Möglicherweise orientieren sie sich nach der ersten Phase der Unsicherheit stärker an den Beobachterbögen und können mit der Beurteilungssituation zunehmend besser umgehen (Übungseffekt).

Bei den Experten sinken die mittleren Korrelationen nach der Übung "Stressinterview" ab. Für sie ist diese Übung möglicherweise die Übung des AC und bewirkt, dass sich die Experten in ihrem Urteil über die Teilnehmer möglicherweise schon nach dieser Übung so sicher waren, dass sie sich in der Folge in ihren Beobachtungen und Beurteilungen weniger genau an die Beobachterbögen gehalten haben und mehr undifferenziert beobachteten.

Dass in der zweiten Gruppenübung die Korrelationen unter denen der ersten Gruppenübung liegen, kann einerseits an dieser vorschnellen Urteilsbildung liegen, andererseits aber auch darauf zurückgeführt werden, dass die Teilnehmergruppen in der zweiten Gruppenübung generell weniger Aktivitäten gezeigt haben als in der ersten Gruppen-

übung. Das Absinken der Gruppenaktivität kann darauf beruhen, dass die Übung schlecht konstruiert war, also nicht zu dem in dem Beobachtungsinventar nahegelegten Verhalten provoziert hat, oder dass die Belastbarkeit der Teilnehmergruppe nach dem anstrengenden AC-Verlauf dazu geführt hat, dass weniger Aktivitäten gezeigt wurden.

Übung	AC-Leiter und...		
	AC-Assistenz	Experten	Novizen
Selbstpräsentation	.320 (91)	.250 (366)	.310 (406)
Gruppenübung 1	.330 (70)	.380 (185)	.355 (207)
Stressinterview	/	.425 (297)	.430 (378)
Zweiergespräch	.425 (76)	.345 (78)	.440 (165)
Gruppenübung 2	.330 (71)	.310 (195)	.415 (208)

in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung der mittleren Korrelationen bilden.

**Tab. 4: Übungsbezogene mittlere Korrelation zwischen AC-Leiter und den anderen Beobachtergruppen**

Daher könnte die Verhaltensbeurteilung erschwert gewesen sein und die Experten bemühten sich, beurteilbare Verhaltensweisen zu entdecken. Dabei griffen sie wahrscheinlich auf ihre eigenen Kategoriensysteme zurück und beurteilten daher die Teilnehmer sehr unterschiedlich. Die Novizen hingegen hielten sich, um gute Beobachter zu sein, mehr an die Beurteilungsskala der Beobachterbögen: kaum gezeigtes Verhalten ist kaum beobachtbar und muss mit einer entsprechenden Beurteilung versehen werden. Offenbar waren in dieser Übung Novizen und AC-Leiter eher einer Meinung als Experten-, bzw. die AC-Assistenz und der AC-Leiter.

Der Trainiertheitsgrad der Beobachter müsste sich in einem signifikanten Unterschied, z.B. zwischen den Experten und den Novizen niederschlagen. Die Unabhängigkeit dieser beiden Beobachtergruppen ist nicht gegeben (s.o.). Dennoch wurde unter der Annahme der Unabhängigkeit eine Signifikanzprüfung vorgenommen. Ein 1%-iges Signifikanzniveau ergibt sich für einen errechneten z-Wert von  $< -2.23$ , ein 5%-iges Signifikanzniveau ergibt sich für einen errechneten z-Wert von  $< -1,64$  und ein 10%-iges Sig-

nifikanzniveau ergibt sich für einen errechneten z-Wert von  $< -1,28$ . Hier zeigt sich, dass bezogen auf die Übung "Zweiergespräch" ( $z = -1,96$ ) und die Gruppenübung 2 ( $z = -2,72$ ) signifikante Unterschiede in der Beobachterübereinstimmung mit dem AC-Leiter bestehen. Bei allen anderen Übungen liegen die errechneten z-Werte ausserhalb der o.a. Grenzen.

#### **4.1.2 Übungsbezogene mittlere Korrelationen zwischen AC-Assistenz - Experten und AC-Assistenz - Novizen**

Tabelle 5 zeigt die übungsbezogenen mittleren Korrelationen zwischen AC-Assistenz und Experten bzw. AC-Assistenz und Novizen (bzgl. der unterschiedlichen Fallzahlen vgl. die Ausführungen oben). Auffällig ist hier, dass sowohl für die Experten wie auch für die Novizen ähnliche Korrelationsmuster gelten, wie in den Auswertungen der Tabelle 4.

Die höchste Übereinstimmung in Bezug auf die AC-Assistenz gibt es für die Experten und die Novizen beim "Stressinterview", wahrscheinlich begünstigt durch die hohe Eigenbeteiligung der Beobachter in der Interviewsituation. Möglicherweise spielen hier zudem noch Identifikationsprozesse der Beobachter mit den Teilnehmern eine Rolle, die in einem besonderen Ausmaß die Sensibilität der Beobachter in der Interviewsituation fördern.

Ähnlich wie mit dem AC-Leiter (vgl. Tabelle 4) korrelieren auch in Tabelle 5 die Experten in der Übung "Selbstpräsentation" mit der AC-Assistenz deutlich niedriger als die Novizen. Dies könnte daran liegen, dass sich die Experten weniger von den im Beobachterbogen zu beobachtenden Verhaltensweisen haben leiten lassen. Da diese Übung von einzelnen Teilnehmern stellenweise in ein bis zwei Minuten bewerkstelligt wird, bleibt wenig Zeit, Verhaltensweisen eindeutig zu registrieren. Jeder Experte hat einen eigenen Blickwinkel, nach dem er das wenige beobachtbare Verhalten beurteilt. Die Novizen mögen hier noch ein Stück offener an die Beobachtungen herangegangen sein, und durch ihre Ungeübtheit mit dem Beobachterbogen undifferenzierter beurteilt haben (Halo-Effekt).

Übung	AC-Assistenz und...	
	Experten	Novizen
Selbstpräsentation	.185 (166)	.280 (188)
Gruppenübung 1	.300 (64)	.385 (79)
Stressinterview	.490 (50)	.500 (40)
Zweiergespräch	.315 (118)	.400 (155)
Gruppenübung 2	.290 (65)	.300 (80)

in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung der mittleren Korrelationen bilden.

**Tabelle 5: Übungsbezogene mittlere Korrelation zwischen AC-Assistenz und den anderen Beobachtergruppen**

Die Überprüfung der Unterschiede zwischen Experten und Novizen ergab folgende z-Werte: Selbstpräsentation: -.95, Gruppenübung 1: -.56, Stressinterview: -.008, Zweiergespräch: -.79, Gruppenübung 2: -.06. In Bezug auf die Übereinstimmung mit der AC-Assistenz sind demnach zwischen den Experten und den Novizen keine signifikanten Unterschiede feststellbar.

#### **4.1.3 Übungsbezogene mittlere Korrelationen zwischen Experten und Novizen**

Die Korrelationsmuster der übungsbezogenen mittleren Korrelationen zwischen Experten und Novizen sind ähnlich denen, die sich auch schon in Tabelle 4 und Tabelle 5 wiederfinden (vgl. Tabelle 6).

Offenbar hat das "Stressinterview" eine Spezifität, die es nahelegt, die zu beobachtenden Verhaltensweisen in einer ähnlichen Art und Weise zu beobachten - unabhängig davon, ob man Experte oder Novize ist. Es fällt auf, dass in der ersten Übung, der "Selbstpräsentation", die Korrelationen niedriger sind als in den anderen Übungen. Möglicherweise ist dies ein Effekt des noch nicht vorhandenen Erfahrungswissens mit dem AC-System seitens der Novizen, der dazu führt, dass diese sich mehr als die Experten am Beurteilungsinventar orientieren müssen um ihre Unsicherheit zu reduzieren.

Übung	Experten / Novizen
Selbstpräsentation	.255 (706)
Gruppenübung 1	.400 (177)
Stressinterview	.485 (112)
Zweiergespräch	.325 (572)
Gruppenübung 2	.390 (181)

in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung mittleren Korrelationen bilden.

**Tabelle 6: Übungsbezogene mittlere Korrelation zwischen Experten und Novizen**

In den Gruppenübungen sind die Korrelationen ähnlich hoch. Lediglich in der Übung "Zweiergespräch" kommt es wieder zu einem leichten Abfall, so dass hier vermutet werden kann, dass Experten und Novizen unterschiedliche internale Konzepte davon haben, was für Verhaltensweisen in dieser Übung gezeigt werden sollten.

Wenn man berücksichtigt, dass auch zwischen AC-Assistenz und den Experten in der Übung "Zweiergespräch" die Korrelationen relativ niedriger sind als zwischen den Novizen und der AC-Assistenz, dann kann vermutet werden, dass der spezifische Blickwinkel der Experten für dieses Ergebnis bedeutsam ist.

Da "Zweiergespräch" bedeutet, dass es sich in diesen Gesprächssituationen um Vertriebssituationen handelt und die Experten auch zu einem großen Teil auf langjährige Vertriebserfahrung zurückblicken können, kann es durchaus sein, dass hier der spezifische Blickwinkel der Vertriebssituation zu einer eher strengen Beurteilung seitens der Experten geführt hat. Demgegenüber könnten Novizen und AC-Assistenz, die über weniger derartige Erfahrungen verfügen, eher Aspekte des "Miteinander in Kontakt-Tretens" in den Vordergrund ihrer Betrachtung gestellt haben.

#### **4.2 Mittlere Korrelationen der Beobachtergruppen untereinander über die Verhaltensmerkmale hinweg**

Wie sieht es nun mit den mittleren, auf die einzelnen Verhaltensmerkmale bezogenen

---

Korrelationen aus? Dem AC wird ja immer der Vorwurf gemacht, dass Beobachter eine globale Eindrucksbildung zu Ungunsten einer differenzierten Eindrucksbildung vornehmen - was eigentlich nicht der Fall sein dürfte, wenn man den Zweck eines derartigen Verfahrens (individuellen Stärken-/Schwächenanalyse, Potenzialdiagnostik) bedenkt. Gemäß Forschungshypothese 1 müsste sich die höhere Expertise der verschiedenen Beobachtergruppen auch in der Höhe der mittleren Korrelationen der Beobachtergruppen untereinander feststellen lassen.

#### **4.2.1 Merkmalsbezogene mittlere Korrelationen zwischen dem AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter - Novizen**

Bezüglich der mittleren Korrelationen zwischen AC-Leiter und AC-Assistenz, AC-Leiter und Experten und AC-Leiter und Novizen ist festzustellen, dass auch hier die Höhe der Korrelationen unter den in der Literatur berichteten Werten liegt (vgl. u.a. die Ergebnisse von Guldin 1996, Anhang 1, S. 123). Bei den Werten handelt es sich um die Berechnung von identischen Verhaltensmerkmalen, die in verschiedenen Übungen jeweils übungsspezifisch operationalisiert beobachtet werden (vgl. Anhang 3).

Die mit einem Stern versehenen Werte bei der AC-Assistenz beziehen sich auf Auswertungen ohne die Daten der Übung "Stressinterview" (vgl. dazu die Ausführungen unter 4.1.1). Insofern beruhen diese Werte auf Beobachtungen von Verhaltensmerkmalen, die nur in einer, nicht aber in zwei oder drei Übungen vorgenommen werden (zur Verteilung der Verhaltensmerkmale auf die verschiedenen Übungen siehe Anhang 2).

In Tabelle 7 fällt auf, dass die merkmalsbezogenen mittleren Korrelationen zwischen AC-Leiter und den anderen Beobachtergruppen die Annahme, dass der Expertenstatus einen Einfluss auf die Beobachterübereinstimmung hat, nicht bestätigen können.

Zwischen AC-Leiter und AC-Assistenz gibt es die höchste Übereinstimmung im Merkmal "Engagement", das vor allem im "Stressinterview" und im "Zweiergespräch" beobachtet wird. Außerdem gibt es eine hohe Übereinstimmung für die Merkmale "Teambildung unterstützen" und "Schrittweises und kontrolliertes Arbeiten" (beobachtet in den Gruppenübungen), sowie "Verbalisierung" (beobachtet im "Zweiergespräch" und in der

"Selbstpräsentation"). Auffällig ist, dass das Merkmal "Feedback geben", das in den Gruppenübungen beobachtet wird, extrem niedrige Übereinstimmungen aufweist. Da dies alle Beobachtergruppen betrifft, ist zu vermuten, dass es entweder schlecht beobachtet werden kann, oder dass der Charakter der Gruppenübungen es nicht nahelegt, entsprechendes Verhalten zu zeigen.

Merkmal	AC-Leiter und...		
	AC-Assistenz	Experten	Novizen
Umgang mit belastenden Situationen	.216* (91)	.320 (444)	.385 (571)
Engagement	.485* (76)	.435 (182)	.435 (344)
Persönliches Auftreten	.442* (91)	.395 (444)	.400 (571)
Verbalisierung	.410 (167)	.240 (470)	.335 (585)
Überzeugungsfähigkeit	.385 (167)	.315 (470)	.375 (585)
Zwischenm. Beziehungen gestalten	.290 (217)	.370 (290)	.360 (358)
Selbständiges Verhalten	.345* (141)	.455 (304)	.455 (434)
Feedback geben	.105 (141)	.200 (226)	.180 (169)
Teambildung unterstützen	.425 (141)	.440 (226)	.465 (169)
Problemlöseverhalten	.375 (141)	.355 (226)	.395 (169)
Schrittweises u. kontrolliertes Arbeiten	.460 (141)	.375 (226)	.450 (169)
Zielorientierung	.200 (141)	.420 (226)	.300 (169)
Geistige Flexibilität	.448* (167)	.375 (548)	.265 (750)

\* die Datenbasis dieser Merkmale ist um die Daten der Übung 3 reduziert (zur Problematik vgl. die Ausführungen weiter oben).; in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung der mittleren Korrelationen bilden

**Tabelle 7: Merkmalsbezogene mittlere Korrelation zwischen AC-Leiter und den anderen Beobachtergruppen**

Die niedrigen Übereinstimmungen im Merkmal "Zielorientierung", das ja, wie weiter oben angesprochen, nur bei der Auswahl von Führungspositionen beobachtet werden sollte, sind wahrscheinlich ein Indikator für eine unterschiedliche Auffassung zum Thema "Führungsverhalten". Zusätzlich scheint die individuelle Erfahrung in Führungssituationen (AC-Leiter und Experten sind führungserfahren) bedeutsam zu sein (füh-

rungsunerfahrene AC-Assistenz und evtl. auch Novizen).

Die Korrelationsmuster zwischen AC-Leiter und Experten, bzw. AC-Leiter und Novizen sind ähnlich: es finden sich vergleichbare Korrelationshöhen in den Merkmalen "Engagement", "Persönliches Auftreten", "Zwischenmenschliche Beziehungen gestalten", "Selbständiges Verhalten", "Feedback geben" und "Teambildung unterstützen" und größere Unterschiede in den anderen Merkmalen. Hierbei betragen die größte Unterschiede .095 beim Merkmal "Verbalisierung" und .120 beim Merkmal "Zielorientierung". Dies bedeutet zunächst, dass AC-Leiter und Experten, bzw. AC-Leiter und Novizen in bestimmten Verhaltensdimensionen ähnliche Blickwinkel haben und in anderen Verhaltensdimensionen unähnliche Blickwinkel.

Die Überprüfung der Übereinstimmung von Experten und Novizen mit dem AC-Leiter ergab signifikante Unterschiede lediglich für die Merkmale "Zielorientierung" ( $z = -1.35$ ), "Umgang mit belastenden Situationen" ( $z = -.132$ ), "Geistige Flexibilität" ( $z = 2.19$ ) und das Merkmal "Verbalisierung" ( $z = -.171$ ), nicht aber für die anderen Verhaltensmerkmale.

#### **4.2.2 Merkmalsbezogene mittlere Korrelationen zwischen AC- Assistenz - Experten und AC-Assistenz - Novizen**

Die höchsten Übereinstimmungen gibt es hier in den Merkmalen "Engagement", "Persönliches Auftreten", "Selbständiges Verhalten", die niedrigsten in den Merkmalen "Problemlöseverhalten" und "Überzeugungsfähigkeit". Es fällt auf, dass zwischen AC-Assistenz und Novizen eine höhere Beobachterübereinstimmung in den Korrelationswerten ausgedrückt wird (sie schwanken zwischen .290 und .480), als zwischen AC-Assistenz und Experten: hier liegt die Schwankungsbreite zwischen .135 und .440. Wie in den Korrelationen zwischen AC-Leiter und den verschiedenen anderen Beobachtergruppen auch (vgl. Tabelle 7) scheinen auch in den Korrelationen zwischen AC-Assistenz und den Experten und den Novizen die Merkmale "Engagement", "Persönliches Auftreten", "Selbständiges Verhalten", "Schrittweises und kontrolliertes Arbeiten" und "Teambildung" die beste Beobachterübereinstimmung nach sich zu ziehen, so dass davon ausgegangen werden kann, dass diese auch am besten zu beobachten und/oder zu bewerten sind.

Merkmal	AC-Assistenz und...	
	Experten	Novizen
Umgang mit belastenden Situationen	.182 (216)	.425 (228)
Engagement	.420 (168)	.480 (195)
Persönliches Auftreten	.420 (216)	.370 (228)
Verbalisierung	.280 (.284)	.350 (343)
Überzeugungsfähigkeit	.185 (284)	.345 (343)
Zwischenm. Beziehungen gestalten	.360 (247)	.310 (347)
Selbständiges Verhalten	.440 (179)	.445 (199)
Feedback geben	.285 (129)	.360 (179)
Teambildung unterstützen	.340 (129)	.290 (179)
Problemlöseverhalten	.135 (129)	.300 (179)
Schrittweises u. kontrolliertes Arbeiten	.310 (129)	.415 (179)
Zielorientierung	.380 (129)	.330 (179)
Geistige Flexibilität	.340 (334)	.325 (383)

in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung der mittleren Korrelationen bilden.

**Tabelle 8: Merkmalsbezogene mittlere Korrelation zwischen AC-Assistenz, Experten und Novizen**

Auch hier wurden signifikante Unterschiede in der Übereinstimmung mit der AC-Assistenz zwischen den Experten und Novizen gefunden, und zwar bezogen auf die Merkmale "Umgang mit belastenden Situationen" ( $z = -1.30$ ), "Überzeugungsfähigkeit" ( $z = -2.16$ ) und "Problemlöseverhalten" ( $z = -1.5$ ).

#### 4.2.3 Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen

Betrachtet man die merkmalsbezogenen mittleren Korrelationen zwischen Experten und Novizen (vgl. Tabelle 9) fällt auf, dass die Schwerpunkte der Übereinstimmung in den Merkmalen "Engagement", "Zwischenmenschliche Beziehungen gestalten", "Überzeugungsfähigkeit" und "Feedback geben" liegen. In diesen Merkmalen liegen die Kor-

relationen über .40.

Merkmal	Experten/Novizen
Umgang mit belastenden Situationen	.367 (883)
Engagement	.432 (684)
Persönliches Auftreten	.356 (883)
Verbalisierung	.377 (1278)
Überzeugungsfähigkeit	.436 (470)
Zwischenm. Beziehungen gestalten	.540 (930)
Selbständiges Verhalten	.381 (293)
Feedback geben	.404 (293)
Teambildung unterstützen	.377 (293)
Problemlöseverhalten	.377 (293)
Schrittweises u. kontrolliertes Arbeiten	.354 (293)
Zielorientierung	.346 (293)
Geistige Flexibilität	.304 (1390)

in Klammern sind jeweils die Fallzahlen aufgeführt, die die Grundlage für die Berechnung der mittleren Korrelationen bilden.

**Tabelle 9 : Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen**

Auffällig die Korrelation von .54 für das Merkmal "Zwischenmenschliche Beziehungen gestalten". Sie könnte ein Hinweis dafür sein, dass Experten und Novizen eine gemeinsame Vorstellung über das Konzept "Zwischenmenschliche Beziehungen gestalten" haben.

Auch in den Beobachterübereinstimmungen zwischen Experten und Novizen sind die Korrelationen wieder unter den in der Literatur berichteten Höhen (vgl. Anhang 1, S. 123). Als erstes Fazit lässt sich daher festhalten: über alle Kombinationen der Beobachterübereinstimmung hinweg (übungsbezogen oder merkmalsbezogen) lassen sich nur vereinzelt Einflüsse der Expertise erkennen. Es gibt eventuell Übungsschwer-

punkte, die zu einer stärkeren Übereinstimmung führen, es gibt auch Verhaltensmerkmale, in denen die Beobachtergruppen unterschiedlich miteinander übereinstimmen - aber ein deutlicher Effekt des Expertenstatus ist zunächst nicht festzustellen. Die Beobachterübereinstimmungen in den verschiedensten Beobachterkombinationen unterscheiden sich kaum. Die Höhe der mittleren Korrelationen deutet nicht darauf hin, dass eine globale Eindrucksbildung über die Teilnehmer stattgefunden hat.

### 4.3 Korrelation der Beobachtergruppen innerhalb der Übungen

Als Begründung für ein Beobachtertraining wird in der Literatur unter anderem angeführt, dass trainierte Beobachter besser als untrainierte Beobachter in der Lage sind zwischen den einzelnen Verhaltensdimensionen zu differenzieren.

Würde dies der Fall sein, dann müsste sich nachweisen lassen, dass untrainierte Beobachter innerhalb der Übungen über die einzelnen Verhaltensmerkmale hinweg höhere Korrelationen aufweisen (als Indikator für einen Halo-Effekt) als trainierte Beobachter.

