

**Assessment-Center in der Personalauswahl:  
Wunsch und Wirklichkeit**

Bachelorarbeit  
im Studiengang B. Sc. Psychologie

FernUniversität in Hagen  
Fakultät Kultur- und Sozialwissenschaften  
Institut für Psychologie  
Lehrgebiet Arbeits- und Organisationspsychologie

Verfasser: Sonja Moreira Coutinho  
Abgabedatum: 20.02.2014

## **Zusammenfassung**

Assessment Center sind beliebte und häufig eingesetzte Verfahren in der Personalauswahl. Richtlinien spezifizieren wie die Anwendung der Methode in der Praxis aussehen soll. Ein Vergleich der Standards der International Task Force on Assessment Center Guidelines (2009), des Arbeitskreis Assessment Center e.V. (2004) und von Kanning (2004) identifiziert Mängel in den deutschen Manuals, besonders bei Kanning in der Detailliertheit. Eine Übersicht aus nationalen und internationalen Studien zeigt, dass viele der postulierten Empfehlungen in den Unternehmen Einzug gefunden haben. Hinsichtlich einiger Aspekte wie der Evaluation oder dem Technologieeinsatz besteht erheblicher Verbesserungsbedarf. Kleine Stichproben und die Verwendung von Fragebögen führen zu einer Diskussion über die Grenzen und Einschränkungen der Studien. Es wird argumentiert, dass die Erarbeitung einheitlicher Standards die Handhabung des Assessment Center in der Praxis verbessern kann.

*Stichwörter:* Assessment Center, Personalauswahl, Standards, Richtlinien, Empfehlung, Organisation

## **Abstract**

Assessment centers are popular and frequently used methods in the personnel selection. Guidelines specify how the application in the practice should be. A comparison of the standards of the International Task Force on Assessment Center Guidelines (2009), the Arbeitskreis Assessment Center e.V. (2004) and Kanning (2004) reveal some faults in the German manuals, especially in the detailedness by Kanning. An overview of national and international studies appears that many of the postulated recommendations arrived in the organizations. In reference to some issues like evaluation or use of technology exists a substantial need for improvement. Small sample sizes and the use of questionnaires lead to a discussion of study limitations. It will be argued, that the development of uniform standards can improve the management of the assessment center in the practice.

*keywords:* assessment center, personnel selection, standard, guideline, recommendation, organization

## Inhaltsverzeichnis

Zusammenfassung .....	ii
Abstract .....	ii
Inhaltsverzeichnis.....	iii
Tabellenverzeichnis.....	v
Abkürzungsverzeichnis .....	vi
1. Einleitung .....	1
2. Methode.....	2
3. Assessment Center .....	3
3.1 Standards des Assessment Center .....	3
3.1.1 Guidelines and Ethical Consideration .....	3
3.1.1.1. Elemente des Assessment Center .....	4
3.1.1.2 Grundsatzerklärung.....	6
3.1.1.3 Informiertheit und Rechte der Teilnehmer .....	6
3.1.1.4 Evaluation.....	7
3.1.1.5 Interkulturelle Anwendung.....	7
3.1.1.6 Abgrenzung des Assessment Center.....	7
3.1.2 Standards des Arbeitskreis Assessment Center e.V. ....	8
3.1.3 Privatstandards am Beispiel Kanning .....	10
3.2 Vergleich der Standards .....	11
4. Umsetzung in der Praxis .....	17
4.1 Überblick der Studien.....	17
4.2. Anforderungsanalyse.....	18
4.3 Verhaltensklassifikation .....	22
4.4 Multiple Beurteilungen insbesondere durch Simulationen .....	24
4.5 Beobachter.....	28
4.6 Beobachtertraining .....	31

4.7 Verhaltensaufnahme .....	32
4.8 Datenintegration .....	33
4.9 Grundsatzerklärung .....	34
4.10 Informiertheit und Rechte der Teilnehmer .....	35
4.11 Evaluation.....	38
4.12 Technologieeinsatz.....	40
4.13 Interkulturalität.....	41
4.14 Entwicklung .....	42
5. Diskussion .....	43
5.1 Zusammenfassung der Ergebnisse .....	43
5.2 Mängel der verwendeten Studien.....	45
5.3 Resümee zu den Standards .....	47
5.4 Einschränkungen dieses Reviews.....	49
5.5 Fazit und Ausblick .....	50
Literaturverzeichnis.....	51
Eidesstattliche Erklärung.....	59

## Tabellenverzeichnis

Tabelle 1: Vergleich der deutschen Richtlinien mit den internationalen Empfehlungen.....	11
Tabelle 2: Überblick der verwendeten Studien zur Umsetzung des Assessment Center in der Praxis.....	18
Tabelle 3: Anwendung der Anforderungsanalyse und deren verschiedenen Methoden in Prozentangaben an der Stichprobengröße n .....	19
Tabelle 4: Arten von verwendeten Simulationsaufgaben in Prozentangaben an der Stichprobengröße n .....	25
Tabelle 5: Einsatz von weiteren Verfahren ohne Simulationscharakter in Prozent als Anteil an der Stichprobengröße n.....	27
Tabelle 6: Beruflicher Hintergrund der Beobachter in Prozentangaben an der Stichprobengröße n .....	29
Tabelle 7: Prozentangaben der berücksichtigten Merkmale bei der Beobachterausswahl als Anteil an der Stichprobengröße n .....	30
Tabelle 8: Varianten und Zeitpunkt des Feedbacks in Prozent als Anteil an der Stichprobengröße n .....	37
Tabelle 9: Prozentangaben der Evaluation und erhobenen Gütekriterien an der Stichprobengröße n .....	39
Tabelle 10: Der Einsatz von Computern in einem Assessment Center in Prozentangaben an der Stichprobengröße n.....	41
Tabelle 11: Entwicklung und Revision des Assessment Center in Prozent an der Stichprobengröße n.....	43

## **Abkürzungsverzeichnis**

AC	Assessment Center
AkAC	Arbeitskreis Assessment Center e. V.
ITFACG	International Task Force on Assessment Center Guidelines

## 1. Einleitung

Jeden Tag sind Unternehmen weltweit auf der Suche nach geeignetem Personal für eine zu besetzende Position. Es ist nicht selten, dass sich auf Stellenanzeigen beliebter Firmen mehrere hunderte Leute bewerben, wodurch sich die Auswahl erschwert. Den falschen Kandidaten<sup>1</sup> einzustellen ist mit finanziellen Kosten verbunden (Nerdinger, Blickle & Schaper, 2011; Obermann, 2013). Wichtig ist deshalb die Selektion anhand geeigneter Personalauswahlverfahren. Eine Möglichkeit hierfür ist das Assessment Center (AC). Hierbei handelt es sich um eine Kombination aus verschiedenen eignungsdiagnostischen Methoden (Höft & Funke, 2006; Schuler, 2007). Ziel war durch die Verbindung verschiedener plausibler Techniken ein Verfahren zu entwickeln, welches eine valide Entscheidung ermöglicht (Obermann, 2013). Während in älteren Metaanalysen eine mittlere operationale Validität von  $\rho = .37$  gefunden wurde (Gaugler, Rosenthal, Thornton & Bentson, 1987; Schmitt & Hunter, 1998), erreicht das AC in neueren Studien nur noch einen mittleren Wert von  $\rho = .26$  (Hardison & Sackett, 2007). Trotzdem gewann das AC weltweit an Beliebtheit, weil Unternehmen von der Zielerreichung überzeugt waren und Bewerber sich wertgeschätzt fühlten (Hell, Boramir, Schaar & Schuler, 2006; Nerdinger et al., 2011; Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Weltweit machten sich Wissenschaftler und Praktiker auf die Suche nach Verbesserungsmöglichkeiten des Verfahrens und seiner diagnostischen Aussagekraft. Als Resultat entstanden spezielle Empfehlungen für die Entwicklung, Anwendung und Auswertung des AC.