Gleichzeitig müssten die Korrelationen innerhalb der einzelnen Verhaltensmerkmale über die Übungen hinweg bei den untrainierten Beobachtern (sie hätten eine geringere Konstruktvalidität) niedriger sein als bei den trainierten Beobachtern, weil die Fähigkeit, ein bestimmtes Verhaltensmerkmal in unterschiedlichen Übungen zu beobachten, bei den trainierten Beobachtern besser ausgeprägt ist.

Forschungshypothese 2 lautet:

Die Korrelationen innerhalb der einzelnen AC-Übungen sind für die Experten geringer als für die Novizen, weil die Experten besser zwischen den Verhaltensmerkmalen differenzieren können als die Novizen (für die Novizen lässt sich ein Halo-Effekt nachweisen).

Zur Überprüfung dieser Annahme wurden die mittleren Korrelationen der einzelnen Beobachtergruppen über die Verhaltensmerkmale hinweg innerhalb der einzelnen Übungen berechnet (Tabelle 10):

	Selbstprä- sentation	Gruppen- übung 1	Stress- interview	Zweier- gespräch	Gruppen- Übung 2
<b>AC-Leiter</b>	.45	.45	.48	.38	.45
<b>AC-Assistenz</b>	.39	.52	.37	.46	.50
<b>Experten</b>	.47	.72	.57	.70	.72
<b>Novizen</b>	.58	.72	.45	.59	.72

**Tabelle 10: Mittlere Korrelationen der Beobachtergruppen innerhalb der einzelnen Übungen**

Die vorliegenden Daten zeigen, dass der AC-Leiter offenbar unabhängig von der jeweiligen Übung die gleichmäßigste Beobachtungsqualität hat (erkennbar an der relativ gleichbleibend geringen Höhe der Korrelationen über die verschiedenen Übungen hinweg). Danach folgt die AC-Assistenz; Experten und Novizen weisen innerhalb der Übungen (bis auf die Übung "Selbstpräsentation" bei den Experten und die Übung "Stressinterview bei den Novizen") die höchsten Korrelationen auf (zwischen .57 bis .72), so dass hier das Vorliegen eines Halo-Effekts vermutet werden kann. Besonders auffällig sowohl für die Experten wie auch für die Novizen sind die hohen Korrelationen von .72 in den Gruppenübungen.

Verantwortlich für dieses Phänomen sind möglicherweise gegenläufige Effekte: seitens der Experten kann man einen höheren Zufallsfehler unterstellen, der dadurch zu Stande kommt, das diese ihre Beurteilungen weniger bewusst abgeben (ihr Urteil über die Teilnehmer steht früher fest als bei den Novizen). Seitens der Novizen kann man einen höheren systematischen Fehler unterstellen, dessen Ursache in deren Unerfahrenheit im Umgang mit dem AC-System und deren Überforderung erklärt werden kann.

#### **4.4 Eindrucksbildung über die Teilnehmer auf Grund von Rankings**

Bei dem vorliegenden AC-Typ ging es darum, nach den Übungen des AC entsprechend der Leistung der Teilnehmer eine Reihenfolge über diese zu bilden, ohne dass die schriftlich gemachten Beurteilungen der Beobachterbögen hinzugezogen wurden. Dabei wurde diese Rangfolge von jedem Beobachter individuell erstellt, und erst dann wurden die individuellen Rangfolgen verlesen und die Rangsummen wurden ermittelt.

Eine Gemeinsamkeit der Beobachter in ihrer Einschätzung von den Teilnehmern sollte sich auch in diesen Daten niederschlagen.

Forschungshypothese 3 lautet:

Hat das Beobachtertraining einen Einfluss auf die Beobachterübereinstimmung, dann bilden die rankings trainierter Beobachter über die Teilnehmer genauer die Reihenfolge der Teilnehmer im AC ab als die rankings untrainierter Beobachter.

Bei dem hier berücksichtigten Datensatz weisen AC-Leiter und Novizen die höchsten Korrelationen mit der von allen Beobachtern unabhängig von den Beobachterbögen ermittelten Gesamtrangfolge über die Teilnehmer auf, gefolgt von AC-Assistenz und Experten (vgl. Tabelle 11). Die allgemeine Höhe der Koeffizienten spricht dabei für eine hohe Übereinstimmung mit der auf dem Gesamteindruck über die Teilnehmer beruhenden Einschätzung durch die unterschiedlichen Beobachtergruppen. Experten und Novizen weisen ähnlich hohe Korrelationen mit dem AC-Leiter, die AC-Assistenz geringere.

	Gesamtrangfolge	AC-Leiter	AC-Assistenz	Experten
AC-Leiter	.90			
AC-Assistenz	.81	.61		
Experten*	.75	.76	.60	
Novizen*	.92	.80	.71	.66

\*diese Werte sind gemittelte Korrelationen, vgl. 4.1

**Tabelle 11: Korrelation der Rangfolge über die Teilnehmer aufgrund des Gesamteindrucks, sowie Korrelation zwischen den Beobachtergruppen bezüglich der Gesamtrangfolge (Rangsummenbildung)**

Vergleicht man die Beobachtergruppen bezüglich dieser Gesamtrangfolgen miteinander, stellt man fest, dass die AC-Assistenz zwar am niedrigsten mit allen anderen Beobachtergruppen korreliert und damit übereinstimmt, in Bezug auf die Gesamtrangfolge jedoch noch höhere Korrelationen aufweist als die Experten. Die Experten haben die

niedrigsten Übereinstimmungen mit der AC-Assistenz und die niedrigsten Korrelationen in Bezug auf die Gesamtrangfolge.

Eigentlich wäre zu erwarten gewesen, dass sie, wenn die Expertise als Beobachter die Beobachterqualität beeinflusst, eine höhere Übereinstimmung mit dem AC-Leiter und der AC-Assistenz und zugleich auch eine höhere Übereinstimmung mit der Gesamtrangfolge erzielen würden als die Novizen.

Mit diesem Datensatz ist die Schwierigkeit verbunden, dass die Beobachter zwar eine Rangfolge über alle Teilnehmer erstellen, aber nicht alle Teilnehmer in allen Übungen gesehen und/ oder beurteilt haben ("Zweiergespräch", "Stressinterview", "Gruppenübungen", vgl. die Ausführungen weiter oben). Von daher muss angenommen werden, dass es für die Beobachter nicht von Bedeutung war, ob die Teilnehmer unterschiedlichen Übungstypen ausgesetzt wurden – ihre Eindrucksbildung war davon möglicherweise nicht abhängig.

#### **4.5 Dieselben Forschungshypothesen - eine andere Stichprobe**

Nun könnte man die Frage stellen, ob nicht Überforderung oder Störeinflüsse, z.B. durch die große Zahl der Teilnehmer (maximal zwölf Teilnehmer in einem AC), die Beobachtgenauigkeit beeinflusst haben, oder ob die o.a. Ergebnisse ein Effekt des vorliegenden Datensatzes sind. Es liegt auf der Hand, dass es einen Unterschied für die Beobachter darstellt, ob sie eine Gruppenübung mit nur vier Teilnehmern beobachten, oder eine Gruppenübung mit zwölf Teilnehmern, von denen sie die vier Teilnehmer beobachten sollen, die sie selber am besten sehen können. Gleichwohl sammeln sie aber Eindrücke über die nicht beobachteten Teilnehmer.

Aus diesem Grunde wurden die eingangs erhobenen Hypothesen an einem Datensatz von insgesamt 26 Teilnehmern überprüft. Dieser Datensatz bezieht sich auf sechs AC, in denen maximal fünf Teilnehmer beobachtet wurden. Letztendlich flossen aber nur die Daten von 25 Teilnehmern auf Grund von missing values in den Datensatz ein. Bei diesem Datensatz wurden die Beobachtungen über die Teilnehmer vom AC-Leiter, der AC-Assistenz und von insgesamt 9 Experten und 15 Novizen erstellt (bezüglich der Unabhängigkeit der Stichproben vgl. die Ausführungen oben).

#### **4.5.1 Mittlere Korrelationen der Beobachtergruppen untereinander über die Übungen hinweg**

Zur Erinnerung noch einmal die Forschungshypothese 1. Sie lautet:

Wenn spezifische trainierbare Beobachterkompetenzen für die Qualität von AC-Ergebnissen ausschlaggebend sollen, dann müssen Unterschiede in der Beobachterübereinstimmung in Abhängigkeit vom Trainiertheitsgrad der Beobachter bestehen. Unternehmensinterne Führungskräfte und/oder Mitarbeiter, die wenigstens einmal bei einem AC als Beobachter dabei waren (= trainierte Experten) korrelieren dann mit einem vielfach trainierten AC-Leiter und einer gut trainierten AC-Assistenz höher als unternehmensinterne Mitarbeiter, die noch nie bei einem AC als Beobachter dabei waren (= untrainierte Novizen).

##### **4.5.1.1 Übungsbezogene mittlere Korrelationen zwischen AC-Leiter - AC-Assistenz, AC-Leiter - Experten und AC-Leiter – Novizen**

Tabelle 12 zeigt die übungsbezogenen mittleren Korrelation des AC-Leiters mit der AC-Assistenz, den Experten und den Novizen. Es zeigt sich, dass die AC-Assistenz mit dem AC-Leiter in allen fünf Übungen höhere Korrelationen aufweist als die Experten oder als die Novizen mit dem AC-Leiter; des weiteren zeigen die Experten (mit Ausnahme für die Übung "Zweiergespräch") mit dem AC-Leiter höhere Korrelationen als die Novizen. Dies deutet darauf hin, dass möglicherweise mit häufigerer Teilnahme als Beobachter an dem AC-Verfahren, unter der Voraussetzung, dass nur maximal 5 Teilnehmer beobachtet werden müssen, die Übereinstimmungen zwischen den verschiedenen Beobachtergruppen und dem AC-Leiter, der das Verfahren konstruiert hat, ansteigen (das gilt in gleichem Maße auch für die AC-Assistenz).

Die mittleren Korrelationen zwischen dem AC-Leiter und den anderen Beobachtergruppen wurden mit einem Lisrel-Modell (DOS Extender LISREL 8.12a) auf Signifikanz überprüft. Von einem signifikanten Unterschied ist dann auszugehen, wenn der nach dem o.a. errechnete Modell  $\chi^2$  Wert signifikant ist. Angesichts der geringen Fallzahl ( $n=25$ ) wird das Signifikanzniveau auf 20% gesetzt. Wie Tabelle 12 zeigt, konnte bei dem vorliegenden Datensatz kein signifikanter Unterschied zwischen den Beobachtergruppen festgestellt werden.

Übung	AC-Leiter und...			chi <sup>2</sup> (df=1)
	AC-Assistenz	Experten (n=9)	Novizen(n=15)	
Selbstpräsentation	.405	.350	.290	.061 (p = .81)
Gruppenübung 1	.405	.350	.245	.160 (p = .69)
Stressinterview	.470	.410	.385	.071 (p = .79)
Zweiergespräch	.255	.150	.215	.190 (p = .67)
Gruppenübung 2	.415	.300	.270	.200 (p = .65)

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 12: Übungsbezogene mittlere Korrelationen zwischen AC-Leiter und den anderen Beobachtergruppen**

Es fällt auf, dass in der Übung "Zweiergespräch" bei allen Beobachtergruppen die mittleren Korrelationen gegenüber denen der anderen Übungen deutlich geringer ausfallen. Wiederum ergeben sich für alle Beobachtergruppen für die Übung "Stressinterview" die höchsten Übereinstimmungen mit dem AC-Leiter (zur Erklärung vgl. die Ausführungen oben).

In Tabelle 12 fällt ferner auf, dass die Novizen niedrigere Korrelationen mit dem AC-Leiter (mit Ausnahme der Übung "Zweiergespräch") aufweisen als die Experten. Deutlich ist dabei, dass im Gegensatz zu der AC-Assistenz und den Experten in der zweiten AC-Übung, der Gruppenübung 1, die Korrelation gegenüber der ersten Übung niedriger ist. Möglicherweise deutet dies auf einen Leistungsabfall der Beobachtungsgüte hin, der damit erklärt werden kann, dass im Gegensatz zur "Selbstpräsentation", bei der nur ein Kandidat beobachtet wird, in der Gruppenübung auf einmal maximal fünf Teilnehmer beobachtet werden müssen, und die Novizen noch nicht so geübt in der Benutzung ihres Beobachtungs- und Beurteilungsinventars sind, wie Experten oder AC-Assistenz.

So sind auch die Übungen, in denen nur ein einzelner Kandidat beobachtet wird ("Selbstpräsentation" und "Stressinterview"), diejenigen, bei denen die höchsten Korrelationen mit dem AC-Leiter zu finden sind. Insofern kann man annehmen, dass die Anzahl der zu beobachtenden Teilnehmer einen Einfluss auf die Beobachtergüte bei den Novizen gehabt haben könnte.

#### 4.5.1.2 Übungsbezogene mittlere Korrelationen zwischen AC-Assistenz – Experten und AC-Assistenz - Novizen

Tabelle 13 zeigt die übungsbezogene mittlere Korrelation zwischen der AC-Assistenz und den Experten bzw. der AC-Assistenz und den Novizen. Danach ist zu entnehmen, dass die Korrelationen von Experten und Novizen in Bezug auf die AC-Assistenz ähnlich hoch sind (mit Ausnahme der Übung "Zweiergespräch"). Bei der Betrachtung der Korrelationskoeffizienten lassen sich die Muster feststellen, wie sie auch in den Korrelationen zwischen dem AC-Leiter und den Experten und Novizen vorzufinden sind. Lediglich für die zweite Gruppenübung zeigen die beiden Beobachtergruppen ein umgekehrtes Korrelationsmuster.

Übung	AC-Assistenz und...	
	Experten (n=9)	Novizen (n=15)
Selbstpräsentation	.355	.335
Gruppenübung 1	.330	.305
Stressinterview	.375	.320
Zweiergespräch	.135	.245
Gruppenübung 2	.410	.420

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 13: Übungsbezogene mittlere Korrelationen zwischen AC-Assistenz und den anderen Beobachtergruppen**

In dieser zweiten Gruppenübung haben die Experten und die Novizen in Bezug auf die AC-Assistenz mit .410 bzw. .420 höhere mittlere Korrelationen als in Bezug auf den AC-Leiter (vgl. Tabelle 12). Eine Signifikanzprüfung ergab für die Korrelationen mit dem größten Unterschied zwischen den Experten und den Novizen (hier bei der Übung "Zweiergespräch") keinen signifikanten Unterschied ( $z = -.23$ ). Insofern wird es auch für die anderen Korrelationsunterschiede, die gleich hoch oder niedriger sind, keine Signifikanzen geben.

#### 4.5.1.3 Übungsbezogene mittlere Korrelationen zwischen Experten und Novizen

Zwischen den Experten und den Novizen sind die mittleren Korrelationen ähnlich hoch wie zwischen der AC-Assistenz und den Novizen (mit Ausnahme der Übung "Stressinterview"; vgl. Tabelle 14). Möglicherweise greifen in der Übung "Stressinterview" die persönlichen berufsspezifischen Erfahrungen der Experten und der Novizen, die der AC-Leiter und die AC-Assistenz nicht haben können: in dieser Übung werden die Merkmale "Umgang mit belastenden Situationen", "Selbständiges Verhalten", "Engagement", "Geistige Flexibilität" und "Persönliches Auftreten" beobachtet.

Das spezifische Bild des Finanzdienstleisters im Außendienst wird davon geprägt, dass er dann erfolgreich ist, wenn er belastbar, selbständig und seinen Kunden mit einer bestimmten Form des persönlichen Auftretens ausgestattet gegenübertritt. Darauf hinarbeitend gibt es eine Fülle von Schulungen, in denen diese Eigenschaften gestärkt werden.

Übung	Experten/Novizen (n=9) / (n=15)
Selbstpräsentation	.300
Gruppenübung 1	.325
Stressinterview	.540
Zweiergespräch	.265
Gruppenübung 2	.410

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 14: Übungsbezogene mittlere Korrelation zwischen Experten und Novizen**

Von daher liegt der Verdacht nahe, dass in der Situation des "Stressinterview", das inhaltlich eine hohe Ähnlichkeit zu berufsfeldtypischen Situationen der unternehmensinternen Beobachter hat, diese nicht in der Lage sind, innerlich in die Distanz zu gehen. Vielmehr identifizieren sie sich mit den Teilnehmern und der Situation und begehen dadurch den Fehler, ihre eigenen Vorstellungen von Belastbarkeit, selbständigem Verhalten und persönlichem Auftreten als Maßstab für die Beurteilung heranzuziehen anstatt auf die Merkmale zu achten, die durch den Beobachtungsbogen vorgegeben werden.

#### **4.5.2 Mittlere Korrelationen der Beobachtergruppen untereinander über die Verhaltensmerkmale hinweg**

Gemäß dem oben beschriebenen Vorgehen beim Datensatz 1 wurde auch bei diesem Datensatz die mittlere Korrelationen der Beobachtergruppen untereinander über die Verhaltensmerkmale hinweg überprüft.

##### **4.5.2.1 Merkmalsbezogene mittlere Korrelationen zwischen AC-Leiter – AC-Assistenz, AC-Leiter - Experten und AC-Leiter - Novizen**

Tabelle 15 zeigt die merkmalsbezogenen mittleren Korrelationen zwischen den unterschiedlichen Beobachtergruppen bezogen auf die Verhaltensmerkmale. Ihr kann man entnehmen, dass die mittleren Korrelationen bezogen auf die einzelnen Merkmale zwischen dem AC-Leiter und der AC-Assistenz in sieben von dreizehn Fällen höher sind (mindestens um .10 höher) als zwischen den Experten und dem AC-Leiter. Im einzelnen handelt es sich um die Merkmale "Persönliches Auftreten", "Zwischenmenschliche Beziehungen gestalten", "Selbständiges Verhalten", "Zielorientierung", "Schrittweises und kontrolliertes Arbeiten" sowie "Geistige Flexibilität".

Lediglich im Punkt "Engagement" zeigt die AC-Assistenz eine niedrigere Übereinstimmung mit dem AC-Leiter als die Experten mit dem AC-Leiter.

Vergleicht man die mittleren Korrelationen der Experten in Bezug auf den AC-Leiter und die mittleren Korrelationen der Novizen in Bezug auf den AC-Leiter, so kann man feststellen, dass in den Merkmalen "Zwischenmenschliche Beziehungen gestalten", "Selbständiges Verhalten", "Feedback geben", "Problemlöseverhalten" sowie "Geistige Flexibilität" die Novizen höhere Korrelationen mit dem AC-Leiter aufweisen als die Experten.

Das Merkmal "Geistige Flexibilität" macht dabei den größten Unterschied aus, er liegt dort bei .225. Die Experten haben wohl in Bezug auf dieses Merkmal einen Blickwinkel, der sich von dem Blickwinkel der AC-Assistenz und dem der Novizen stark unterscheidet.

Größere Unterschiede in der Beobachterübereinstimmung mit dem AC-Leiter gibt für die anderen Beobachtergruppen noch beim Merkmal "Engagement": hier übertreffen in Bezug auf den AC-Leiter die mittleren Korrelationen der Experten die der Novizen und die der AC-Assistenz.

Merkmal	AC-Leiter und...			chi <sup>2</sup> (df=1)
	AC-Assistenz	Experten (n=9)	Novizen (n=15)	
Umgang mit belastenden Situationen	.465	.436	.385	.018 (p = .89)
Engagement	.315	.430	.285	.24 (p = .62)
Persönliches Auftreten	.545	.410	.410	.35 (p = .56)
Verbalisierung	.380	.325	.300	.043 (p = .84)
Überzeugungsfähigkeit	.195	.205	.165	.00 (p = 1.00)
Zwischenm. Beziehungen gestalten	.400	.210	.235	.58 (p = .45)
Selbständiges Verhalten	.450	.275	.290	.55 (p = .46)
Feedback geben	.215	.125	.135	.11 (p = .74)
Teambildung unterstützen	.480	.455	.305	.007 (p = .93)
Problemlöseverhalten	.360	.350	.405	.002 (p = .96)
Schrittweises u. kontrolliertes Arbeiten	.585	.375	.285	.91 (p = .34)
Zielorientierung	.375	.225	.190	.34 (p = .56)
Geistige Flexibilität	.270	.055	.270	.74 (p = .39)

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 15: Merkmalsbezogene mittlere Korrelation zwischen AC-Leiter und den anderen Beobachtergruppen**

Zwischen den Beobachtergruppen gibt es keine signifikanten Unterschiede. Allgemein liegen die Übereinstimmungen zwischen AC-Leiter und den anderen Beobachtergruppen deutlich unter den in der Literatur (vgl. Anhang 1, S. 123) berichteten Werten: die Experten korrelieren in vier Merkmalen über .40 mit dem AC-Leiter, die Novizen in zwei Merkmalen; insofern ist anzunehmen, dass die Beobachter eine genügend gute merkmalsbezogene Betrachtung der Teilnehmer vorgenommen haben.

Dabei muß man jedoch bedenken, dass die Datenbasis für diese Werte mit insgesamt 25 AC-Teilnehmern relativ gering ist.

#### 4.5.2.2 Merkmalsbezogene mittlere Korrelation zwischen AC-Assistenz - Experten und AC-Assistenz - Novizen

Ein ähnliches Bild zeigt sich, wenn man die merkmalsbezogenen mittleren Korrelationen von Experten und Novizen in Bezug auf die AC-Assistenz analysiert (vgl. Tabelle 16). Bezüglich der Merkmale "Umgang mit belastenden Situationen", "Engagement", "Persönliches Auftreten", "Überzeugungsfähigkeit", "Problemlöseverhalten" und "Zielorientierung" haben die Experten in Bezug auf die AC-Assistenz auch hier, wie in Be-

zug auf den AC-Leiter (vgl. Tabelle 15) die höchsten Übereinstimmungen.

Die höchsten mittleren Korrelationen der Novizen in Bezug auf den AC-Leiter beziehen sich bis auf das Verhaltensmerkmal "Umgang mit belastenden Situationen" auf andere Verhaltensmerkmale ("Persönliches Auftreten", "Problemlöseverhalten") als in Bezug auf die AC-Assistenz (hier sind es "Zwischenmenschliche Beziehungen gestalten", "Selbständiges Verhalten" und "Zielorientierung").

Der "Umgang mit belastenden Situationen" gehört sowohl in Bezug auf den AC-Leiter wie auch in Bezug auf die AC-Assistenz zu den Verhaltensmerkmalen mit der größten Beobachterübereinstimmung.

Merkmal	AC-Assistenz und ...	
	Experten (n=9)	Novizen (n=15)
Umgang mit belastenden Situationen	.430	.395
Engagement	.425	.270
Persönliches Auftreten	.405	.335
Verbalisierung	.125	.255
Überzeugungsfähigkeit	.190	.240
Zwischenm. Beziehungen gestalten	.305	.395
Selbständiges Verhalten	.320	.385
Feedback geben	.280	.325
Teambildung unterstützen	.330	.325
Problemlöseverhalten	.375	.315
Schrittweises u. kontrolliertes Arbeiten	.510	.290
Zielorientierung	.260	.370
Geistige Flexibilität	.090	.390

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 16: Merkmalsbezogene mittlere Korrelation zwischen AC-Assistenz und den anderen Beobachtergruppen**

Für die Verhaltensmerkmale "Verbalisierung", "Überzeugungsfähigkeit", "Zwischenmenschliche Beziehungen gestalten", "Selbständiges Verhalten", "Feedback geben", "Zielorientierung" und "Geistige Flexibilität" zeigen sich für die Novizen in Bezug auf die AC-Assistenz höhere Korrelationen als für die Experten in Bezug auf die AC-Assistenz. Es fällt auf, dass auch hier die Korrelationen zwischen den Experten und der AC-Assistenz bzw. den Novizen und der AC-Assistenz unter den in der Literatur berichteten Werten liegen (vgl. Anhang 1, S. 123).

Die Überprüfung der Übereinstimmung der Experten und Novizen mit der AC-Assistenz ergab keine signifikanten Unterschiede zwischen den beiden Beobachtergruppen (vgl. dazu die Ausführungen unter 4.4.1.2).

#### 4.5.2.3 Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen

Merkmals	Experten/Novizen (n=9) / (n=15)
Umgang mit belastenden Situationen	.440
Engagement	.360
Persönliches Auftreten	.550
Verbalisierung	.400
Überzeugungsfähigkeit	.230
Zwischenm. Beziehungen gestalten	.325
Selbständiges Verhalten	.450
Feedback geben	.375
Teambildung unterstützen	.375
Problemlöseverhalten	.355
Schrittweises u. kontrolliertes Arbeiten	.330
Zielorientierung	.330
Geistige Flexibilität	.340

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 17: Merkmalsbezogene mittlere Korrelation zwischen Experten und Novizen**

Betrachtet man Tabelle 17 fällt auf, dass Experten und Novizen mit Ausnahme des Verhaltensmerkmals "Persönliches Auftreten" (.550) in allen anderen Verhaltensmerkmalen mittelhohe Korrelationen zeigen. Die nächst höheren Übereinstimmungen gibt es dabei in Bezug auf "Selbständiges Verhalten" (.450) und "Umgang mit belastenden Situationen" (.440). Nicht geklärt werden kann die Frage, ob sich die Beobachter tatsächlich von den zu beobachtenden Verhaltensmerkmalen leiten ließen, oder ob sie lediglich die Überschriften nutzten, um sich dann bei ihren Beurteilungen an ihren spezifischen Auffassungen (vergleichbar mit impliziten Persönlichkeitsannahmen) zu orientieren.

#### 4.5.3 Korrelation der Beobachtergruppen innerhalb der Übungen

Die Differenzierungsfähigkeit der einzelnen Beobachtergruppen zwischen den einzel-

nen Verhaltensmerkmalen wurde auch für den Datensatz 2 an Hand der Höhe der mittleren Korrelationen innerhalb der Übungen überprüft. Bei den diesem Datensatz zu Grunde liegenden AC sind die Beobachter weniger stark belastet in den Einzelübungen (weil weniger Teilnehmer zu beobachten sind) und stärker belastet in den Gruppenübungen, wo wegen der geringeren Anzahl an Teilnehmern die Gruppenaktivität höher ist.

Forschungshypothese 2 lautet:

Die Korrelationen innerhalb der einzelnen AC-Übungen sind für die Experten geringer als für die Novizen, weil die Experten besser zwischen den Verhaltensmerkmalen differenzieren können als die Novizen (für die Novizen lässt sich ein Halo-Effekt nachweisen).

Unter dieser Bedingung zeigen sich für den AC-Leiter und die AC-Assistenz in der ersten Übung (Selbstpräsentation) niedrigere Korrelationen als bei den AC aus dem Datensatz 1. Bis auf die Übung "Stressinterview", bei der für die Novizen die mittleren Korrelationen aus dem Datensatz 2 gegenüber dem Datensatz 1 ansteigen, sinken für Experten und Novizen die mittleren Korrelationen in den Übungen ab, in denen mehrere Teilnehmer beobachtet werden.

	Selbstpräsentation	Gruppenübung 1	Stressinterview	Zweiergespräch	Gruppenübung 2
<b>AC-Leiter</b>	.34	.49	.55	.45	.56
<b>AC-Assistenz</b>	.29	.51	.56	.40	.59
<b>Experten</b>	.48	.58	.50	.26	.64
<b>Novizen</b>	.55	.67	.66	.47	.58

Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 18: Mittlere Korrelationen der Beobachtergruppen innerhalb der einzelnen Übungen**

Nach den vorliegenden Daten differenzieren AC-Leiter und AC-Assistenz in der Übung "Selbstpräsentation" am besten zwischen den einzelnen Verhaltensmerkmalen, die

Experten tun das vermutlich in der Übung "Zweiergespräch". Ansonsten ist in den Mehrpersonenübungen sowohl für die Experten wie auch für die Novizen eine geringere Höhe der mittleren Korrelationen als beim Datensatz 1 festzustellen.

Die Anzahl an beobachteten Fällen (der Datensatz 2 basiert auf 25 Teilnehmern) ist zwar gering, aber dennoch kann die Aussage gewagt werden, dass mit der Verringerung der AC-Teilnehmer vor allem in den Gruppenübungen die Qualität der Beobachtungen der Experten und der Novizen gegenüber Datensatz 1 (vgl. Tabelle 10) steigt.

#### **4.5.4 Eindrucksbildung über die Teilnehmer auf Grund von Rankings**

Auch bei dem vorliegenden, zweiten Datensatz wurden wieder gemittelte Korrelationen als Maß für die Übereinstimmung der einzelnen Beobachtergruppen mit der unabhängig von den Beobachterbögen erstellten Gesamtrangfolge berechnet (zum Vorgehen vgl. 4.3). Zur Erinnerung noch einmal Forschungshypothese 3:

Hat das Beobachtertraining einen Einfluss auf die Beobachterübereinstimmung, dann bilden die rankings trainierter Beobachter über die Teilnehmer genauer die Reihenfolge der Teilnehmer im AC ab als die rankings untrainierter Beobachter.

In Tabelle 19 ist zu erkennen, dass der AC-Leiter und die Experten die höchsten Übereinstimmungen mit der Gesamtrangfolge aufweisen, gefolgt von den Novizen und der AC-Assistenz. Die Höhe der mittleren Korrelationen ist dabei geringer als bei dem ersten Datensatz (vgl. Tabelle 11). AC-Assistenz und Experten weisen mit .65 und .68 vergleichbar hohe Übereinstimmungen mit dem AC-Leiter auf, während die Novizen hier nur eine Übereinstimmung von .38 erreichen. Die insgesamt niedrigere Höhe der Korrelationen zwischen den einzelnen Beobachtergruppen im Vergleich zum ersten Datensatz deutet darauf hin, dass sich die Beobachter in ihrer der Beurteilung dieser Teilnehmer, über die sie sich in jeder Übung einen Eindruck verschaffen konnten, voneinander unterscheiden.

	Gesamtrangfolge	AC-Leiter	AC-Assistenz	Experten
AC-Leiter	.84			
AC-Assistenz	.51	.65		
Experten (n=9)*	.75	.68	.19	
Novizen (n=15)*	.58	.38	.01	.21
		chi <sup>2</sup> (df=1) .0031 (p=.86)		

\* Datenbasis sind die Beobachtungen über 25 Teilnehmer

**Tabelle 19: Korrelation der Rangfolge über die Teilnehmer aufgrund des Gesamteindrucks, sowie Korrelation zwischen den Beobachtergruppen bezüglich der Gesamtrangfolge (Rangsummenbildung)**

Bei der Überprüfung der Übereinstimmung der Experten und Novizen mit der AC-Assistenz konnten keine signifikanten Unterschiede festgestellt werden ( $z = -.37$ ).

Als zweites Fazit kann festgehalten werden, dass die Beobachterübereinstimmungen gegenüber dem ersten Datensatz niedriger sind, signifikante Beobachterunterschiede aber nicht feststellbar waren. Möglicherweise hat bei diesen Ergebnissen die niedrigere Fallzahl gegenüber dem Datensatz 1 einen Einfluss.

## 5 Diskussion

Zu Beginn der vorliegenden Arbeit wurde die Frage aufgeworfen, ob es denn unter ökonomischen Gesichtspunkten nicht möglich sei, auf ein Beobachtertraining zu verzichten, zumal die von einem derartigen Training erwarteten Effekte nicht uneingeschränkt positiv bewertet werden können. Zur Prüfung dieser Frage wurde ein AC konstruiert, das eingesetzt wurde, um Außendienstmitarbeiter für ein Versicherungsdienstleistungsunternehmen zu rekrutieren. Das Beobachtungsinventar für dieses AC wurde mit Mitarbeitern des Unternehmens zusammen entwickelt, um dessen Normen und Wertvorstellungen abbilden zu können.

Die Übungsinhalte wurden so gestaltet, dass Thematiken aus dem Berufsalltag verarbeitet wurden. Die Hypothese war, dass, wenn ein Beobachtertraining einen Effekt hat, die Expertise der Beobachter (hier: AC-Leiter, AC-Assistenz, Experten und Novizen) nachweisbar sein muss. Diese Expertise beeinflusst zum einen den Grad der Beobachterübereinstimmung zwischen den einzelnen Beobachtergruppen, zum anderen die Güte des Zusammenhangs der Verhaltensmerkmale über die einzelnen Übungen des AC hinweg. Geht man davon aus, dass der AC-Leiter quasi als "Oberexperte" und Konstrukteur des Verfahrens die im Beobachtungsinventar und den Beobachterbögen vorgegebenen Verhaltensmerkmale und deren Operationalisierungen am besten, am genauesten beobachten kann, dann sollten bspw. von den anderen Beobachtern die Novizen am wenigsten mit diesen "Oberexperten" übereinstimmen, weil sie mit der AC-Thematik bisher noch nicht in Berührung gekommen waren. Genau das konnte in der vorliegenden Untersuchung nicht bestätigt werden.

Zwei Datensätze wurden untersucht: beim Datensatz 1 gab es Teilnehmer, die nicht von allen Beobachtern in jeder Übung beurteilt wurden (die fehlenden Beobachterdaten wurden mit den Beurteilungsdaten anderer Personen der selben Beobachtergruppe aufgefüllt, vgl. 3.8), beim Datensatz 2 handelt es sich um Beobachterdaten aus AC, in denen alle Beobachter alle Teilnehmer in jeder Übung beurteilt haben.

Zunächst wurde Datensatz 1 untersucht. Folgende Berechnungen wurden vorgenommen:

- übungsbezogene mittlere Korrelationen zwischen allen Beobachtergruppen
- merkmalsbezogene mittlere Korrelationen zwischen allen Beobachtergruppen
- mittlere Korrelationen innerhalb der Übungen für jede Beobachtergruppe

- Korrelation der Gesamtrangfolge über die Teilnehmer hinweg mit der von den jeweiligen Beobachtergruppen ermittelten Rangfolge.

Hinsichtlich der übungsbezogenen mittleren Korrelationen zwischen den Beobachtergruppen korrelierten in Bezug auf den AC-Leiter die verschiedenen Beobachtergruppen in den einzelnen Übungen mit diesem unterschiedlich hoch. Die AC-Assistenz und die Novizen korrelierten am höchsten mit dem AC-Leiter in der Übung "Zweiergespräch". Die AC-Assistenz weist zudem in dieser Übung mit dem AC-Leiter höhere Korrelationen auf als in allen anderen Übungen – hier treffen am stärksten gemeinsame Blickwinkel zusammen, begünstigt durch die Erfahrungen in der Zusammenarbeit mit dem AC-Leiter. In Anlehnung an die Befunde von Exler und Kleinmann (1997) kann vermutet werden, dass die AC-Assistenz sich vor allem in dieser Übung bemüht, die spezifischen Vorstellungen des AC-Leiters von dieser Übung zu erfüllen.

Die Experten korrelierten am höchsten mit dem AC-Leiter in der Übung "Stressinterview." Dort ist der gemeinsame Blickwinkel auch nahezu identisch mit dem Blickwinkel zwischen AC-Leiter und Novizen. Als Erklärung dafür mag die Überlegung gelten, dass in dieser Übung von den Beobachtern eine teilnehmende Beobachtung abgefordert wird: sie müssen sich aktiv, die Teilnehmer konfrontierend, in das Diskussionsgeschehen einschalten. Von dieser konfrontativen Grundhaltung (die Übung heißt ja "Stressinterview") müssen sie abweichen, wenn sie merken, dass der AC-Leiter oder die AC-Assistenz sich einschalten und sich von ihrem Befragungsstil her mehr wohlwollend / fürsorglich verhalten - was sie machen, wenn sie merken, dass der Teilnehmer die Situation nicht mehr gestalten kann. Da die Beobachter wissen, dass der AC-Leiter oder die AC-Assistenz ihr Verhalten nur in diesem Fall ändern, urteilen sie wahrscheinlich korrespondierend zur Intervention des AC-Leiters/der AC-Assistenz, dass der Teilnehmer „schwach“ ist. Insofern könnte es sich bei diesen übungsbezogenen Korrelationen um einen Versuchsleitereffekt handeln.

Andererseits sind die mittleren Korrelationen in Bezug auf den AC-Leiter für die Novizen in dieser Übung höher als in den beiden vorangegangenen Übungen – das erscheint plausibel, wenn man bedenkt, dass diese keinerlei Erfahrung mit dem "Instrument AC" hatten. Die ersten beiden Übungen könnten als Trainingsphase gewertet werden, in der die Novizen am "Life-Fall" lernen, Beobachterbögen zu führen und Beurteilungen unter Zeitdruck zu machen. Die höhere Korrelation in den Übungen "Zwei-

ergespräch" und "Gruppenübung 2" gegenüber den anderen Beobachtergruppen ist möglicherweise darauf zurückzuführen, dass die Novizen sich mehr bemühen, ihre Unsicherheit zu reduzieren. Daher nutzen sie die Beobachterbögen stärker als die Experten oder die AC-Assistenz, die sich sicherer fühlen und nähern sich im Laufe der zunehmenden Vertrautheit mit dem AC-System und mit der Beobachterrolle dem Blickwinkel des AC-Leiters als "Oberexperten" an.

In Bezug auf die AC-Assistenz zeigte sich, dass diese in der Übung "Stressinterview" die höchste gemeinsame Übereinstimmung mit den Experten, bzw. den Novizen hat, die sich in der Höhe ihrer mittleren Korrelationen kaum voneinander unterscheiden (.49, bzw. .50). Generell lassen die mittleren Korrelationen für die Experten, bzw. Novizen in Bezug auf die AC-Assistenz keine bedeutsamen Unterschiede untereinander erkennen. Das bedeutet, dass die Novizen in Bezug auf die AC-Assistenz eine Übereinstimmung vergleichbar mit der der Experten hat.

Auch bei der Betrachtung der mittleren Korrelationen zwischen Experten und Novizen fällt auf, dass diese Beobachtergruppen in der Übung "Stressinterview" die größte Übereinstimmung haben. Die Beobachterübereinstimmungen von Experten und Novizen sind dabei auf vergleichbarer Höhe wie die Beobachterübereinstimmungen zwischen AC-Leiter und Experten/Novizen, bzw. AC-Assistenz und Experten/Novizen. Würde die Expertise der Beobachter eine Auswirkung auf die Beobachterübereinstimmung haben, dann wäre zu erwarten, dass die Novizen mehr fehlerbehaftete Beurteilungen vornehmen und deswegen weniger mit den anderen Beobachtergruppen übereinstimmen. Das scheint jedoch nicht der Fall zu sein: die Höhe der mittleren Korrelationen legt dies nicht nahe. Es gibt im Gegenteil signifikante Unterschiede dahingehend, dass die Experten signifikant niedriger mit dem AC-Leiter korrelieren als die Novizen (zur Problematik der Signifikanzprüfung vgl. die Ausführungen unter 3.9).

Forschungshypothese 1 kann daher in dieser Phase der Untersuchung nicht bestätigt werden, da sich die unterschiedlichen Expertisen der Beobachter nicht nachweisen lassen. Die Korrelationen sind mittelhoch; auch gibt es wie in der Literatur berichtet keine Werte von .60 oder mehr (vgl. z.B. Guldin, 1996, Anhang 1). Es scheint Übungen zu geben, die aufgrund ihres übungsspezifischen Charakters eine deutliche gemeinsame Eindrucksbildung für alle Beobachtergruppen ermöglichen (z.B. das Stressinterview). Möglicherweise wird diese Übung als die Übung schlechthin angesehen, weil

der Übungsinhalt die Berufsrealität treffend abbildet, dadurch das Erfahrungswissen der Beobachter nachhaltig angesprochen wird und implizite Persönlichkeitstheorien der unternehmensinternen Beobachter (Experten/Novizen) zum Tragen kommen.

Offensichtlich gibt es auch für die anderen Übungstypen Übereinstimmungen, die, betrachtet man die Höhe der mittleren Korrelationen, den Schluss zulassen, dass die Beobachter differenziert und nicht global beobachtet haben.

Die Analyse der merkmalsbezogenen mittleren Korrelationen zwischen dem AC-Leiter und den anderen Beobachtergruppen zeigt auf, dass auch hier die Beobachtergruppen bezüglich der Verhaltensmerkmale unterschiedliche hohe Übereinstimmungen mit diesem haben. Es gibt Verhaltensmerkmale, für die die Übereinstimmung unabhängig von der Beobachtergruppe niedrig ist (z.B.: Feedback geben) und umgekehrt, und es gibt bestimmte Verhaltensmerkmale, in denen eine Beobachtergruppe höher als die anderen Beobachtergruppen mit dem AC-Leiter übereinstimmt (z.B.: Zielorientierung für die Experten). Verantwortlich dafür sind die Übereinstimmung im Erfahrungswissen, die sich hierin niederschlägt, möglicherweise weniger gut definierte und eindeutig operationalisierte Verhaltensmerkmale, Übungen und/oder Teilnehmerzusammensetzungen, die es nicht ermöglichten, entsprechendes Zielverhalten zu zeigen.

Vor allem für die schwieriger zu beobachtenden Verhaltensmerkmale ist das Erfahrungswissen im Bezug auf den Umgang mit dem AC-System wichtig.

Für das Merkmal "Zielorientierung" z.B. beträgt in Bezug auf den AC-Leiter die Beobachterübereinstimmung mit der AC-Assistenz .20, mit den Experten .42 und mit den Novizen .30. Die AC-Assistenz hat möglicherweise am wenigsten Zugang zu Führungsthemen (zur Erinnerung: "Zielorientierung" wurde nur in den AC beobachtet, in denen Vertriebsleiter als potenzielle Führungskräfte ausgewählt wurden). Experten, die schon einmal am AC teilgenommen hatten, wurden schon einmal bezüglich ihrer Beobachtertätigkeit evaluiert – sie erlebten direktes und indirektes Feedback im Rahmen der Auseinandersetzung über ihre Beurteilungen während ihrer ersten AC-Teilnahme und kannten daher den Blickwinkel des AC-Leiters zum Thema "Mitarbeiterführung". Da sie zudem auch selber als Führungskräfte über Erfahrungswissen verfügten, konnten sie den Erkenntnisgewinn aus der Reflexion über ihre Sichtweise beim nächsten AC umsetzen und hatten daher eine bessere Übereinstimmung mit dem AC-Leiter.

Vielleicht ist die stärkere Übereinstimmung mit dem AC-Leiter zusätzlich darauf zurückzuführen, dass unter den Experten einige Beobachter waren, die schon als Novizen spontan gut beobachten konnten (die Höhe der mittleren Korrelationen von .30 für die Novizen in diesem Merkmal legt das nahe), und die von sich aus Feedback aufsuchten, um ihre Beobachterqualität zu erhöhen (weniger gute Beobachter würden aus Selbstwertschutzgründen diesen Weg nicht gehen). Allein die erneute Teilnahme an einem AC als Beobachter, bei der die einmal gemachte Beobachtererfahrung wiederholt wird, führt aller Voraussicht nach noch nicht zu einer Verbesserung der Ergebnisse.