In der vorliegenden Arbeit werden zunächst nach einem Überblick über die Literaturrecherche diese Standards erläutert. Dabei werden die Richtlinien der International Task Force on Assessment Center Guidelines (ITFACG, 2009), des Arbeitskreis Assessment Center e.V. (AkAC, 2004) und Privatstandards von Kanning (2004) genauer beleuchtet. Ein Vergleich der drei Standardgruppen schließt das Kapitel ab. Hierbei werden wichtige Resultate aus der Forschung zur Untermauerung der Relevanz der Richtlinien für die Anwendung des AC genannt. Anschließend werden empirische Befunde zur Umsetzung dieser Empfehlungen in der Praxis referiert. Im letzten Kapitel werden Mängel in der Umsetzung zusammengefasst und die verschiedenen Standards kritisch hinterfragt sowie eigene Gedanken zur Verbesserung geäußert. Einschränkungen der verwendeten Untersuchungen werden ebenfalls diskutiert. Ein Fazit und ein Ausblick bilden den Abschluss der Thesis.

---

<sup>1</sup> Aus Gründen der besseren Lesbarkeit wird auf die konsequente Nennung der männlichen und weiblichen Form verzichtet. Gemeint sind jedoch, wenn nicht explizit anders erwähnt, stets beide Geschlechter.

## 2. Methode

Es handelt sich bei der vorliegenden Thesis um eine theoretische Arbeit ohne eigene Datenerhebung, stattdessen wurde eine Literaturrecherche in Datenbanken und Bibliotheken durchgeführt. Zuerst wurden die Kataloge der Bibliotheken der FernUniversität in Hagen, der Universität des Saarlandes in Verbindung mit der Hochschule für Technik und Wirtschaft des Saarlandes und der Universität Koblenz-Landau mit Standort Landau nach Standardwerken und Lehrbüchern zum Thema durchforstet. Als Suchbegriffe dienten die Wörter *Assessment Center*, *Eignungsdiagnostik*, *Personalauswahl*, *Personalselektion* und *Personalpsychologie*.

Außerdem kam die Metadatenbank EBSCO zum Einsatz. Es wurden die Datenbanken PsycINFO, PSYINDEX, PsycARTICLES, Psychology and Behavioral Sciences und Business Source Complete ausgewählt. Da die Suche mit der Eingabe Assessment Center mehr als 70.000 Ergebnisse lieferte, wobei der Begriff auch beispielsweise in der klinischen Psychologie verwendet wird, wurden verschiedene Varianten durchlaufen. Es wurden alle Kategorien gewählt, die sich dem arbeits- und organisationspsychologischen Bereich zuordnen ließen. So wurden 375 Treffer identifiziert, häufig jedoch mit geringer Relevanz für diese Abhandlung. Zur weiteren Eingrenzung wurden verschiedene Kombinationen mit den Suchbegriffen *Assessment Center* und *guideline*, *standard*, *recommendation*, *Task Force*, *Arbeitskreis*, *practice*, *organization*, *personnel selection* eingegeben. Dabei wurde der Boolesche Operator *and* verwendet. Ebenso wurde bei einem Suchdurchgang festgelegt, dass die verwendeten Begriffe im Titel enthalten sein mussten. Zusätzlich wurden die Literaturverzeichnisse der so recherchierten Texte auf weitere interessante Quellen untersucht. Dabei wurden für Kapitel 4 nur Studien einbezogen, die sich einerseits mit mindestens einem Standard in Verbindung setzen ließen und andererseits sich auf mehrere Organisationen beziehen, um ein repräsentativeres Bild für ein Land zu erhalten. Für die empirischen Grundlagen der Standards wurde eine richtlinienspezifische Suche mit den Begriffen *Assessment Center* und *job analysis*, *observation*, *exercice*, *dimension*, *validity*, *participants*, *assessor*, *culture* durchgeführt

Insgesamt wurden durch dieses Vorgehen 78 relevante Quellen ausgemacht. Davon gingen 17 Arbeiten in das Kapitel Umsetzung in der Praxis ein. Die restlichen Quellen werden in den weiteren Kapiteln referiert. Zwei Untersuchungen werden nur in Sekundärliteratur aufgeführt, da die entsprechenden Studien bisher ausschließlich auf Kongressen berichtet wurden.

### **3. Assessment Center**

Ein AC stellt eine standardisierte Verhaltensbewertung dar, die auf zuvor definierten Dimensionen, mehreren Beobachtern und verschiedenen Verfahren beruht (ITFACG, 2009). Das AC lässt sich den simulationsorientierten Verfahren zuordnen. Typische Bestandteile sind Rollenspiele, Gruppendiskussionen, Postkorbübung, Präsentationen, Fallstudien und Planspiele. Zusätzlich können Tests und Interviews zur Anwendung kommen (Höft & Funke, 2006; Marcus, 2011). Von jedem Bewerber wird ein Profil erstellt, indem Assessoren die Kandidaten auf den relevanten Dimensionen mehrfach einschätzen (ITFACG, 2009). Am Ende werden die einzelnen Beurteilungen entweder durch eine Beobachterkonferenz oder durch statistische Aggregation zu einem Globalurteil zusammengetragen (Höft & Funke, 2006; ITFACG, 2009; Marcus, 2011; Schuler, 2007).

#### **3.1 Standards des Assessment Center**

Die bekanntesten Standards stammen von der ITFACG (2009), welche ursprünglich aus rein US-amerikanischen Mitgliedern bestand, sich mittlerweile aber zu einem internationalen Zusammenschluss entwickelte (Lievens & Thornton, 2005). In Deutschland haben Wissenschaftler und Praktiker die Standards des AkAC (2004) erarbeitet. Neben diesen beiden großen Gruppen von Richtlinien wurden weitere nationale Empfehlungen entwickelt. Dabei orientierten sich die Arbeitsgruppen aus Indonesien, Südafrika und Großbritannien an den Empfehlungen der ITFACG (2009) (Assessment Centre Study Group, 2007; British Psychological Society, 2003; Pedit, 2011). Die Schweizer Richtlinien dagegen bauen auf den deutschen Standards des AkAC (2004) auf, wobei lediglich eine sprachliche Anpassung vorgenommen wurde (Swiss Assessment, 2007). Aus diesem Grund wird auf eine Ausführung der Richtlinien dieser Gruppen verzichtet. Kanning (2004) hat als Einzelperson ebenfalls den Versuch unternommen deutsche Standards abzuleiten. Diese werden zur Abgrenzung von Gremienentscheidungen als Privatstandards bezeichnet (Kersting, 2009).

Im Folgenden werden die Empfehlungen der ITFACG (2009) erläutert. Anschließend erfolgt ein Überblick über die Standards des AkAC (2004) und von Kanning (2004).

##### **3.1.1 Guidelines and Ethical Consideration**

Die internationale Arbeitsgruppe hat auf Basis von Beobachtungen und Erfahrungen von Experten im Jahr 1975 die ersten Richtlinien verabschiedet. Die aktuellen Empfehlungen stammen aus dem Jahr 2009. Bis zu dieser Zeit wurden sie regelmäßig überarbeitet, um Verbesserungen vorzunehmen und die Aktualität zu garantieren (ITFACG, 2009).

Nach Auffassung der Vereinigung ist die Verwendung des Titels „Assessment Center“ auf jene Methoden beschränkt, die allen diesen Richtlinien folgen. Zusätzlich beziehen sich diese Empfehlungen nur auf das AC und nicht auf ähnliche Verfahren wie beispielsweise die Verhaltenssimulation bei einer Individualbewertung. Die Zielgruppe sind Personen, welche ein AC entwerfen oder Entscheidungen über die Einführung treffen sowie Assessoren. Insgesamt besteht das Manual aus neun Empfehlungen, die zur Charakterisierung eines AC herangezogen werden können. Darüber hinaus werden noch weitere zu berücksichtigende Aspekte thematisiert. Wichtig ist die Konzeptualisierung des AC entsprechend seiner definierten Ziele wie beispielsweise Verhaltensvorhersage oder Identifikation von Entwicklungsbedürfnissen (ITFACG, 2009).