Auch hier muss Forschungshypothese 1 verworfen werden, weil die Beobachterexpertise, die sich in den Höhen der Korrelationskoeffizienten niederschlagen sollte, nicht nachgewiesen werden konnte. Sowohl in Bezug auf den AC-Leiter, wie auch in Bezug auf die AC-Assistenz läßt die Höhe der ermittelten merkmalsbezogenen mittleren Korrelationen für Experten und Novizen nur sporadisch signifikante Unterschiede erkennen. Ob die durch das AC gemachte "on-the-job-Erfahrung" einen Einfluss haben könnte, kann daher für diesen Datensatz nicht eindeutig beantwortet werden.

Hinsichtlich der mittleren Korrelationen innerhalb der Übungen gibt es eine Bestätigung der Befunde von Sackett und Dreher (1982), die in den Gruppenübungen bei ihrer Studie 3 die Korrelationen über die Verhaltensmerkmale hinweg erhoben und dabei auf Werte von .69, bzw. .77 kamen. Die Werte sind ähnlich hoch wie die Werte der Experten und der Novizen (in dieser Studie liegen die Korrelationen bei .72 in beiden Gruppenübungen). Auch die von Sackett und Dreher (1982) berichtete Höhen des Korrelationskoeffizienten für das Interview von .47 ist vergleichbar mit dem für die Novizen ermittelten Wert von .45 für das Stressinterview (die Experten haben mit .57 einen höheren Wert); lediglich der in Studie 1 (vgl. Sackett und Dreher, 1982) berichtete Wert für die Präsentation weicht mit .71 von den Werten der Experten und Novizen mit .47, bzw. .58 ab. Die Beobachter in den Studien von Sackett und Dreher (1982) waren trainiert; ihre Diskriminanzfähigkeit der einzelnen Verhaltensmerkmale der jeweiligen Übungen ist jedoch ähnlich der Diskriminanzfähigkeit der untrainierten Novizen in der vorliegenden Untersuchung.

Es fällt auf, dass die Experten der vorliegenden Studie in den Gruppenübungen ähnlich globale Einschätzungen über die Teilnehmer abgeben wie die Novizen. Dies ist ein

Hinweis dafür, dass allein ein Erfahrungswissen im Umgang mit der AC-Situation und den Beobachterbögen nicht ausreicht, um besser zwischen den Verhaltensmerkmalen diskriminieren zu können. Innerhalb der Übung "Selbstpräsentation", sind die Einschätzungen der Experten differenzierter; dort weisen sie ähnlich hohe mittlere Korrelationen (.47) auf wie der AC-Leiter (.45). Möglicherweise können erfahrenere Beobachter (AC-Leiter, AC-Assistenz, Experten) in dieser Übung, in der Beurteilungen in einem engen Zeitfenster von 3-5 Minuten abgegeben werden müssen, besser diskriminieren als die Novizen, weil sie mit der Situation und dem Verfahren schon vertraut sind, während die Novizen noch lernen müssen. Andererseits kann vermutet werden, dass sich die Experten zu sicher fühlen und im weiteren Verlauf des AC zu sehr von ihren eigenen Beobachtungskriterien und zu wenig von den Verhaltensmerkmalen der Beobachterbögen leiten lassen und von daher ähnlich ungenau beurteilen wie die Novizen.

Generell steigen die mittleren Korrelationen innerhalb der Übungen bei AC-Assistenz, Experten und Novizen in den Übungen an, in denen mehr als ein Teilnehmer zu beobachten ist (für die AC-Assistenz und den AC-Leiter liegen die Korrelationen innerhalb der Übung für alle Übungstypen unter den von Sackett und Dreher (1982) berichteten Werten).

In der Übung "Stressinterview", der dritten Übung des AC, haben die Novizen niedrigere Korrelationen als in den anderen Übungen, und in der Übung "Zweiergespräch" beobachten sie sogar differenzierter als die Experten. Ihre bis zu diesem Zeitpunkt gemachten Erfahrungen im Umgang mit dem AC-System führen wahrscheinlich dazu, dass sie beobachtetes Verhalten einer einzelnen Person oder von zwei Personen an Hand von Beobachterbögen differenziert beurteilen können. In den Gruppenübungen beobachten sie ähnlich differenziert wie die Experten. Das bedeutet, dass die Reflexionen über das AC, die die Experten schon in den Beobachterkonferenzen gemacht haben hinsichtlich der Gruppenübungen keinen Effekt in Bezug auf eine Verbesserung der Beobachtergenauigkeit für die Experten gehabt hatten. Für die Gruppenübungen scheint das Erfahrungswissen der Beobachter nur dann bedeutsam zu sein, wenn sie eine Expertise wie die AC-Assistenz oder der AC-Leiter haben.

Mit Blick auf Forschungshypothese 2 muss bestätigt werden, dass Beobachter wie der AC-Leiter oder die AC-Assistenz innerhalb der Übungen über die Verhaltensmerkmale hinweg niedrigere Korrelationen aufweisen als die anderen Beobachtergruppen, für die unternehmensinternen Experten betätigt sich diese Annahme gegenüber den Novizen

jedoch nicht (ausgenommen in der Übung "Selbstpräsentation").

Von daher muss angenommen werden, dass die Beobachter, die noch kein AC als Beobachter erlebt haben, innerhalb der Übungen die einzelnen Verhaltensmerkmale genauso gut und/oder schlecht differenzieren wie Beobachter (Experten) mit dem größeren Erfahrungswissen. Es scheint übungsabhängig für die einzelnen Beobachtergruppen Schwerpunkte zu geben. Die Beobachtung von Beutel (1983), der feststellte, dass die Dimensionen in den Mehrpersonenübungen höher untereinander korreliert sind als bei den Einzelübungen oder den Zweierbeziehungen, muss für diesen Datensatz betätigt werden.

Die Höhe der mittleren Korrelationen innerhalb der Übungen (Forschungshypothese 2) liegt für die Experten und für die Novizen beim Datensatz 1 generell höher als für die AC-Assistenz oder den AC-Leiter (vgl. Tabelle 10) so dass hier der Einfluss der unterschiedlichen Expertisen vermutet werden könnte - Experten und Novizen haben einen größeren Halo-Effekt als AC-Assistenz oder AC-Leiter, die differenzierter beobachten können, weil sie in der Personalbeurteilung mit dem AC-System vertrauter sind. Wahrscheinlich können sie auch besser als Experten und Novizen damit umgehen, dass sie mehrere Teilnehmer beobachten müssen. Beim Datensatz 1 gibt es maximal 12 Teilnehmer, die die Beobachtung beeinflussen, auch wenn sie nicht alle, wie beispielsweise in den Gruppenübungen, beurteilt werden (es werden dort immer nur 4 Teilnehmer beurteilt). Es ist anzunehmen, dass durch die Teilnehmer, die nicht beobachtet werden, Störeffekte ausgelöst werden. Zudem gibt es, bedingt durch die Gruppengröße, spezifische gruppenspezifische Prozesse (bei 8 Teilnehmern kommt statistisch gesehen der einzelne Teilnehmer weniger oft zu Wort etc.), die die Urteilsbildung beeinträchtigen. In diesem Moment der Verunsicherung konzentrieren sich die Experten und die Novizen weniger auf die Operationalisierungen, die anhand der Beobachterbögen nahegelegt werden, sondern eher auf die Oberbegriffe der Operationalisierungen und dabei darauf, was sie individuell darunter verstehen - und differenzieren sie ungenauer.

Dieses Vorgehen führt möglicherweise dazu, dass relativ schnell für die Experten die Urteile über einen Teilnehmer, vielleicht sogar schon nach der zweiten Übung, feststehen. Sie fühlen sich sicherer, da sie sich in ihren eigenen Beobachungskriterien bestätigt fühlen, beobachten also individueller (undifferenzierter gemäß dem Beobachtungsinventar), und haben daher weniger hohe mittlere Korrelationen bezogen auf den AC-Leiter oder die AC-Assistenz, die mehr AC-systemkonformes Beobachterverhalten

zeigen.

Es muss zudem berücksichtigt werden, dass möglicherweise Experten und Novizen mehr Probleme haben, die vierstufige Beurteilungsskala voll auszuschöpfen als AC-Assistenz und AC-Leiter. Rückmeldungen von Experten und Novizen deuten darauf hin, dass deren Maßstäbe für die Vergabe der Beurteilung "sehr oft beobachtbar (4)" so hoch waren, dass sie diese Beurteilung kaum vergeben haben. Mit einer Einschränkung der Beurteilungsskala verändern sich auch die Beurteilungen, so dass gewöhnt werden kann, die Beobachtungen wären undifferenziert. Das Problem ist jedoch die Handhabung der Beurteilungsskala in Abhängigkeit vom eigenen Beurteilungsmaßstab.

Des Weiteren kommt dazu, dass spezifische Erwartungen der Experten an die seitens der Teilnehmer zu leistende inhaltliche Gestaltung der Übungen den Blickwinkel der Experten einfärben. Wenn man vermutet, dass die Experten die Übung "Zweiergespräch" in erster Linie vor dem Hintergrund eines potenziellen Verkaufsgesprächs sehen, und daher in ihrem Blickwinkel eher von ihren Erwartungen an die Übung geprägt wurden als von dem Gedanken, Teilnehmer in einer spezifischen Situation bezüglich spezifischer Verhaltensoperationalisierung zu beobachten, dann kann das ebenfalls eine Erklärung dafür sein, warum die mittleren Korrelationen der Experten für die Übung "Zweiergespräch" deutlich ansteigen.

Bei der Betrachtung der Übereinstimmung der von den einzelnen Beobachtergruppen erstellten Rangfolge mit der Gesamtrangfolge über alle Teilnehmer stellt sich heraus, dass die Experten die geringste Übereinstimmung mit dieser Gesamtrangfolge haben. Novizen und Experten haben in Bezug auf den AC-Leiter mit .80, bzw. .76 eine nahezu identische Übereinstimmung, und auch die Übereinstimmung mit der AC-Assistenz lässt keinen statistisch bedeutsamen Unterschied vermuten. Offensichtlich hat die Beobachterexpertise bei der Ermittlung des Gesamteindrucks weniger Einfluss als das Erfahrungswissen in der Personalauswahl. Interessant sind diese Befunde vor allem deswegen, weil bei der Ermittlung der Übereinstimmung der von den Beobachtergruppen ermittelten Rangfolge mit der Gesamtrangfolge im Gegensatz zum Vorgehen bei den Gruppenübungen und beim Zweiergespräch hier keine Daten von anderen Beobachtern genutzt wurden, um den Datensatz 1 aufzufüllen. Es wurden die unvermischten Daten der an Stelle eins und Stelle zwei des Basisdatensatzes stehenden Exper-

ten, bzw. Novizen genommen. Daher beziehen sich diese Daten auf die Einschätzung von Teilnehmern, die die Beobachter entweder in einer oder in beiden Gruppenübungen gesehen hatten (abhängig von der Teilnehmerzahl) oder im Stressinterview und nicht im Zweiergespräch oder umgekehrt. Es fehlen demnach Beobachterinformationen aus Übungen - und trotzdem ist die Übereinstimmung der Experten und Novizen in Bezug auf den AC-Leiter, die AC-Assistenz und auch untereinander hoch.

Eine Erklärung für dieses Phänomen besteht darin, dass Experten und Novizen möglicherweise auf ein gemeinsames Erfahrungswissen im Bereich der Personalauswahl und/oder Einschätzung von Menschen im Versicherungsaußendienst verfügen, das sie befähigt, über Bewerber und deren Berufseignungspotenziale ähnlich zu urteilen. Erfahrungswissen ist Wissen, das im praktischen Handeln erworben und angewandt wird. Es ist personengebunden und auf den situativen Kontext bezogen (vgl. Straka, 2000). Nach einem traditionellen Verständnis ist ein Kennzeichen von Erfahrungswissen dessen Objektivierbarkeit, seine rationale Begründung und die auf Wissen beruhende Planung des Handelns (vgl. Fischer, 2000). Nach einem erweiterten Verständnis von Erfahrungswissen müssen Aspekte berücksichtigt werden, die Volpert (1999) als intuitiv-improvisierendes Handeln bezeichnet. Demnach sind auch Empfindungen und sinnliche Wahrnehmungen von Bedeutung für die Anwendung des Erfahrungswissens. Diese Facetten sind nicht objektivierbar.

Bezogen auf die hohe Übereinstimmung zwischen Experten und Novizen im ranking über die Teilnehmer ist anzunehmen, dass die Erfahrung mit den betrieblichen Gegebenheiten der Arbeitssituation im Versicherungsaußendienst und ein durch die hohe Fluktuation von Mitarbeitern im Versicherungsaußendienst geprägtes Gespür für "gute" Mitarbeiter bewirken, dass intuitiv die "richtigen" Mitarbeiter identifiziert werden. Das würde bedeuten, dass die Übungssituationen im AC realistisch genug sind, um einen Teil der Arbeitsrealität abzubilden. Dadurch ist es den Beobachtern möglich, ihr Erfahrungswissen zu aktivieren und für den Personalbeurteilungsprozess zu nutzen. Das Erfahrungswissen ist von größerer Bedeutung als die zu beobachtenden Verhaltensmerkmale der Beobachterbögen (Fixierung auf das Vorwissen).

Damit ist erklärbar, warum Experten, die eigentlich eine höhere Expertise haben müssten als Novizen, innerhalb der Übungen ähnlich gut oder schlechter als diese zwischen den Verhaltensmerkmalen diskriminieren, und trotzdem, wie die Novizen auch,

eine hohe Übereinstimmung mit dem AC-Leiter haben. Andererseits wäre auch denkbar, dass die AC-Teilnahme, die zu der Einteilung in die Kategorie "Experten" führt, keinen Effekt hinsichtlich einer Verbesserung der Beobachterexpertise gehabt hat.

Möglicherweise ist Erfahrungswissen für die Novizen bedeutsamer als für die Experten, denn die Novizen stimmen höher als die Experten mit der Gesamtrangfolge über die Teilnehmer überein. Geht man davon aus, dass die Novizen zu Beginn des AC mit ihrer Beobachteraufgabe überfordert sind, dann müssen sie sich Anker suchen, die ihnen Sicherheit geben. Diese Anker kann das Erfahrungswissen bieten. Es ist zu vermuten, dass gute Beobachter versuchen, die Beobachtungen aus den AC-Übungen, die durch die Beobachterbögen gesteuert werden, und das Erfahrungswissen miteinander zu verbinden um zu guten Entscheidungen zu gelangen. Einmal bei einem AC dabeigewesen, glauben die ehemaligen Novizen, die nun als Experten aufgeführt werden, zu wissen, "wie das AC funktioniert" (meta-kognitive Kontrolle). Dadurch sind sie nachlässiger in ihrer Aufmerksamkeit – vielleicht, weil sie auch gelernt haben, dass für die Auswahlentscheidung nicht ihr individuelles Urteil sondern das Urteil der gesamten Beobachtergruppe entscheidend ist, und kommen dann sogar zu schlechteren Ergebnissen als die "frisch motivierten" Novizen.

Für den Datensatz 2 wurden die selben Analysen vorgenommen wie für den Datensatz 1. Im Unterschied zu diesem handelt es sich hier um AC, in denen alle Beobachter alle Teilnehmer in jeder Übung beurteilt haben. Bei der Analyse der übungsbezogenen mittleren Korrelationen zwischen dem AC-Leiter und den anderen Beobachtergruppen stellte sich heraus, dass die AC-Assistenz hier die höchste Übereinstimmung mit dem AC-Leiter aufweist, dann kommen die Experten und dann die Novizen (einzige Ausnahme: die Übung Zweiergespräch; hier liegen die mittleren Korrelationen der Novizen in Bezug auf den AC-Leiter höher als die der Experten). Dies könnte ein Hinweis darauf sein, dass die unterschiedliche Expertise der Beobachter ihren Niederschlag findet. Allerdings ist dieser Unterschied nicht signifikant.

Die höchste Übereinstimmung der Beobachtergruppen untereinander besteht in Bezug auf das Stressinterview; damit ist eine Parallele zu den Ergebnissen aus Datensatz 1 gegeben. Diese Übung scheint von zentraler Bedeutung für die übereinstimmende Urteilsbildung zu sein.

Der Abfall der Beobachterübereinstimmung in der Übung "Zweiergespräch" wird möglicherweise durch die gegenüber den anderen Übungen veränderte Gestaltung der Beobachterbögen hervorgerufen: im Unterschied zu den anderen Übungen war der Beobachtungsbogen für das Zweiergespräch so aufgebaut, dass jeweils auf einem Bogen beide Teilnehmer beurteilt werden mußten. Dadurch könnten sich die Beobachter verleitet gefühlt haben, die Teilnehmer gegeneinander zu gewichten und dies in ihren Beurteilungen auch auszudrücken. Dies ist natürlich leichter möglich, wenn man zwei Beurteilungsreihen auf einem Formular nebeneinander stehen hat, als wenn man im Verlauf seiner Urteilsbildung zwischen mehreren Formularen hin und her blättern muss.

Andererseits gilt dieses Vorgehen auch für Datensatz 1 und dort sind die Ergebnisse so, dass die Übereinstimmungen höher sind. Entweder haben die Beobachter des Datensatzes 2 schlechter beobachtet, oder die Teilnehmer waren schlechter und haben es den Beobachtern schwerer gemacht, zu (übereinstimmenden) Beobachterurteilen zu kommen.

Es kann auch sein, dass diese Übung am ehesten geeignet ist, einen Teil der Berufsrealität (Vertriebstätigkeit im Finanzdienstleistungsgewerbe) widerzuspiegeln; hier handelt es sich, wie im Berufsleben auch, um eine Zweiergesprächssituation, die starken Verkaufscharakter hat. Die Experten haben hier möglicherweise ihre spezifischen Erwartungen an das Verhalten eines Versicherungsaußendienstmitarbeiters in einer derartigen Situation ("...man muss den Kunden etwas verkaufen...") auf diese Zweiergesprächssituation übertragen und sind in Bezug auf diese Erwartungen von den Teilnehmern enttäuscht wurden. Dass dieser Effekt bei den Novizen nicht auftrat liegt möglicherweise daran, dass diese noch unvorbelastet (sie konnten sich die Situation nur schwer vorstellen) in diese Übung hineingegangen sind, und sich konsequent an den Beobachterbögen orientiert haben.

Eine weitere Erklärung für die oben beschriebenen Ergebnisse bietet Kleinmann (1997) an, der herausfand, dass Teilnehmer in einem AC schlechter beurteilt werden, wenn sie die relevanten Anforderungsdimensionen der Übung nicht erkennen. Dieses Phänomen könnte auch hier eingetreten sein und dazu geführt haben, dass die Teilnehmer sich so verhalten haben, dass sie von AC-Assistenz, Experten und Novizen (vorausgesetzt, der AC-Leiter hat die größte Beurteilerkonstanz) schwer eingeschätzt werden konnten, weil zufällig diese Teilnehmer die Verhaltensmerkmale, die ihnen am

Vorabend des AC in der Vorbereitung auf den AC-Tag präsentiert wurden, den Übungen nicht mehr gut zuordnen konnten.

Zusammengefasst deuten die Ergebnisse darauf hin, dass sich unabhängig vom Übungstyp mit zunehmender Beobachterexpertise im Rahmen eines AC die Übereinstimmung mit dem AC-Leiter, der als "Oberexperte" das AC konzipiert hat, erhöht. Die vorliegende Untersuchung liefert Hinweise dafür, dass mit zunehmender Übung die Beurteilungs- und Beobachtungsqualität der Beobachter in einem AC steigt. Gleichwohl führt eine höhere Beobachterexpertise aller Voraussicht nach nicht zu signifikanten Unterschieden zwischen den Beobachtergruppen – der durch die AC-Teilnahme vermittelte Erfahrungsgewinn bietet den Experten keine Vorteile gegenüber den Novizen.

Die übungsbezogenen mittleren Korrelationen zwischen der AC-Assistenz und Experten bzw. Novizen lassen den Schluss zu, dass in Bezug auf die AC-Assistenz sowohl Experten wie auch Novizen ähnlich gut beobachten. Im Unterschied zu Datensatz 1 ist die Übereinstimmung zwischen Experten AC-Assistenz bzw. Novizen und AC-Assistenz ähnlich hoch. Auch hier ist ein Rückgang in der Übereinstimmung bei der Übung "Zweiergespräch" und eine stärkere Übereinstimmung mit der AC-Assistenz in der zweiten Gruppenübung gegenüber den anderen Übungen festzustellen. Novizen beobachten ähnlich gut wie die Experten in Bezug auf die AC-Assistenz. Gegenüber der vergleichbaren Auswertung in Datensatz 1 ist die Höhe der Korrelationen weniger stark übungsabhängig. Es ist denkbar, dass die geringere Anzahl an zu beobachtenden Teilnehmern und die damit verbundene weniger starke Belastung der Beobachter dazu geführt hat, dass sie sich besser/konsequenter an den Beobachterbögen orientiert haben.

Auch zwischen Experten und Novizen sind die übungsbezogenen mittleren Korrelationen vom Muster her vergleichbar mit den mittleren Korrelationen von Experten oder Novizen in Bezug auf die AC-Assistenz. Auffällig ist in der Übung "Stressinterview" der starke Anstieg der Übereinstimmung. Es ist denkbar, dass die Reduktion der zu beobachtenden Teilnehmer und die damit verbundenen geringere Aufmerksamkeitsbelastung für die Beobachter von diesen nicht dazu genutzt wurde, sich noch mehr auf die zu beobachtenden Verhaltensmerkmale (vorgegeben durch die Beobachterbögen) zu konzentrieren. Andererseits könnte auch schlechter beobachtbares Verhalten der Teil-

nehmer (z.B. verursacht durch ein weniger aktives Verhalten in der Übungssituation), dazu geführt haben, dass die Beobachter die Teilnehmer vor dem Hintergrund ihrer berufsspezifischen Erfahrungen bewerten (diesen Effekt nannten Klimonski und Brickner, 1987, "indirekte Kriterienkontamination"; sie deuteten ihn als Erklärung für die hohe prädiktive Validität bei geringer Konstruktvalidität des AC).

Hinsichtlich der merkmalsbezogenen mittleren Korrelationen läßt sich kein signifikanter Unterschied zwischen den Beobachtergruppen in bezug auf den AC-Leiter feststellen. Augenscheinlich hat die AC-Assistenz die größte Übereinstimmung mit dem AC-Leiter, und es gibt einzelne Merkmale, für die die Übereinstimmung mit dem AC-Leiter in Abhängigkeit von der Beobachtergruppe größer ist. Bezüglich des Merkmals "Geistige Flexibilität" gibt es für die Experten in Bezug auf den AC-Leiter und auch in Bezug auf die AC-Assistenz die geringsten Übereinstimmungen (möglicherweise können es die Beobachter am schlechtesten beobachten/beurteilen), und in Bezug auf die AC-Assistenz sticht das Merkmal "Schrittweises und Kontrolliertes Arbeiten" mit .51 deutlich hervor. Die Höhen der Korrelationskoeffizienten bewegen sich generell bei allen Kombinationen der Beobachtergruppen auf dem Niveau der mittelhohen Übereinstimmung, so dass die Übereinstimmung der Beobachter als Ausdruck einer globalen Eindrucksbildung ausgeschlossen werden kann.

Auch in Bezug auf den Datensatz 2 muss die Forschungshypothese 1 verworfen werden – es gibt keine signifikanten Unterschiede in der übungs- und/oder merkmalsbezogenen Beobachterübereinstimmung zwischen den einzelnen Beobachtergruppen.

Die mittleren Korrelationen innerhalb der Übungen beim Datensatz 2 sind für die Experten und Novizen vor allem in den Mehrpersonenübungen niedriger als beim Datensatz 1. Die Novizen haben im Unterschied zum Datensatz 1 beim Datensatz 2 höhere Korrelationen in der Übung "Stressinterview". Sie machen hier möglicherweise einen größeren Beobachterfehler. Die Experten haben beim Datensatz 2 in der Übung "Zweiergespräch" niedrigere Korrelationen als beim Datensatz 1. Da sich ihre Korrelationshöhen beim Datensatz 2 deutlich von denen der anderen Beobachtergruppen unterscheiden, ist anzunehmen, dass die Experten eine bessere Differenzierung zwischen den einzelnen Verhaltensmerkmalen vornehmen. Da der AC-Leiter quasi als Oberexperte diese Art der Differenzierung offenbar nicht vornimmt, und die AC-Assistenz und die Novizen ähnlich hohe Korrelationskoeffizienten haben wie der AC-

Leiter, ist zu vermuten, dass die Experten zufällig völlig anders beobachten als die anderen Beobachtergruppen.

Die Novizen zeigen in den beiden ersten Übungen des Datensatzes 2 eine ähnliche Diskriminierungsfähigkeit wie beim Datensatz 1, erreichen sich aber in den letzten beiden Übungen das Niveau von AC-Assistenz und AC-Leiter. Die geringere Anzahl an zu beobachtenden Teilnehmern scheint dahingehend zu wirken, dass mit zunehmender Dauer des AC Ermüdung weniger stark eintritt, die Novizen ihre Lernerfahrungen im Umgang mit dem AC-System besser umsetzen und ihre Beobachterrolle besser wahrnehmen können als beim Datensatz 1, wo die zu beobachtenden Teilnehmer schon allein wegen der größeren Teilnehmergruppe weniger oft zu Wort kommen können und daher auch für Ungeübte schlechter zu beobachten sind.

Insgesamt deuten die Ergebnisse an, dass auch für den Datensatz 2 Forschungshypothese 2 verworfen werden muss: eine höhere Beobachterexpertise ist nicht feststellbar.

Bezüglich der Gesamtrangfolge ist die Übereinstimmung von AC-Leiter und Experten beim Datensatz 2 ähnlich wie beim Datensatz 1. Die Übereinstimmung der einzelnen Beobachtergruppen untereinander unterscheidet sich jedoch erheblich von der Übereinstimmung gemäß Datensatzes 1. Die Übereinstimmung von AC-Assistenz und Novizen mit der Gesamtrangfolge ist bspw. schlechter als beim Datensatz 1. Es ist anzunehmen, dass sich bei der geringeren Anzahl von Teilnehmern beim Datensatz 2 zwar die Beobachterübereinstimmung an Hand der Beobachterbögen für Beobachter mit einer geringeren Expertise verbessert, die Schlussfolgerung über die Teilnehmer aber verschlechtert, weil sie von Kriterien geleitet wird, die außerhalb des Beobachtungsinventars liegen. Vielleicht ist dieses Vorgehen dadurch motiviert, dass die Teilnehmer generell niedrigere Leistungspotenziale hatten als beim Datensatz 1, so dass die AC-Assistenz und die Novizen für eine Rangbildung der Teilnehmer untereinander nicht zuletzt auch auf Grund ihrer möglicherweise geringeren Erfahrung in Personalauswahlsituationen auf ihnen mehr vertraute Kriterien zurückgriffen.

In Bezug auf den Datensatz 2 zeigen sich hinsichtlich der Übereinstimmung in der individuellen Rangfolge mit der Rangfolge des AC-Leiters für die anderen Beobachtergruppen keine signifikanten Unterschiede. Bei der Betrachtung der Übereinstimmung

der einzelnen Beobachtergruppen untereinander fällt auf, dass die Übereinstimmung der einzelnen Beobachtergruppen hinsichtlich der Gesamtrangfolge sowohl für sich allein genommen, wie auch im Vergleich untereinander gegenüber den Ergebnissen aus Datensatz 1 absinkt (vgl. Tab. 15). Gerade hier wäre zu erwarten, dass die Übereinstimmungen der Beobachter größer werden, weil weniger Störeinflüsse durch andere Teilnehmer, die nicht beobachtet werden aber zugegen sind (vgl. die Gruppenübungen), wegfallen und die Belastung der Beobachter insgesamt durch die geringere Anzahl an Teilnehmern geringer ist als bei den AC von Datensatz 1.

Bei dem dieser Untersuchung zu Grunde liegenden Datenmaterial konnte festgestellt werden, dass sich Experten und Novizen in ihrer Beobachterkompetenz nur unwesentlich voneinander unterscheiden. Es gibt übungsspezifische Schwerpunkte, und es gibt auch verhaltensmerkmalsbezogene Schwerpunkte (vgl. auch Schmitt, 1977); diese unterschiedlichen Schwerpunkte beeinflussen jedoch nicht die generelle Entscheidung, ob ein Kandidat in einem AC als "gut" oder "schlecht" beurteilt werden muss. So konnte die Analyse der Gesamtrangfolge (gemäß Forschungshypothese 3) zeigen, dass beim Datensatz 1 insgesamt die Übereinstimmung der Beobachtergruppen mit der Gesamtrangfolge recht hoch ist; am niedrigsten ist die Übereinstimmung der Experten mit der Gesamtrangfolge (.75).

Nur untereinander variieren die Blickwinkel der einzelnen Beobachtergruppen stärker (vgl. Abb. 8). Das bedeutet jedoch, dass durch das AC-System Verhalten bei den Teilnehmern provoziert werden konnte, dass Eindrücke hinterlassen hat, die die Beobachter befähigten, eine nach einem bestimmten Vorgehen ermittelte Gesamtrangfolge über die Teilnehmer mit einer hohen Übereinstimmung einzuschätzen.

Ein Ziel von Beobachtertrainings ist die Verbesserung der Qualität der Beurteilungen. Ein derartiger Qualitätsindikator ist die Genauigkeit, mit der ein bestimmtes Verhaltensmerkmal in verschiedenen Übungen beobachtet werden kann. Bei dem vorliegenden AC-Typ war die Philosophie jedoch, dass zwar dieselben Verhaltensmerkmale, jedoch mit übungsspezifischen Operationalisierungen versehenen beobachtet werden sollen (ausgenommen davon sind die Gruppenübungen, in denen identische Beobachterbögen benutzt wurden). In diesem Fall kann es nicht erwünscht sein, dass Verhaltensmerkmale, die in unterschiedlichen Übungen mit einem unterschiedlichen Verhaltensaufforderungscharakter und dort mit verschiedenen Operationalisierungen be-

obachtet werden, hohe mittlere Korrelationen aufweisen. Das Verhaltensmerkmal "Engagement" bspw. in einem Zweiergespräch wird anders operationalisiert sein müssen als in einer Gruppenübung. Von daher müssen die mittleren Korrelationen zwangsläufig niedrig sein. Das konnte die vorliegende Untersuchung bestätigen: die mittleren Korrelationen der Verhaltensmerkmale sind über die Übungen hinweg niedriger als in der Literatur berichtet.

Ein anderes Ziel von Beobachtertrainings besteht darin, die Beobachterübereinstimmung zu verbessern. In der vorliegenden Untersuchung konnte nachweisen, dass sich Experten und Novizen voneinander kaum unterscheiden, dass sie aber in einigen Übungen schlechter beobachten als die AC-Assistenz oder der AC-Leiter. Sie beobachten zudem mit unterschiedlichen Schwerpunkten. Die Novizen kommen wider Erwarten zu ähnlichen Gesamtergebnissen wie die anderen Beobachtergruppen und konnten zudem zeigen, dass sie während des AC Lerneffekte hatten, die sie befähigten, auf einem ähnlich hohen Qualitätsstandard zu beobachten, wie Experten, AC-Assistenz oder AC-Leiter. Bezüglich der Experten muß vermutet werden, dass diese unter bestimmten Bedingungen ihre Beurteilungen unabhängig von den in den Beobachterbögen nahegelegten Operationalisierungen treffen. Möglicherweise nutzen sie lediglich die in den Beobachterbögen vermerkten Überschriften über den Verhaltensoperationalisierungen, um ihre Eindrucksbildung in Form von Beurteilungen zu Papier zu bringen, und beachten weniger die zu beurteilenden Operationalisierungen.

Wenn als Beobachtertraining die Teilnahme an einem AC gilt, dann muss diskutiert werden, wie oft denn ein Beobachter an einem AC teilnehmen muß, um die Expertise der AC-Assistenz oder des AC-Leiters zu erlangen, der als Psychologe einen anderen Hintergrund und einen anderen Zugang zur "Materie AC " hat als die meisten Beobachter. Können andere Beobachter eine derartige Expertise überhaupt durch ein einziges Beobachtertraining erlangen? Das Erfahrungswissen hinsichtlich der Personalauswahl mittels AC scheint bedeutsam zu sein – zumindest gibt es Anzeichen dafür, dass bei dieser einfachen Art von AC das Erfahrungswissen einen Effekt haben könnte.

Die Ergebnisse dieser Arbeit müssen vor dem Hintergrund der Restriktionen, denen die Untersuchung ausgesetzt ist, verstanden werden. Vorgehensweisen, wie man sie in Laborsituationen gestalten könnte, waren hier nicht möglich. Die Untersuchung stützt

sich auf AC, die ein Versicherungsunternehmen durchgeführt hat, um geeignetere Bewerber für den Versicherungsaußendienst zu rekrutieren als in der Vergangenheit. Geeigneter bedeutet in diesem Fall eine geringere Fluktuation und eine größere Produktivität. Es verbergen sich also handfeste ökonomische Interessen hinter diesem Vorgehen. Seitens der Bewerber muss betont werden, dass diese entweder in einem überwiegend branchenfremden festen Arbeitsverhältnis waren oder aber eine Anstellung suchten. Mit dem AC sollte ihnen die Risikoentscheidung, sich auf eine Tätigkeit im Versicherungsaußendienst einzulassen, erleichtert werden. Sie versprachen sich von einer derartigen Tätigkeit eine deutliche und schnelle Verbesserung ihrer eigenen wirtschaftlichen Verhältnisse – möglicherweise über die Grenze hinaus, die ihnen ihr erlernter Beruf gesetzt hätte. Beiderseits waren die Erwartungen also hoch. Damit unterscheiden sich dieses AC von vielen AC, in denen studentische Kandidaten als AC-Teilnehmer unter kontrollierten Bedingungen ohne existenzielle Konsequenz teilnahmen. Das macht den Charme dieser Untersuchung aus: sie behandelt eine Personalauswahlsituation aus dem Leben.

Dadurch rücken die oben angesprochenen Restriktionen ein wenig in den Hintergrund. Natürlich wäre es interessanter gewesen, ein Beobachtungsinventar neu, vielleicht leichter verständlich zu konstruieren. Das lehnte das Unternehmen aus Gründen der Vergleichbarkeit mit den bisher stattgefundenen AC ab. Möglicherweise konnte auch mit den beschriebenen Anforderungen nicht immer die Komplexität des gewünschten Zielverhaltens abgebildet werden, was eventuell für den Zweck der Personalauswahl zu verschmerzen ist; für Personalentwicklungsziele müsste dieser Aspekt jedoch überdacht werden.

Der untersuchte Datensatz besteht aus teilweise abhängigen, teilweise unabhängigen Stichproben. Statistisch gesehen gibt es für diese Fragestellung keine gute Lösung. Es wäre sicherlich besser gewesen, nur Experten und nur Novizen miteinander zu vergleichen.

Auch die Zuordnung der Beobachter zu den einzelnen Kandidaten ist nicht kontrollierbar gewesen. Sie war mehr zufällig und unterscheidet sich daher von Untersuchungen, in denen seitens des Settings eine eindeutige Zuordnung von Beobachtern zu Kandidaten vorgenommen werden konnten. Das könnte dazu geführt haben, dass trotz der Rotation der Teilnehmer ein Teilnehmer in den Gruppenübungen von einem Beobach-

ter in beiden Übungen beobachtet wurde. Dadurch ist in der vorliegenden Untersuchung nicht immer in Vorfeld eindeutig geklärt, wie viele Beobachtungen letztendlich zum erhobenen Wert beigetragen haben. Dieser ermittelt sich aus den Daten verschiedener Beobachter.

Manchmal gab es in den AC auch keine Experten, bzw. Novizen oder unterschiedlich viele Experten und Novizen. und unterschiedlich viele Teilnehmer. Urteile über Teilnehmer wurden gefällt in dem Wissen, dass die Beobachter nicht alle Teilnehmer in allen Übungen gesehen haben (aus Gründen der Zeitökonomie wurden in der Übung "Zweiergespräch" und in der Übung "Stressinterview" die Teilnehmergruppen getrennt). Zudem könnten sich durch die Unterschiede in der Moderation dieser beiden Übungen (eine Gruppe wurde vom AC-Leiter betreut, die andere Gruppe von der AC-Assistenz) Versuchsleitereffekte niedergeschlagen haben, die die Beurteilungen der Beobachter beeinflusst haben könnten.

Ferner war es aus Gründen der Durchführungsökonomie nicht möglich, die Anzahl der zu beobachtenden Teilnehmer, Verhaltenanforderungen und deren Operationalisierungen in den Gruppenübungen zu reduzieren. Ungeübte Beobachter könnten sich hier schnell überfordert fühlen; ebenso könnte die gemeinsame Beurteilung beider Teilnehmer in den "Zweiergesprächen" auf einem Beobachterbogen Kontrasteffekte begünstigt haben.

Dennoch: viele Aspekte konnten realisiert werden, die auch in der Literatur als Qualitätsindikatoren und/oder als qualitätssteigernd dargestellt werden. Das Ziel der vorliegenden Untersuchung bestand darin herauszufinden, ob es möglich sei, getrieben von ökonomischen Gesichtspunkten, auf ein Beobachtertraining in einem Personalauswahl-AC zu verzichten. Es konnte mit dem Experten / Novizen-Vergleich gezeigt werden, dass dies bei den gegebenen Voraussetzungen, für diese spezifische Art von AC, möglich ist, weil die Beobachter in ihrer Beurteilung eine hohe Übereinstimmung mit dem Gesamtergebnis haben.

Letztendlich wird es nicht möglich sein aufzuklären, welches Ausmaß an Beobachterübereinstimmung mindestens erforderlich ist und höchstens erforderlich sein darf, um den Beurteilungsprozess im AC als qualitativ gut zu bezeichnen. Menschliches Verhalten wird individuell wahrgenommen – und individuelle Maßstäbe der Beobachter

beeinflussen seine Beurteilung im AC. Das AC ist ein situatives Verfahren, das zu Ergebnissen führt, die wegen der für Beobachter und Teilnehmer einsetzenden Lerneffekte nicht reproduzierbar sind. Nichts desto trotz kann die Prozeßsicherheit, um hier einen Begriff aus dem Qualitätswesen zu gebrauchen, optimiert werden, indem...

- die Eindeutigkeit des Beobachtungsinventars dadurch erhöht wird, dass die sprachliche Abfassung von Definitionen und Operationalisierungen die Normen und Werte des Unternehmens abbilden;
- eine Abbildung der branchenspezifischen Arbeitsrealität in den Übungen erfolgt;
- die Steuerung der Beobachter entsprechend dem weiter oben skizzierten Vorgehen (keine Diskussionen untereinander während des AC, Beurteilung des gesehenen Verhaltens unverzüglich mit Korrekturmöglichkeit) vorgenommen wird;
- der Prozess der Urteilsbildung anonymisiert und individualisiert, mechanisch und emotional gestaltet wird;
- eine ausführliche Vorinformation der Teilnehmer über das AC erfolgt;
- die Steuerung des Teilnehmer durch die Teilnehmerinstruktion ermöglicht, das gewünschte Zielverhalten zu zeigen.

Berufliche Erfahrungen kommen zum Tragen und können im Falle der Personalauswahl, wenn es darauf ankommt, einen globalen Eindruck über Teilnehmer zu erlangen, hilfreiche Ergänzung sein.

Nun kann es sein, dass das vorliegende Beobachtungsinventar relativ restriktiv ist, so dass Beobachter wie die Novizen spontan zu guten Ergebnisse kommen. Ob diese Ergebnisse auch bei komplexeren Beobachtungsinventaren, wie sie z.B. in Personalentwicklungs-AC benutzt werden, nachgewiesen werden können, ist zu prüfen. Wie die Ergebnisse in Datensatz 2 gezeigt haben, kann der Blickwinkel der Beobachter sehr unterschiedlich sein. Mehr Gemeinsamkeit wäre wünschenswert. Hier würde es sich anbieten, vor allem für die, die sich in der Personalauswahl/Mitarbeiterbeurteilung und/oder im Umgang mit dem AC-System sehr sicher fühlen (Experten) eine Sensibilisierung vorzunehmen mit dem Ziel, mehr gemäß den Vorgaben der Beobachterbögen zu beurteilen und schnelle Schlußfolgerungen zu vermeiden.

Durch eine veränderte Gestaltung der Beobachterbögen (keine Oberbegriffe) und des Beobachtermaterials (keine Oberbegriffe und Definitionen aufführen) könnten die Be-

obachter noch mehr gezwungen werden, sich auf die einzelnen Verhaltensoperationalisierungen zu konzentrieren, und würden dadurch genauer die Beobachterbögen für den Beurteilungsprozeß nutzen. Andererseits muss bei dieser Einschätzung berücksichtigt werden, dass sie auf Ergebnissen beruht, die dadurch zu Stande kommen, dass die unterschiedlichen Beobachtergruppen in ihrer Übereinstimmungsqualität immer mit dem AC-Leiter in Verbindung gebracht werden. Der AC-Leiter wird als "Oberexperte" angesehen wurde; seine Expertise wurde jedoch nicht überprüft.

Die Forschung müsste die Frage nach dem Grad der Beobachterübereinstimmung als Indikator für einen Beobachterfehler neu diskutieren. Wird in unterschiedlichen Typen von Übungen ein und dasselbe Merkmal mit identischen Operationalisierungen beobachtet, dann liegt es auf der Hand, dass die Beobachter dazu verleitet werden, zu raten, weil das übungsabhängige Verhalten der Teilnehmer nicht zu dem zu beobachtenden Verhalten passt.

Wünschenswert wäre auch, eine Hilfestellung dafür zu bekommen, das Leistungsverhalten einer Teilnehmer-, bzw. Beobachtergruppe beschreiben zu können, um Vergleichbarkeit mit unterschiedlich leistungsstarken Gruppen herzustellen zu können. Gerade für Beobachter bei Personalentwicklungs-AC könnten z.B. Pre-Tests im Vorfeld zu individuellen Qualifizierungsmaßnahmen im Rahmen von Online-Programmen führen, was ökonomischer ist als ein aufwändiges Beobachtertraining.

## Zusammenfassung

Für ein Assessment Center zur Auswahl von Versicherungs-Außendienstmitarbeitern wurde geprüft, ob Beobachterexperten und Beobachternovizen zu ähnlichen oder zu unterschiedlichen Schlussfolgerungen über die Teilnehmer kommen. Eine ähnliche Schlussfolgerung sollte dabei Indikator dafür sein, dass ein Beobachtertraining (hier: Verbesserung der Beobachterexpertise durch wenigstens einmalige Teilnahme an einem AC als Beobachter) ohne Auswirkungen bleibt. Verglichen wurden die Ergebnisse dieser Beobachtergruppen mit dem AC-Leiter als "Oberexperten" und der AC-Assistenz. Im Rahmen der vorliegenden Untersuchung wurden die Daten von 37 AC mit insgesamt 267 Kandidaten und 171 Beobachtern in zwei unterschiedlichen Datensätzen überprüft. Es konnte nachgewiesen werden, dass sowohl Experten wie auch Novizen eine hohe Übereinstimmung mit der Gesamtrangfolge über die Teilnehmer hatten; hinsichtlich der übungsbezogenen und der merkmalsbezogenen mittleren Korrelationen sowie der mittleren Korrelationen innerhalb der einzelnen Übungen wurden kaum signifikante Unterschiede zwischen den einzelnen Beobachtergruppen festgestellt. Aus diesem Grunde kann geschlossen werden, dass die Teilnahme an einem AC als Beobachter nicht notwendigerweise die Beobachterexpertise verbessert. Angesichts der hohen Übereinstimmung der Beobachtergruppen mit der Gesamtrangfolge über die Teilnehmer kann geschlossen werden, dass für den vorliegenden einfachen AC-Typ verbunden mit der spezifischen Art der Ergebniserhebung der Verzicht auf ein Beobachtertraining nicht notwendigerweise eine Verschlechterung der AC-Ergebnisse bedeutet.

## Literatur

- Arbeitskreis Assessment Center e.V. (1996) Standards der Assessment Center-Technik. in: Arbeitskreis Assessment Center e.V. (Hrsg.), Assessment Center als Instrument der Personalentwicklung, (S. 485 -493), Hamburg: Windmühle.
- Asch, S.E. (1946). Forming impressions of personality. *Journal of Abnormal Social Psychology*, 41, 258- 290.
- Bender, E. & Hastorf, A.H. (1953). On measuring generalized empathic ability (social sensitivity). *Journal of Abnormal Social Psychology*, 48, 503-506.
- Benedetti, D.T. & Hill, J.G. (1960). A determiner of the centrality of a trait in impression formation. *Journal of Abnormal Social Psychology*, 60, 278-280.
- Bernardin, H.J. & Pence, E.C. (1980). Effects on rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, 65, 60-66.
- Bernardine, H.J. & Buckley, M.R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205-212.
- Beutel, K., Fischer, H.P., Heinecke, H.J., Moeckel, K., Uhlenbrock, H.J., Vornberger, E. (1983). Validierung des betrieblichen Diagnoseinstruments Assessment Center. 12. Kongreß für Angewandte Psychologie, bdp, Band 3, 3- 21.
- Bittner, R.H. & Rundquist, E.A. (1950). The rank-comparison rating method. *Journal of Applied Psychology*, 34, 171 –177.
- Bobbitt, R.A., Gordon, B.N., & Jensen, G.D. (1966). Development and application of an observational method: Continuing reliability testing. *Journal of Psychology*, 63, 83-88.
- Bormann, W. (1975). Effects of instruction to avoid halo error on reliability and validity of performance evaluating ratings. *Journal of Applied Psychology*, Vol. 60, No.5, 556-560.
- Borman, W. (1977). Consistency of rating accuracy and rating error in the judgement of human performance. *Organizational Behavior and Human Performance*, 20, 238-252.
- Bormann, W. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, Vol. 64, No. 4, 410- 421.
- Bortz, J. (1979). *Lehrbuch der Statistik. Für Sozialwissenschaftler*. Berlin: Springer-Verlag.
- Bortz, J. & Döring, N. (1995); *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin: Springer-Verlag.
- Brandstätter, H. (1979). Die Beurteilung von Mitarbeitern. in: Mayer, A. & Herwig, B.(Hrsg.): *Handbuch der Psychologie*, Bd 9, Betriebspsychologie, 668 - 734.
- Bray, D.W. & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, 80, Whole 625, 1-27.

- Buzby, D.E. (1924). The interpretation of facial expression. *American Journal of Psychology*, 35, 602-604.
- Byham, W.C. (1977). Assessor selecting training. in: Moses, J.L. & Byham, W.C. (Eds.), *Applying the Assessment Center Method*. New York, 89-126.
- Campbell, D.T. & Fiske, D.W. (1958). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, Vol.56, No.2, 81-105.
- Cattell, R.B. (1957). *Personality and motivation-Structures and measurement*. Yonkers on Hudson.
- Comelli, G. (1973). Warum wir Mitarbeiter falsch beurteilen. *Plus, Zeitschrift für Führungspraxis*, 1, 1-11.
- Cooper, H.W. (1981). Conceptual similarity as a source of illusory halo in job performance ratings. *Journal of Applied Psychology*, Vol. 66, No.3, 302-307.
- De Soto, C.B. (1961). The predilection for single orderings. *Journal of Abnormal and Social Psychology*, Vol. 62, No.1, 16-23.
- Dorsch, H. & Jochum, J. (1982). Beobachtungen im Assessment Center. *Zeitschrift für betriebswirtschaftliche Forschung*, 374-383.
- Donat, M. & Moser, K. (1989). Die Arbeits- und Anforderungsanalyse als Grundlage der Gestaltung von Assessment Centern. in: Lattmann, Ch. (Hrsg.), *Das Assessment Center-Verfahren der Eignungsbeurteilung*, Heidelberg: Physica-Verlag, 155-182.
- Exler, C. & Kleinmann, M. (1997). Die Wirkung impliziter Kriterien bei einem Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 41, Jg. (N.F.15) 2, 86-90.
- Feldman, J.M. (1981). Performance Appraisal. *Journal of Applied Psychology*, 66, 1-148.
- Feltham, R. (1988). Assessment centre decision making: Judgemental vs. mechanical. *Journal of Occupational Psychology*, 61, 237-241.
- Fischer, M. (2000). *Von der Arbeitserfahrung zum Arbeitsprozesswissen. Rechnergestützte Facharbeit im Kontext beruflichen Lernens*. Opladen: Leske+Budde.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Fox, S., Bissman, A. & Herrman, E. (1983). The halo-effect. Is it a unitary concept? *Journal of Occupational Psychology*, 61, 129-144.
- Franke, J. (1959). Die wechselseitige Abhängigkeit von Persönlichkeitsbeurteilungen. *Psychologie und Praxis*, 3, 178-83.
- Franke, J., Hoffmann, K. & Kamphaus, G. (1974). Zum Training gegen Überstrahlungseffekte bei der Beurteilung von Personen. *Psychologische Beiträge*, Band 16, 188-202.
- Gage, N.L. (1952). Judging interests from expressive behavior. *Psychological Monographs*, 66, No.18.

- Gaugler, B.B., & Thornton 111, G.C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, Vol. 74, No.4, 611-618.
- Gerpott, T.J. (1985). Training von Beurteilern zur Verbesserung von Leistungsbeurteilungsprozessen in Organisationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 29 (N. F. 3), 116-128.
- Ghiselli, E.E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Goodstone, M.S. & Lopez, F.E. (2001). The Frame of Reference Approach as a Solution to an Assessment Center Dilemma. *Consulting Psychology Journal: Practice and Research*, Vol. 53, No. 2, 96-107.
- Greenwood, J.M. & McNamara, W.J. (1967). Interrater reliability in situational tests. *Journal of Applied Psychology*, Vol. 51 No.2, 101-106.
- Guldin, A. (1996). *Konstruktvalidität des AC-Verfahrens: Empirische Untersuchung zum modifizierten Konzept der diskriminanten und konvergenten Validität*. Frankfurt/M.: Peter Lang.
- Halpern, D.F. (1997). Sex differences in intelligence. Implications for education. *American Psychologist*, 52, 1091-1102.
- Hartmann, D.P. (1982). Assessing the dependability of observational data. in: Hartmann, D.P. (Ed.). *Using observers to study behavior*. *New Directions for Methodology of Social and Behavioral Science*, No 14, San Francisco: Jossey-Bass.
- Hasemann, K. (1983). Verhaltensbeobachtung und Ratingverfahren. in: Groffmann, K. & Lothar, M. (Hrsg.), *Verhaltensdiagnostik*, Göttingen: Hogrefe, 434-488.
- Heitmeyer, K. & Thom, N. (1988). *Assessment-Center Gestaltungs- und Anwendungsmöglichkeiten*. 3. Aufl., Köln
- Hennessey, J., Maybe, B. & Warr, P. (1998). Assessment Centre Observation Procedures: An Experimental Comparison of Traditional, Checklist and Coding Methods. *International Journal Of Selection And Assessment*, Volume 6, Number 4, 222-231.
- Hinrichs, J.R. & Haanperä, S. (1976). Reliability of measurement in situational exercises: An assessment of the assessment centre method. *Personnel Psychology*, 29, 31-40.
- Höft, S. & Funke, U. (2001). Simulationsorientierte Verfahren der Personalauswahl. in: Schuler, H. (Hrsg.): *Lehrbuch der Personalpsychologie* (S. 135-173). Göttingen: Hogrefe.
- Hofsommer, W. (1980). Personalbeurteilung als Kommunikationsproblem. in: Neubauer, R. & Rosenstiel, L. v. (Hrsg.): *Handbuch der Angewandten Psychologie*, Bd. 1: Arbeit und Organisation, München: Verlag Moderne Industrie.
- Huck, J.R. (1973). Assessment centers: A review of the external and internal validities. *Personnel Psychology* 26, 191-212.
- Hunter, J.E. & Hunter, E.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 1984, 96, 72-98.
- Hoyos, C. Graf (1974). *Arbeitspsychologie*. Stuttgart: Kohlhammer

- International Task Force on Assessment Center Guidelines (2000). Guidelines and Ethical Considerations for Assessment Center Operations. 28<sup>th</sup> International Congress on Assessment Center Method, San Francisco.
- Ivancevich, J.M. (1979). A longitudinal study of the effects of rater training on psychometric error in rating. *Journal of Applied Psychology*, 6,4, 502-508.
- Jäger, A.O. (1970). Personalauslese. in: Mayer, A. & Herwig, B. (Hrsg.), *Betriebspsychologie*, (S. 613-667). Göttingen: Hogrefe.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern: Assessment Center Verfahren*. München: Hanser.
- Jeserich, W. & Grube, S. (1989). Stand der Assessment Center Methode in der Bundesrepublik Deutschland - Auswertung einer Umfrage 1988. in: Arbeitskreis Assessment Center (Hrsg.), *Das Assessment Center in der betrieblichen Praxis*. Hamburg: Windmühle GmbH.
- Jones, E.E. (1954). Authoritarianism as a determinant of first impression formation. *Journal of Personality*, 23, 107-127.
- Kagan, J. (1967). On the need for relativism. *American Psychologist*, 22, 131-142.
- Kazdin, A.E. (1982). *Single case research design: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kelly, E.L. & Fiske, D.W. (1951). *The prediction of performance in Clinical Psychology*. Ann Arbor.
- Klimoski, R.J. & Strickland, W.J. (1977). Assessment centers valid or merely precient. *Personnel Psychology*, 30, 353-361.
- Kleinmann, M., Exler, C., Kuptsch, C. & Köller, O. (1995). Unabhängigkeit und Beobachtbarkeit von Anforderungsdimensionen im Assessment Center als Moderatoren der Konstruktvalidität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, Jg. (N.F.13) 1, 22-28.
- Kleinmann, M. (1997). *Assessment Center: Stand der Forschung – Konsequenzen für die Praxis*. Göttingen: Hogrefe.
- Kolk, N.J. (2001). *Assessment Centers. Understanding and improving construct-related validity*. Enschede: Printpartners Ipskamp BV.
- Koltuv, B.B. (1962). Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs*, 76, No.33, Whole No.552, 1-33.
- Kompa, A. (1989) *Assessment Center: Bestandsaufnahme und Kritik*. München u. Mering: Hampp.
- Kremers, J. (1962). The influence of studying psychology on the process of interpersonal perception. in: *Berichte des 16. Internationalen Kongresses für Psychologie*, Amsterdam, 583-585.
- Lammers, F. (1992). Zur Problematik des Beobachterverhaltens im Assessment Center. *Forschungsberichte aus dem Fachbereich Psychologie der Universität Osnabrück*, Nr. 82.
- Lammers, F. & Holling, H. (2000). Der Einfluss der Beobachterrotation auf die Konstruktvalidität des Assessment Centers. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 21, 270-278.
- Landy, F.J. & Farr, J.L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-107.

- Landy, F.J. (1985). *Psychology of work behavior*. (3rd ed.). Chicago: Dorsey Press.
- Latham, G.P., Wexley, K.N. & Pursell, E.D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, Vol. 60 No.5, 550-555.
- Lewry, P.E. (1993). The Assessment Center: An Examination of the Effects of Assessor Characteristics on Assessor Scores. *Public Personnel Management*, Vol. 22 No. 3, 487-499.
- Levine, J. & Butler, J. (1952). Lecture versus group discussion in changing behaviors. *Journal of Applied Psychology*, 36, 29-33.
- Lievens, F. (1998). Factors which Improve the Construct Validity of Assessment Centers: A Review. *International Journal Of Selection And Assessment*, Volume 6, Number 3, 141-152.
- Lievens, F. (2001). Assessor Training Strategies and Their Effects on Accuracy, Inter-rater Reliability, and Discriminant Validity. *Journal of Applied Psychology*, Vol. 86, No. 2, 255-264.
- Lilly, W. & Frey, D. (1993). Die Hypothesentheorie der Sozialen Wahrnehmung. in: D. Frey & M. Irle (Hrsg.), *Kognitive Theorien der Sozialpsychologie* (S. 49-78). Bern: Huber.
- Luft, J. (1950). Implicit hypotheses and chemical predictions. *Journal of Abnormal and Social Psychology*, 45, 756- 760.
- Lorenzo, R.V. (1984). Effects of assessorship on managers proficiency in acquiring, evaluating, and communicating information about people. *Personnel Psychology*, 37, 617-643.
- Mash, E.J. & McElwee, D. (1974). Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. *Child Development*, 45, 367-377.
- Maukisch, K. (1989). Informationswert und Ökonomie der diagnostischen Prinzipien von Assessment Center Systemen zur Erfassung von Management Potential. in: Arbeitskreis Assessment Center (Hrsg.), *Das Assessment Center in der betrieblichen Praxis*. Harnburg: Windmühle GmbH.
- McClelland, W. (1951). A preliminary test of role-playing ability. *Journal of Consulting Psychology*, 15, 102-108.
- McLaughlin, B. (1979). Similarity, recall, and reappraisal of others. *Journal of Personality*, 38, 106-110.
- Meehl, P.E. (1954). *Clinical vs. statistical prediction*. Minneapolis.
- Michel (1964). Allgemeine Grundlagen psychometrischer Tests. in: Heiß, R. (Hrsg.). *Psychologische Diagnostik, Handbuch der Psychologie*, Bd. 6 (S.19-70), Göttingen: Hogrefe.
- Moskowitz, D.S. & Schwarz, J.C. (1982). Validity comparison of behavior counts and ratings by knowledgeable informants. *Journal of Personality and Social Psychology*, Vol. 42, No.3, 518-528.
- Murphy, K.R., Martin, C. & Garcia, M. (1982). Do behavioral observations measure observation? *Journal of Applied Psychology*, Vol. 67, No.5, 562-567.
- Murray, H.M. (1938). *Explorations in Personality*. New York.

- Neubauer, O. (1989). Assessment Center - Ein Handel mit Illusionen? in: Arbeitskreis Assessment Center (Hrsg.), Das Assessment Center in der betrieblichen Praxis. Harnburg: Windmühle GmbH.
- Neubauer, R. (1990). Frauen im Assessment Center - ein Gewinn?. Zeitschrift für Arbeits- und Organisationspsychologie, 34, 29-36.
- Newcomb, T. (1931). An experiment designed to test the validity of the rating technique. Journal of Educational Psychology, 22, 279- 289.
- Pawlik, K. (1976). Modell- und Praxisdimensionen psychologischer Diagnostik. in: Pawlik, K. (Hrsg), Diagnose der Diagnostik (S. 13-43). Stuttgart: Klett .
- Pulakos, E.D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 58 -5881.
- Randhofer, T. (1995). The social acceptance of the assessment center technique: an empirical analysis. *Risorsa Uomo*, Vol. 3, n. 1, 25-37.
- Reid, J.B. (1982). Observer training in naturalistic research. in: Hartmann, D.P. (Ed.): Using observers to study behavior. *New Directions for Methodology of Social and Behavioral Science*, No 14, San Francisco: Jossey-Bass.
- Reilly, R.R., Henry, S., & Smither, J.W. (1990). An examination of the effects of using behavior checklists on the construct validity dimensions. *Personnel Psychology*, 43, 71-84.
- Rojahn, J. (1984). Inter- und Intra-Beobachterübereinstimmung als Gütekriterien der systematischen Verhaltensbeobachtung. in: Bericht über den 34. Kongreß der Deutschen Gesellschaft für Psychologie in Wien, Band 2: Anwendungsbezogene Forschung, Hogrefe, 575-576.
- Saal, F.E., Downey, R.G. & Lahey, M.A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88, 413-428.
- Sackett, P.R. & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 400-410.
- Sackett, P.R., & Wilson, M.A. (1982). Factors affecting the consensus judgement process in managerial assessment center. *Journal of Applied Psychology*, Vol. 67 No.1, 10-17.
- Sackett, P.R. (1987). Assessment centers and content validity: some neglected issues. *Personnel Psychology*, 40, 13-25.
- Sagie, A. & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103-108.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 66, 178-200.
- Schleicher, D.J., Day, D.V. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behaviour and Human Decision Process*, 73, 76-101.
- Schleicher, D.J., Day, D.V., Bronston, T.M. & Riggio, R.E. (2002). A New Frame for Frame-of-Reference Training: Enhancing the Construct Validity of Assessment Centers. *Journal of Applied Psychology*, Vol. 87, No. 4, 735-746.

- Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgements. *Journal of Applied Psychology*, Vol. 62, No.3, 171-176.
- Schuler, H. & Stehle, W.(1983). Neuere Entwicklungen des Assessment Center-Ansatzes - beurteilt unter dem Aspekt der Sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- Schuler, H. (1989). Die Validität des Assessment Centers. in: Arbeitskreis Assessment Center (Hrsg.), *Das Assessment Center in der betrieblichen Praxis*. Hamburg: Windmühle GmbH.
- Schuler, H. & Funke, U. (1989). Berufseignungsdiagnostik. in: E. Roth, (Hrsg.): *Enzyklopädie der Psychologie*, Themenbereich D, Serie III, Bd. 3,281-320, Göttingen: Hogrefe.
- Schuler, H. & Prochaska, M. (1999). Ermittlung personaler Merkmale: Leistung und Potenzialbeurteilung bei Mitarbeitern. in: Kh. Sonntag (Hrsg.), *Personalentwicklung in Organisationen*, 2. Auflage, S. 181-210, Göttingen: Hogrefe.
- Schuler, H. & Höft, S. (2001). The conceptual basis of assessment center ratings. *International Journal of Selection and Assessment*, 9, 114-123.
- Stevens, N. & Wonderlic, E.F. (1934). An effective revision of the rating technique. *Personnel Journal*, 13, 125-134.
- Stamoulis, D.T. & Hauenstein, N.M.A. (1993). Rater training and rating accuracy: Training for dimensional accuracy vs. Training for ratee differentiation. *Journal of Applied Psychology*, 78, 994-1003.
- Spychalski, A.C., Quinones, M.A., Gaugler, B.B. & Pohley, K. (1997). A Survey of assessment center practice in organizations in the United States. *Personnel Psychology*, 50, 71-90.
- Straka, G.A. (2000). Lernen unter informellen Bedingungen (informelles Lernen). in: QUEM Arbeitsgemeinschaft Qualifikationsentwicklungen Management (Hrsg.): *Kompetenzentwicklung 2000, Lernen im Wandel, Wandel durch Lernen*. Berlin: Waxmann.
- Taft, R. (1959). Multiple methods of personality assessment. *Psychological Bulletin*, 56, 333-356.
- Taylor, E.K. & Hastman, R. (1956). Relation of format and administration to the characteristics of graphic rating scales. *Personnel Psychology*, 9, 181-206.
- Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Thornton, G.C., Gaugler, B., Rosenthal, D.B. & Bentson, C. (1987). Die prädiktive Validität des Assessment Centers - eine Meta-Analyse. in: Schuler, H. & Stehle, W. (Hrsg.), *Assessment Center als Methode der Personalentwicklung*. Stuttgart: Verlag für Angewandte Psychologie, 36-60.
- Turnage, J.J. & Muchinsky, P.M. (1984). A comparison of predictive validity of assessment center evaluations versus traditional measures in forecasting supervisory job performance: Interpretative implications of criterion distortion for the assessment paradigm. *Journal of Applied Psychology*, Vol. 64, No.4, 595-602.

- Uhlenbrock, H.-G. & Vornberger, E. (1990). Beobachterurteile. in: Sarges, W. (Hrsg.), Management Diagnostik, (S.554-563). Göttingen: Hogrefe.
- Volpert, W. (1999). Wie wir handeln, was wir können. Ein Disput als Einführung in die Handlungspsychologie, 2. Auflage, Sottrum: Artefact-Verlag.
- Wells, W. D. & Smith, G. (1960) Four semantic rating scales compared. Journal of Applied Psychology, Vol. 44, No. 6, 393-397.
- Wexley, K., Yuki, G. Kovacs, S. & Sanders, R. (1972). Importance of contrast effects in employment interviews. Journal of Applied Psychology, 56,45 -48 .
- Willingham W.W. & Jones, M.B. (1958). On the identification of halo through analysis of variance. Educational and Psychological Measurement, Vol. XVIII No.2, 403- 407.
- Woehr, D.J. & Huffcutt, A.I. (1994). Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organisational Psychology, 67, 189-205.

**Anhang 1**

**Greenwood, & McNamara (1967,**

TABLE 1  
RELIABILITY COEFFICIENTS OF OBSERVER RATINGS AND RANKINGS

Division	Task Force		Leaderless		Manufacturing	
	Ratings Mean $r$ (Z) <sup>a</sup>	Rankings Mean $r$ (Z) <sup>b</sup>	Ratings Mean $r$ (Z)	Rankings Mean $r$ (Z)	Ratings Mean $r$ (Z)	Rankings Mean $r$ (Z)
A (N = 144)						
I (N = 48)	.75	.81	.70	.74	.88	.84
II (N = 48)	.62	.70	.57	.62	.77	.77
III (N = 48)	.86	.82	.82	.83	.88	.89
B (N = 48)	.74	.75	.58	.60	.77	.83
C (N = 60)	.65	.65	.66	.68	.74	.81
D (N = 36)	.84	.78	.48	.48	.49	.49

<sup>a</sup> Pearson product-moment correlations were used to determine the degree of rater agreement for an experimental group (in each experimental group three observers independently evaluated six participants who were completing a particular situational test). In order to determine an estimate of rater agreement for an entire division program, the interrater reliabilities obtained from experimental groups in a particular division were then averaged (after converting to Fisher Z transformations).  
<sup>b</sup> The same statistical procedure was used to determine reliability of ranking except that in each group the observers ranked the participants and Kendall's coefficient of concordance was used to obtain the degree of ranking agreement for an experimental group.

Division = die in der Untersuchung der Autoren analysierte Unternehmenseinheit

Guldin , A. (1996, S. 327)

Tabelle 14.5            Interrater-Übereinstimmung, AC-Verfahren I  
(AC-Teilnehmer N=89, eingeschätzt von 26 Beobachtern)

ÜBUNG:	GDUN <sup>76</sup>	MA	PRO	VK	VT	Mittleres r (pro Skala)
Skala 1	.668**	.742**	.818**	.706**	.650**	.723
Skala 2	.803**	.509**	.674**	.754**	.726**	.705
Skala 3	.599**	.729**	.664**	.655**	.733**	.679
Skala 4	.624**	.667**	.722**	.640**	.631**	.658
Skala 5	.515**	.576**	.708**	.604**	.758**	.641
Skala 6	.714**	.618**	.761**	.582**	.714**	.683
Mittleres r (pro Übung)	.664	.648	.729	.661	.705	

Skala= Verhaltensmerkmal

**Anhang 2**

## Übung 1

### "Selbstpräsentation" (Beobachtermaterial)

**Aufgabe:** Die Teilnehmer präsentieren sich vor den Kollegen und zeigen, dass sie gerne im Mittelpunkt der Aufmerksamkeit stehen.

**Dauer:** 10 Minuten Einzelarbeit.  
Jeweils 5 Minuten Präsentation.

**Kriterien:** **Umgang mit belastenden Situationen**  
Den geistigen, emotionalen und körperlichen Anforderungen des Verfahrens standhalten.

**Engagement**

Sich ehrgeizig und kontinuierlich am Geschehen beteiligen - auch, wenn es schwer fällt, sich mit fremden Ideen zu identifizieren.

**Persönliches Auftreten**

Selbstsicher und selbstbewusst anderen Menschen gegenüber-treten.

**Verbalisierung**

Sich sprachlich auch bei schwierigen Sachverhalten klar und deutlich ausdrücken.

**Überzeugungsfähigkeit**

So geschickt argumentieren, dass andere sich in ihrer Meinung anschliessen.

## Übung 1

### "Selbstpräsentation" (Teilnehmermaterial)

**Instruktion:** In dieser Übung geht es darum, dass Sie Ihre Zuhörer von sich überzeugen und für sich gewinnen sollen. Bitte präsentieren Sie sich dazu so, wie Sie einen, dass es für Sie am sinnvollsten und am vorteilhaftesten ist. Sie können dabei auf die folgenden Punkte eingehen, müssen es aber nicht.  
Für Ihre Präsentation dürfen Sie 5 Minuten nutzen.

Ich stelle Ihnen mit grosser Freude Herrn / Frau \_\_\_\_\_ vor.

1) Ich lege grossen Wert darauf, dass

---

---

---

---

2) Ich habe überhaupt kein Verständnis dafür, dass

---

---

---

---

3) Wenn ich wählen könnte,

---

---

---

---

4) Meine Zukunft stelle ich mir so vor,

---

---

---

---

5) Was ich sonst noch sagen möchte:

---

---

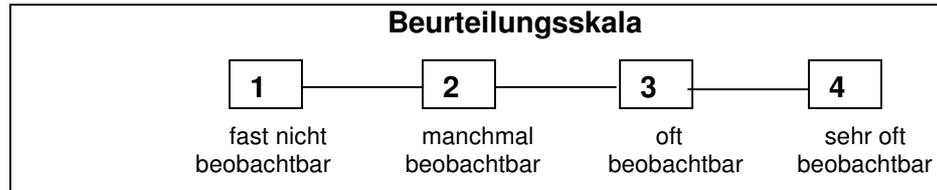
---

---

## Übung 1

### "Selbstpräsentation" (Beobachterbogen)

Beobachter: \_\_\_\_\_ Teilnehmer: \_\_\_\_\_



#### **Umgang mit belastenden Situationen**

- Der/die T. wirkt ruhig und gelassen. \_\_\_\_\_
- Der/die T. lässt keine körperlichen Stressmerkmale (z.B. zitternde Hände) erkennen. \_\_\_\_\_

#### **Engagement**

- Der/die T. geht spontan an die Aufgabe heran. \_\_\_\_\_
- Der/die T. kämpft für die eigenen Einschätzungen. \_\_\_\_\_

#### **Persönliches Auftreten**

- Der/die T. wirkt souverän. \_\_\_\_\_
- Der/die T. hat eine sympathische Ausstrahlung. \_\_\_\_\_

#### **Verbalisierung**

- Der/die T. bindet die Aufmerksamkeit der anderen durch überzeugenden sprachlichen Ausdruck. \_\_\_\_\_
- Der/die T. strukturiert die eigenen Aussagen. \_\_\_\_\_

#### **Überzeugungsfähigkeit**

- Der/die T. argumentiert auf der Sachebene. \_\_\_\_\_
- Der/die T. vertritt hartnäckig den eigenen Standpunkt. \_\_\_\_\_

---

**Gesamtdurchschnitt:** \_\_\_\_\_

---

## Übung 2

### "Gruppenübung 1" (Beobachtermaterial)

- Aufgabe:** Die TeilnehmerInnen diskutieren in der Gruppe über Anforderungen, die an Versicherungs - AussendienstmitarbeiterInnen gestellt werden.
- Typus:** Gruppenübung mit Konkurrenzcharakter.
- Dauer:** 10 Minuten Einzelarbeit  
45 Minuten Gruppenarbeit
- Kriterien:**
- Gestaltung zwischenmenschlicher Beziehungen**  
Zwischenmenschliche Beziehungen anbahnen, entwickeln und gestalten.
  
  - Selbständiges Verhalten**  
Meinungen vertreten, mit denen die anderen nicht notwendigerweise übereinstimmen müssen, jedoch auch Bereitschaft zeigen, diese Meinungen im Bedarfsfall zu revidieren.
  
  - Feedback übermitteln**  
Eigene Einschätzungen, Empfindungen, Bewertungen und Positionen übermitteln.
  
  - Teambildung unterstützen**  
Gruppenmitglieder in den Gruppenprozess integrieren und ihnen ihre Verantwortung für den Gruppenerfolg deutlich machen.
  
  - Problemlöseverhalten**  
Probleme Prozeduren unterziehen, die deren Lösung gewährleisten
  
  - Aufgaben systematisch erledigen**  
Komplexe Sachverhalte systematisch bearbeiten und erledigen.
  
  - Zielorientierung (nur Vertriebsleiter)**  
Gruppen- und Einzelaktivitäten hinsichtlich ihrer Bedeutung für das zu erreichende Ziel beeinflussen.

## Übung 2

### "Gruppenübung 1" (Teilnehmermaterial)

#### **Anforderungen an Versicherungs - AussendienstmitarbeiterInnen**

Bitte stellen Sie sich einmal die folgende Situation vor:

Sie sind Mitglied einer Projektgruppe Ihres Unternehmens, die einen Katalog über Anforderungen an Versicherungs - AussendienstmitarbeiterInnen, speziell für Versicherungsverkäufer, erstellen soll. Mit diesem Katalog sollen Richtlinien für die zukünftige Personalauswahl erstellt werden.

Ihr Ziel wird es sein, in der anschliessenden Besprechung mit der Projektgruppe, sprich, Ihren KollegInnen, eine Rangreihe über die vor Ihnen liegenden Anforderungen zu erstellen.

**Instruktion:** Bitte gehen Sie bei der Bearbeitung der Problematik folgendermassen vor:

1. Lesen Sie sich bitte die folgenden Anforderungen durch und erstellen für sich persönlich eine Rangfolge. Das Ihnen am wichtigsten erscheinende Kriterium erhält dabei den Rangplatz 1, das am wenigsten wichtige Kriterium erhält den Rangplatz 13. Tragen Sie die Rangplatzzahlen hinter den Anforderungen auf dem vorliegenden Formular ein.
2. Diskutieren Sie danach mit der Gruppe über die Anforderungen und erstellen eine Rangfolge, die von allen Mitgliedern getragen wird.

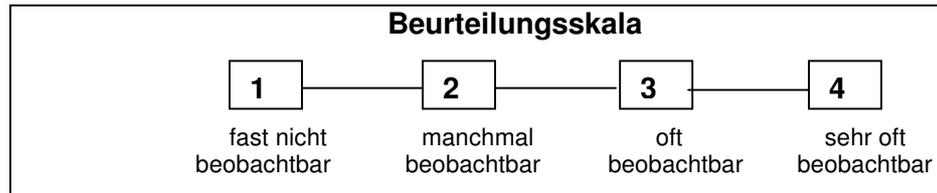
Während der Besprechung sollten Sie dafür eintreten, dass die Rangfolge der Projektgruppe möglichst genau mit der Rangfolge, die Sie selbst erstellt haben übereinstimmt!

	Mein Rangplatz
Geistige Flexibilität	
Problemlöseverhalten	
Schrittweises u. kontrolliertes Arbeiten	
Verbalisierung	
Gestaltung zwischenmenschliche Beziehungen	
Feedback übermitteln	
Teambildung unterstützen	
Selbständiges Verhalten	
Überzeugungsfähigkeit	
Zielstrebiges Verhalten	
Umgang mit belast. Situationen	
Engagement	
Persönliches Auftreten	

## Übung 2

### "Gruppenübung 1" (Beobachterbogen)

Beobachter: \_\_\_\_\_ Teilnehmer: \_\_\_\_\_



#### Gestaltung zwischenmenschlicher Beziehungen

- Der/die T. geht Konflikte produktiv ein. \_\_\_\_\_
- Der/die T. ist darum bemüht, dass andere die eigenen Positionen und Einschätzungen verstehen. \_\_\_\_\_

#### Selbständiges Verhalten

- Der/die T. vertraut der eigenen Urteilsfähigkeit. \_\_\_\_\_
- Der/die T. kann die eigene Meinung differenziert darstellen. \_\_\_\_\_

#### Feedback übermitteln

- Der/die T. äussert Kritik, ohne andere zu verletzen \_\_\_\_\_
- Der/die T. reflektiert Gruppenprozesse auf der Basis eigener Empfindungen. \_\_\_\_\_

#### Teambildung unterstützen

- Der/die T. bezieht Gruppenmitglieder in die Diskussion mit ein. \_\_\_\_\_
- Der/die T. sorgt dafür, dass sich alle Gruppenmitglieder mit dem Ergebnis identifizieren können. \_\_\_\_\_

#### Problemlöseverhalten

- Der/die T. bildet Zwischenziele. \_\_\_\_\_
- Der/die T. erkennt die Struktur des vorliegenden Problems. \_\_\_\_\_

#### Aufgaben systematisch erledigen

- Der/die T. sorgt dafür, dass die Gruppe die Aufgabenstellung konsequent beachtet. \_\_\_\_\_
- Der/die T. behält den Gesamtzusammenhang im Auge und verliert sich nicht in Nebensächlichkeiten. \_\_\_\_\_

#### Zielorientierung (nur Vertriebsleiter)

- Der/die T. achtet auf die Einhaltung des mit der Gruppe abgestimmten Weges \_\_\_\_\_
- Der/die T. analysiert Barrieren, die die Zielerreichung behindern. \_\_\_\_\_

Gesamtdurchschnitt: \_\_\_\_\_

### Übung 3

#### **"Stressinterview" (Beobachtermaterial)**

**Aufgabe:** Die TeilnehmerInnen diskutieren mit den Beobachtern über einen konfliktreichen Fall und müssen dabei ihre Position überprüfen.

**Dauer:** Jeweils 20 Minuten.

**Kriterien:** **Umgang mit belastenden Situationen**  
Den geistigen, emotionalen und körperlichen Anforderungen des Verfahrens standhalten.

**Selbständiges Verhalten**

Meinungen vertreten, mit denen die Anderen nicht notwendigerweise übereinstimmen müssen, jedoch auch Bereitschaft zeigen, diese Meinungen im Bedarfsfall zu revidieren.

**Engagement**

Sich ehrgeizig und kontinuierlich am Geschehen beteiligen - auch, wenn es schwer fällt, sich mit fremden Ideen zu identifizieren.

**Geistige Flexibilität**

Sich mit seinen Einschätzungen in Frage stellen, und neue Aspekte anerkennen, die diese modifizieren können.

**Persönliches Auftreten**

Selbstsicher und selbstbewusst anderen Menschen gegenüber treten.

### Übung 3

#### **"Stressinterview" (Teilnehmermaterial)**

##### Fall 3

Sie sind Aussendienstmitarbeiter bei einer Versicherung, die in starkem Konkurrenzkampf zu Mitbewerbern auf dem Markt steht. Sie persönlich konnten sich in diesem Konkurrenzkampf immer gut behaupten und sind dadurch ihrem Vorgesetzten positiv aufgefallen.

Nach einiger Zeit teilte er Ihnen mit, dass er Sie aufgrund Ihrer Erfolge für höhere Aufgaben vorsehen würde.

Sie sitzen gerade im Büro an Ihrem Arbeitsplatz, als Ihnen ein Telefax auf den Tisch gelegt wird. Folgendes lesen Sie darin:

**"Bitte nehmen Sie den Termin bei meinem Freund Peters von der Maschinenbau AG wahr. Ich muss leider in der Zentrale noch einiges regeln und kann nicht rechtzeitig vor Ort sein."**

Da Sie sehr eng mit Ihrem Vorgesetzten zusammenarbeiten, wissen Sie, dass es sich bei dem Termin bei Herrn Peters um ein Verkaufsgespräch handelt, bei dem noch einige strittige Konditionen geklärt werden müssen, um einen Abschluss in Millionenhöhe zu erreichen.

Unverzüglich machen Sie sich auf den Weg. Herr Peters empfängt Sie sehr freundlich und bewirtet Sie mit Kaffee und Kuchen. Sie sind mitten im Gespräch, als Herr Peters von seiner Sekretärin aus dem Arbeitszimmer gerufen wird.

Sie betrachten sich das Arbeitszimmer etwas genauer und sehen plötzlich neben sich auf einem Beistelltisch Unterlagen eines Ihnen bekannten Mitbewerbers liegen. Handschriftliche Ausführungen auf den Unterlagen deuten darauf hin, dass es sich um ein Angebot handelt.

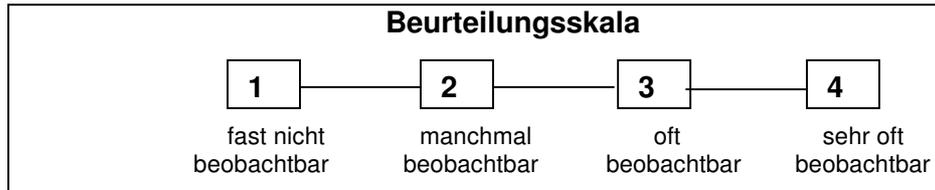
**'Wüsste ich, was die Konkurrenz macht, könnte ich leichter verhandeln' denken Sie sich.**

Was machen Sie?

### Übung 3

#### "Stressinterview" (Beobachterbogen)

Beobachter: \_\_\_\_\_ Teilnehmer: \_\_\_\_\_



#### **Umgang mit belastenden Situationen**

- Der/die T. kann sich nach frustrierenden Erlebnissen wieder aufbauen \_\_\_\_\_
- Der/die T. lässt keine körperlichen Stressmerkmale (z.B. zitternde Hände) erkennen \_\_\_\_\_

#### **Selbständiges Verhalten**

- Der/die T. stellt die eigene Meinung differenziert dar. \_\_\_\_\_
- Der/die T. vertraut der eigenen Urteilsfähigkeit. \_\_\_\_\_

#### **Engagement**

- Der/die T. geht spontan an die Aufgabe heran. \_\_\_\_\_
- Der/die T. kämpft für die eigenen Einschätzungen. \_\_\_\_\_

#### **Geistige Flexibilität**

- Der/die T. erkennt die Problematik, die sich aus der Aufgabenstellung heraus ergibt. \_\_\_\_\_
- Der/die T. korrigiert die eigenen Einschätzungen, wenn triftige Gründe dafür sprechen. \_\_\_\_\_

#### **Persönliches Auftreten**

- Der/die T. wirkt souverän. \_\_\_\_\_
- Der/die T. stellt sich bestimmt und ohne Scheu dar. \_\_\_\_\_

---

**Gesamtdurchschnitt:** \_\_\_\_\_

---

## Übung 4

### "Zweiergespräch" (Beobachtermaterial)

**Aufgabe:** Die TeilnehmerInnen führen ein Kundengespräch mit dem Ziel, ihre Position zu definieren und zu behaupten.

**Dauer:** Jeweils 10 Minuten.

**Kriterien:** **Gestaltung zwischenmenschlicher Beziehungen**  
Zwischenmenschliche Beziehungen anbahnen, entwickeln und gestalten.

**Engagement**

Sich ehrgeizig und kontinuierlich am Geschehen beteiligen - auch, wenn es schwerfällt, sich mit fremden Ideen zu identifizieren.

**Geistige Flexibilität**

Sich mit seinen Einschätzungen in Frage stellen, und neue Aspekte anerkennen, die diese modifizieren können.

**Verbalisierung**

Sich sprachlich auch bei schwierigen Sachverhalten klar und deutlich ausdrücken.

**Überzeugungsfähigkeit**

So geschickt argumentieren, dass andere sich in ihrer Meinung anschliessen.

Thema 1: Erschwerung bei einer PKV  
Thema 2: Auszahlung der LV  
Thema 3: Kündigung der LV  
Thema 4: Mahn- und Kündigungsschreiben  
Thema 5: Verstoss gegen die Bestimmungen der Privathaftpflicht

## Übung 4

### "Zweiergespräch" (Teilnehmermaterial)

#### Thema 1 / Kunde

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Bei der.....Versicherung haben Sie einen Antrag auf eine private Krankenversicherung gestellt. Nach einiger Zeit erhalten Sie die Mitteilung, dass diesem Antrag nur zu erschwerten Bedingungen stattgegeben werden kann.

Sie verstehen das nicht. Manchmal überkommt Sie etwas Hustenreiz, aber Ihr Arzt hat Ihnen gesagt, dass alles in Ordnung ist und Sie kerngesund seien.

Sie gehen nun zum Mitarbeiter der .....Versicherung in die Geschäftsstelle, um den Grund der Erschwerung zu erfahren.

→ Gehen Sie in dem Gespräch so weit, dass Sie drohen, zu einer anderen Versicherung zu gehen

---

#### Thema 1 / Mitarbeiter

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Aufgrund der Auskunft seines Hausarztes konnten Sie den Antrag eines alten Kunden bezüglich einer Krankenversicherung nur unter erschwerten Konditionen annehmen. Der Grund: dem Bericht zufolge leidet der Kunde an einer chronischen Bronchitis. Diese wiederum ist darauf zurückzuführen, dass er ein starker Raucher ist.

Sie sind auf der Geschäftsstelle und erfahren, dass der Kunde mit Ihnen reden möchte.

## Übung 4

### "Zweiergespräch" (Teilnehmermaterial)

#### Thema 2 / Kunde

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel! Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Kunde bei der .....Versicherung. Nun haben Sie ein neues Haus gebaut und brauchen dringend finanzielle Mittel.

Um diese zu bekommen, wollen Sie Ihre Lebensversicherung kündigen und die Beiträge, die Sie in den letzten zwei Jahren eingezahlt haben, zurückverlangen. Dieses Geld würde Ihnen aus den größten Schwierigkeiten heraushelfen. Sie gehen auf die Geschäftsstelle.

-> Werden Sie bei Problemen ärgerlich (z.B.: Wenn man von Versicherungen etwas will, kneifen die immer.... bauen selbst Paläste, aber wir bleiben auf der Strecke....wollen nur unser Geld.. .... hauen uns mit Ihrem Kleingedruckten über' s Ohr etc.!

---

#### Thema 2 / Mitarbeiter

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel! Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Mitarbeiter bei der .....Versicherung und vor Ihnen ist ein Versicherungsnehmer.

Er hat folgenden Vertrag:

Versicherungsnummer: 0012 354 12 / w  
Versicherungssumme: 150.000 DM  
Versicherungsdauer: 40 Jahre  
Rückkaufwert: gering

Der Vertrag besteht seit zwei Jahren.

## Übung 4

### "Zweiergespräch" (Teilnehmermaterial)

#### **Thema 3 / Kunde**

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Seit fünf Jahren sind Sie Kunde bei der .....Versicherung. Ihre dort laufende Lebensversicherung hat eine Versicherungssumme von 25.000 DM bei einer jährlichen Zahlung von 944 DM.

Wegen einer Umschulung an einem anderen Ort werden Sie in den nächsten 3 Jahren jeden Pfennig brauchen. Aus dem Grund konnten Sie auch den letzten Beitrag nicht bezahlen.

Am liebsten wäre es Ihnen, wenn Sie die Beiträge nicht mehr zahlen müssten, also die Lebensversicherung kündigen könnten.

Sie gehen auf die Geschäftsstelle.

---

#### **Thema 3 / Mitarbeiter**

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Mitarbeiter bei der .....Versicherung. Vor Ihnen sitzt ein Kunde, der eine Lebensversicherung bei Ihnen besitzt. Seine Daten:

- Zahlungsweise: 1/1 944 DM
- Versicherungssumme: 25.000 DM
- seit 5 Jahren im Bestand

An die Zahlung des letzten Beitrags musste erinnert werden!

Dieser Kunde sitzt vor Ihnen.

## Übung 4

### "Zweiergespräch" (Teilnehmermaterial)

#### Thema 4 / Kunde

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Kunde bei der .....Versicherung und haben vor einem Monat ein Erinnerungsschreiben erhalten. Heute erhielten Sie mit der Post ein Mahn- und Kündigungsschreiben, dessen Inhalt Sie nicht verstehen.

Das Schreiben wundert Sie ganz besonders, weil Sie seit 10 Jahren nie etwas an Ihrer KFZ- Versicherung (bezogen auf die Beitragserstattung) ändern mussten. Und es ist auch nie etwas bemängelt worden.

Wegen hoher Kontoführungsgebühren haben Sie Ihr bisheriges Konto aufgelöst und ein Postscheckkonto eröffnet. Übrigens: vor dem Erinnerungsschreiben hatten Sie einen Brief wegen Ihrer Lastschrift erhalten; die Angelegenheit hatten Sie aber damals aus Zeitgründen nicht weiter verfolgt.

-> Fragen Sie bitte zuerst nach der Bedeutung des heutigen Schreibens.

---

#### Thema 4 / Mitarbeiter

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel!  
Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Mitarbeiter der .....Versicherung und vor Ihnen sitzt ein Kunde. Einige Daten zu seiner KFZ - Versicherung:

- 10 Jahre im Bestand
- Bankeinzug durch Lastschrift

Die von Ihnen vorgenommene Lastschrift wurde nicht eingelöst. Auch ein zweiter Versuch verlief bei der Bank ergebnislos. Der Kunde reagierte nicht. Auch ein Erinnerungsschreiben nützte nichts. Deshalb haben Sie vor ein paar Tagen ein Mahn- und Kündigungsschreiben nach Paragraph 39 VVG an den Kunden abgesandt.

## Übung 4

### "Zweiergespräch" (Teilnehmermaterial)

#### Thema 5 / Kunde

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel! Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind alter Kunde bei den .....Versicherung. Vor zwei Monaten haben Sie Ihrem zuständigen AD-Mitarbeiter angekündigt, dass Sie in einem Radio- und Fernsehgeschäft aus Unachtsamkeit eine Videokamera umgeworfen haben. Den entstandenen Schaden sollte die Haftpflichtversicherung zahlen.

Sie haben diesbezüglich grünes Licht bekommen und ein Schadensformular ausgefüllt. Das Geschäft schickte an Sie die Reparaturrechnung in Höhe von 675,40 DM, die Sie auch prompt bezahlten.

Nun schicken Sie die Rechnung an die Versicherung in der Erwartung, dass Ihnen die Schadenssumme gutgeschrieben wird. Statt dessen erhalten Sie vom Fernsehgeschäft einen an das Geschäft adressierten Orderscheck mit der Schadenssumme.

Was das nur soll? Sie gehen auf die Geschäftsstelle.

---

#### Thema 5 / Mitarbeiter

**Instruktion:** Bitte versetzen Sie sich kurz in die vor ihnen liegende Rollenvorgabe und entwickeln daraus ein Gespräch mit einem Ihrer Kollegen. Lesen Sie sich die Rollenvorgabe genau durch. Phantasieren Sie nicht zuviel! Für das eigentliche Gespräch haben Sie 10 Minuten lang Zeit.

Sie sind Mitarbeiter der .....Versicherung und haben sich an Vorschriften zu halten.

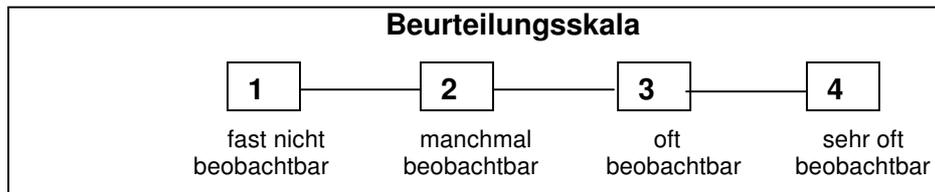
Dazu zählt auch, dass Sie Kunden bspw. darauf hinweisen, dass eigenmächtige Schadensregulierungen nach Paragraph 7 der allgemeinen Geschäftsbedingungen einen Vertragsbruch darstellen. Der Kunde verliert damit zusätzlich seinen Anspruch aus dem Versicherungsvertrag - er geht leer aus (die Versicherung zahlt die vom Kunden an den Geschädigten erstattete Schadenssumme an den Kunden in der Regel nicht aus) und verliert seine Police.

Ein alter Kunde hat Sie um ein Gespräch gebeten.

## Übung 4

### "Zweiergespräch" (Beobachterbogen)

Beobachter: \_\_\_\_\_ Teilnehmer: \_\_\_\_\_



#### **Gestaltung zwischenmenschlicher Beziehungen**

- Der/die T. vermittelt emotionale Wärme. \_\_\_\_\_ / \_\_\_\_\_
- Der/die T. zeigt durch verbales und nonverbales Verhalten Interesse am Gesprächspartner. \_\_\_\_\_ / \_\_\_\_\_

#### **Engagement**

- Der/die T. geht spontan an die Aufgabe heran. \_\_\_\_\_ / \_\_\_\_\_
- Der/die T. kämpft für die eigenen Einschätzungen. \_\_\_\_\_ / \_\_\_\_\_

#### **Geistige Flexibilität**

- Der/die T. betrachtet die Argumente des Gegenübers von mehreren Seiten. \_\_\_\_\_ / \_\_\_\_\_
- Der/die T. stellt eigene Argumente zur Diskussion. \_\_\_\_\_ / \_\_\_\_\_

#### **Verbalisierung**

- Der/die T. baut seine Aussagen didaktisch geschickt auf \_\_\_\_\_ / \_\_\_\_\_
- Der/die T. stellt eigene Gedanken nachvollziehbar dar. \_\_\_\_\_ / \_\_\_\_\_

#### **Überzeugungsfähigkeit**

- Der/die T. steuert das Gespräch. \_\_\_\_\_ / \_\_\_\_\_
- Der/die T. entkräftet Einwände. \_\_\_\_\_ / \_\_\_\_\_

---

**Gesamtdurchschnitt:** \_\_\_\_\_ / \_\_\_\_\_

---

## Übung 5

### "Gruppenübung 2" (Beobachtermaterial)

- Aufgabe:** Die TeilnehmerInnen erfahren von den Problemen eines Bekannten. Sie ergründen mögliche Ursachen und entwickeln gemeinsam Lösungsvorschläge. Das Ergebnis der Diskussion wird von einem Gruppenmitglied präsentiert.
- Typus:** Gruppenübung ohne Konkurrenzcharakter.
- Dauer:** 15 Minuten Einzelarbeit  
45 Minuten Gruppenarbeit
- Kriterien:**
- Gestaltung zwischenmenschlicher Beziehungen**  
Zwischenmenschliche Beziehungen anbahnen, entwickeln und gestalten.
  - Selbständiges Verhalten**  
Meinungen vertreten, mit denen die anderen nicht notwendigerweise übereinstimmen müssen, jedoch auch Bereitschaft zeigen, diese Meinungen im Bedarfsfall zu revidieren.
  - Feedback übermitteln**  
Eigene Einschätzungen, Empfindungen, Bewertungen und Positionen übermitteln.
  - Teambildung unterstützen**  
Gruppenmitglieder in den Gruppenprozess integrieren und ihnen ihre Verantwortung für den Gruppenerfolg deutlich machen.
  - Problemlöseverhalten**  
Probleme Prozeduren unterziehen, die deren Lösung gewährleisten
  - Aufgaben systematisch erledigen**  
Komplexe Sachverhalte systematisch bearbeiten und erledigen.
  - Zielorientierung (nur Vertriebsleiter)**  
Gruppen- und Einzelaktivitäten hinsichtlich ihrer Bedeutung für das zu erreichende Ziel beeinflussen.

## Übung 5

### "Gruppenübung 2" (Teilnehmermaterial)

Ein Bekannter hat Probleme

Irgendwann einmal werden Sie von einem guten Bekannten abends auf ein Glas Bier eingeladen. Mit fortschreitender Dauer des Abends wächst das ohnehin schon zwischen Ihnen bestehende Vertrauen. Schliesslich schüttet sich Ihr Bekannter bei Ihnen das Herz aus:

*"Also, ich sage Dir, so kann es nicht weitergehen. Ich habe auch schon mit Inge darüber gesprochen, aber die weiss auch keinen Rat. Mein Chef spinnt. Dauernd liegt er mir mit seinen Projektberichten in den Ohren. Und dann nervt der mich auch noch mit meinen Mitarbeiterproblemen. Teufel auch noch, woher soll ich die Leute denn nehmen? Und dann stellen die sich auch noch so blöde an. Was kann ich denn dazu? Ich kenne das doch schon lange. Der Markt gibt nun einmal nichts Besseres her. Schliesslich muss mein Chef das doch auch wissen, und er muss wissen, dass ich etwas kann. Nicht umsonst hat er mich zum Projektleiter gemacht. Und die Inge wollte ja auch, dass ich diesen Job annehme. Ich habe keine Probleme mit der Entscheidung gehabt.*

*Alle Aufgaben erfülle ich gewissenhaft und sorgsam. Da kann kommen, was wolle. Ich verbürge mich dafür. Du kennst mich ja. Was soll das ganze Trara meines Chefs also? Ich kann nicht mehr tun, als mich krummlegen und schuffen. Nicht umsonst habe ich im letzten Jahr auf meinen Urlaub verzichtet. Dem Unternehmen und der eigenen Karriere zuliebe muss man nun 'mal Opfer bringen. Inge und die Kinder haben zwar gemeutert, aber das war auch bald ausgestanden. Schliesslich wollen ja alle den derzeitigen Lebensstandard halten.*

*Seit ich Beta-Blocker nehme, geht es mit meinen Herzrythmusstörungen auch besser. Nur machen die Pillen einen so schrecklich müde und depressiv. Ich fühle mich apathisch und schlapp, nicht mehr so leistungsfähig wie früher. Ich kann nicht mehr so von einem Termin zum anderen hetzen.*

*Und trotzdem packe ich es noch. Doch was macht mein Chef? Der nervt mich ständig. Und auch die Inge nörgelt in letzter Zeit an mir herum. Das macht mich noch ganz fertig. Was soll ich tun?"*

Was raten Sie Ihrem Bekannten?

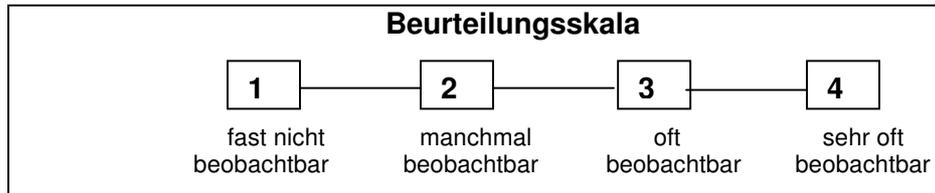
- Instruktion**      Gehen Sie bei der Beantwortung der Frage bitte folgendermassen vor:
- 1) Analysieren Sie zuerst alleine und dann in der Gruppe den Fall unter folgenden Gesichtspunkten:
    - a) Welche Ursachen haben zur jetzigen Situation geführt?
    - b) Was für ein Typ Mensch scheint Ihr Bekannter zu sein?
    - c) Was raten Sie Ihrem Bekannten?
  - 2) Legen Sie bitte am Ende der Gruppendiskussion fest, wer das Ergebnis präsentieren soll.

Sie haben für die Einzelarbeit 15 Minuten, für die Gruppenarbeit 45 Minuten lang Zeit.

## Übung 5

### "Gruppenübung 2" (Beobachterbogen)

Beobachter: \_\_\_\_\_ Teilnehmer: \_\_\_\_\_



#### Gestaltung zwischenmenschlicher Beziehungen

- Der/die T. geht Konflikte produktiv ein. \_\_\_\_\_
- Der/die T. ist darum bemüht, dass andere die eigenen Positionen und Einschätzungen verstehen. \_\_\_\_\_

#### Selbständiges Verhalten

- Der/die T. vertraut der eigenen Urteilsfähigkeit. \_\_\_\_\_
- Der/die T. kann die eigene Meinung differenziert darstellen. \_\_\_\_\_

#### Feedback übermitteln

- Der/die T. äussert Kritik, ohne andere zu verletzen \_\_\_\_\_
- Der/die T. reflektiert Gruppenprozesse auf der Basis eigener Empfindungen. \_\_\_\_\_

#### Teambildung unterstützen

- Der/die T. bezieht Gruppenmitglieder in die Diskussion mit ein. \_\_\_\_\_
- Der/die T. sorgt dafür, dass sich alle Gruppenmitglieder mit dem Ergebnis identifizieren können. \_\_\_\_\_

#### Problemlöseverhalten

- Der/die T. bildet Zwischenziele. \_\_\_\_\_
- Der/die T. erkennt die Struktur des vorliegenden Problems. \_\_\_\_\_

#### Aufgaben systematisch erledigen

- Der/die T. sorgt dafür, dass die Gruppe die Aufgabenstellung konsequent beachtet. \_\_\_\_\_
- Der/die T. behält den Gesamtzusammenhang im Auge und verliert sich nicht in Nebensächlichkeiten. \_\_\_\_\_

#### Zielorientierung (nur Vertriebsleiter)

- Der/die T. achtet auf die Einhaltung des mit der Gruppe abgestimmten Weges \_\_\_\_\_
- Der/die T. analysiert Barrieren, die die Zielerreichung behindern. \_\_\_\_\_

Gesamtdurchschnitt: \_\_\_\_\_

**Anhang 3**

## Verhaltensbegriffe des Basis-Beobachtungsinventars und ihre Definitionen

### **Umgang mit belastenden Situationen**

Den geistigen, emotionalen und körperlichen Anforderungen des Verfahrens standhalten.

### **Gestaltung zwischenmenschlicher Beziehungen**

Zwischenmenschliche Beziehungen anbahnen, entwickeln und gestalten.

### **Selbständiges Verhalten**

Meinungen vertreten, mit denen die anderen nicht notwendigerweise übereinstimmen müssen, jedoch auch Bereitschaft zeigen, diese Meinungen im Bedarfsfall zu revidieren.

### **Feedback übermitteln**

Eigene Einschätzungen, Empfindungen, Bewertungen und Positionen übermitteln.

### **Teambildung unterstützen**

Gruppenmitglieder in den Gruppenprozess integrieren und ihnen ihre Verantwortung für den Gruppenerfolg deutlich machen.

### **Engagement**

Sich ehrgeizig und kontinuierlich am Geschehen beteiligen - auch wenn es schwer fällt, sich mit fremden Ideen zu identifizieren.

### **Geistige Flexibilität**

Sich mit seinen Einschätzungen in Frage stellen, und neue Aspekte anerkennen, die diese modifizieren können.

### **Problemlöseverhalten**

Probleme Prozeduren unterziehen, die deren Lösung gewährleisten.

### **Persönliches Auftreten**

Selbstsicher und selbstbewußt anderen Menschen gegenüberzutreten.

### **Schrittweises und kontrolliertes Arbeiten**

Komplexe Sachverhalte systematisch bearbeiten und erledigen.

### **Verbalisierung**

Sich sprachlich auch bei schwierigen Sachverhalten klar und deutlich ausdrücken.

### **Überzeugungsfähigkeit**

So geschickt argumentieren, dass andere sich in ihrer Meinung anschließen.

### **Zielorientierung**

Gruppen- und Einzelaktivitäten hinsichtlich ihrer Bedeutung für das zu erreichende Ziel beeinflussen.

**Verhalten als Moderator**

Gruppenaktivitäten koordinieren, so dass die Gruppendiskussion zu einem Ergebnis geführt wird.

**Präsentationsverhalten**

Eigene Meinungen und Sachverhalte kompetent und transparent einem Zuhörerkreis nahe bringen.

**Kooperationsverhalten**

Sich anderen als Ansprechpartner und Vertrauensperson zur Verfügung stellen.

**Konkurrenzverhalten**

Konkurrenz zu anderen Gruppenmitgliedern aufnehmen und austragen.

**Sensibilität**

Einfühlungsvermögen in Menschen und im Miteinander beweisen.

**Führungsverhalten**

Die Rolle des Führers als 'Entwickler und Partner' vorleben.

**Risikobereitschaft**

Entscheidungen treffen, deren Auswirkungen nicht genau vorhersehbar sind.

**Ergebnisprofil für Vertriebsleiter Soll / Ist**

	1	1,5	2	2,5	3	3,5	4
Geistige Flexibilität							
Problemlöseverhalten							
Schrittweises u. kontr. Arbeiten							
Verbalisierung							
Gestaltung zwm. Beziehungen							
Feedback übermitteln							
Teambildung unterstützen							
Selbständiges Verhalten							
Überzeugungsfähigkeit							
Zielorientierung							
Umgang mit belast. Situationen							
Engagement							
Persönliches Auftreten							

Name des / der Teilnehmer(in)s: \_\_\_\_\_ Datum \_\_\_\_\_

**Ergebnisprofil für Versicherungsverkäufer Soll / Ist**

	1	1,5	2	2,5	3	3,5	4
Geistige Flexibilität							
Problemlöseverhalten							
Schrittweises u. kontr. Arbeiten							
Verbalisierung							
Gestaltung zwm. Beziehungen							
Feedback übermitteln							
Teambildung unterstützen							
Selbständiges Verhalten							
Überzeugungsfähigkeit							
Zielorientierung							
Umgang mit belast. Situationen							
Engagement							
Persönliches Auftreten							

Name des / der Teilnehmer(in)s: \_\_\_\_\_ Datum \_\_\_\_\_

**Verteilung der Verhaltensdimensionen auf die Übungen  
Vertriebsleiter | Versicherungsverkäufer**

	SE	GR1	ST	GE	GR2
Geistige Flexibilität					
Problemlöseverhalten					
Schrittweises u. kontr. Arbeiten					
Verbalisierung					
Gestaltung zwm. Beziehungen					
Feedback übermitteln					
Teambildung unterstützen					
Selbständiges Verhalten					
Überzeugungsfähigkeit					
Zielorientierung					
Umgang mit belast. Situationen					
Engagement					
Persönliches Auftreten					

**GR1= Gruppenarbeit 1**  
**GE = Zweiergespräch**  
**ST = Stressinterview**  
**SE = Selbstpräsentation**  
**GR2= Gruppenarbeit 2**

-> Zielorientierung ist für die Versicherungsverkäufer nicht relevant! !

**Ergebnisprotokoll qualitative Aussage**

**Datum:** \_\_\_\_\_ **TeilnehmerIn:** \_\_\_\_\_

Gesprächspartner des Rückmeldegesprächs: \_\_\_\_\_  
\_\_\_\_\_

Der/Die TeilnehmerIn wird.....:

- **ohne Einschränkungen** für eine Einstellung empfohlen.  
Begründung

---

---

---

---

- **mit Einschränkungen** für eine Einstellung empfohlen.  
Begründung:

---

---

---

---

- **nicht** für eine Einstellung empfohlen.  
Begründung:

---

---

---

---

**Unterschrift Beobachtkonferenz:**

---

**Zeitplan zum Auswahl - AC  
für Versicherungsverkäufer**

**1. Tag**

12.00 Uhr	Anreise, Vorbereitungen
13.00 Uhr	gemeinsames Mittagessen
14.15 Uhr	Beobachterinstruktion
18.00 Uhr	Teilnehmerinstruktion
19.30 Uhr	gemeinsames Abendessen

**2. Tag**

08.30 Uhr	<b>01 Selbstpräsentation</b> 10 Minuten Einzelarbeit 10 * 5 Minuten Präsentation
09.45 Uhr	<b>02 Gruppenübung 1</b> 15 Minuten Einzelarbeit 45 Minuten Gruppendiskussion
10.45 Uhr	<b>Pause</b>
11.00 Uhr	<b>03 Stressinterview</b> 2 Gruppen á 5 Teilnehmer á 15 Minuten <b>06 Postkorb</b>
12.30 Uhr	<b>Mittagessen</b>
13.30 Uhr	<b>04 Zweiergespräche</b> 5 * 10 Minuten
14.45 Uhr	<b>05 Gruppenübung 2</b> 15 Minuten Einzelarbeit 45 Minuten Gruppenarbeit
15.45 Uhr	<b>07 Planungsaufgabe</b> <b>08 Selbsteinschätzung</b>
	<b>Beobachterkonferenz</b>
17.00 Uhr	<b>Rückmeldegespräche</b>
18.00 Uhr	<b>Abschließende Beobachterkonferenz</b>
21.00 Uhr	<b>Ende des AC</b>

## **EIGENSTÄNDIGKEITSERKLÄRUNG**

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und dabei keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Die Stellen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, wurden durch Quellenangaben im Text deutlich gemacht.

Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde bzw. Promotionskommission vorgelegen.

Bad Honnef, den 27. Oktober 2004

Thomas Randhofer