### **3.1.1.1. Elemente des Assessment Center**

*Anforderungsanalyse / Kompetenzmodellierung.* Die ITFACG (2009) definiert die Anforderungsanalyse als einen Prozess der Ableitung von erfolgsrelevanten, jobbezogenen Verhaltensdimensionen. Zur Datensammlung können verschiedene Methoden wie Interviews, Beobachtungen oder Stellenbeschreibungen eingesetzt werden. Die Anzahl und Kombination der verwendeten Verfahren hängen von der spezifischen Stelle ab. Am einfachsten ist die Ableitung der Anforderungen, wenn eine Position bereits in dem Unternehmen existiert und lediglich ein Nachfolger gesucht wird. Bei der Erschaffung einer neuen Stelle müssen die Erwartungen an den zukünftigen Stelleninhaber präzisiert werden. Hilfreich können ähnliche Tätigkeiten, Visionen, Strategien, Stellung oder Verantwortung sein. Die Verwendung von Ergebnissen aus früheren Anforderungsanalysen sollten nur bei vergleichbaren Stellen eingesetzt werden. Auch kann man sogenannte *subject matter experts*, also Experten, zu den Anforderungen und Kompetenzen im Kontext einer Kompetenzmodellierung befragen. Am Ende müssen klar definierte Verhaltensdimensionen zielrelevanter Kompetenzen bzw. Anforderungen vorliegen, die während eines AC beobachtet werden können (ITFACG, 2009).

*Verhaltensklassifikation.* Das von Teilnehmern gezeigte Verhalten muss relevanten und bedeutungsvollen Kategorien zugeordnet werden. Beispiele hierfür sind Verhaltenskomponenten, Eigenschaften, Charakteristika, Fähigkeiten, Fertigkeiten oder Wissen (ITFACG, 2009).

*Beurteilungsmethoden.* Mit den verwendeten Verfahren müssen Informationen zur Beurteilung der zuvor definierten Anforderungen gesammelt werden können. Hierfür bietet sich die Verwendung einer Matrix an. Auf diese Weise werden die verschiedenen Methoden und Dimensionen miteinander verbunden (ITFACG, 2009).

*Multiple Beurteilungen.* Verschiedene Methoden wie Simulationen, Tests oder Interviews sind zu verwenden. Ziel dieser Heterogenität ist die Sammlung einer großen Variabilität von Verhaltensweisen und wichtigen Informationen. Zusätzlich kann man als weitere Datenquellen eine Selbstbeurteilung oder das 360-Grad-Feedback verwenden. Nach der Verfahrensauswahl werden diese in einem Pretest auf ihre Reliabilität, Objektivität und Validität überprüft. Dazu gehören ein Probedurchlauf mit vergleichbaren Probanden, die Einschätzung der Akkuratheit und Repräsentativität der Verhaltenserhebung durch Experten sowie Evidenz aus der Anwendung dieser Methode für ähnliche Positionen in vergleichbaren Organisationen. Auf diese Weise kann sichergestellt werden, dass auch tatsächlich alle relevanten Informationen mit den geplanten Verfahren gesammelt werden (ITFACG, 2009).

*Simulation.* Das AC beinhaltet eine ausreichende Anzahl arbeitsbezogener Verhaltenssimulationen. Dabei müssen alle vorher definierten Dimensionen in verschiedenen Übungen beobachtbar sein. Die Anzahl der Verfahren hängt von der Komplexität der Arbeitstätigkeit ab. Bei der Konstruktion der Verfahren sollte unbedingt darauf geachtet werden, wie viele Dimensionen durch eine Aufgabe abgedeckt werden. Sind dies zu viele, kann es sehr leicht zu einer kognitiven Überlastung bei den Assessoren kommen, sodass diese nicht in der Lage sind, die wichtigen Informationen herauszulesen. Außerdem müssen die Simulationen unterschiedliche Antworten zulassen, sodass die Teilnehmer nicht aus einer bestimmten Anzahl von Antworten die richtige auswählen, sondern sich selbst Gedanken um einen Lösungsweg machen müssen. Die Abgrenzung von Simulationen zu anderen Verfahren wie dem Situationalen Interview erfolgt durch die Darbietung von tatsächlichem und nicht von hypothetischem Verhalten. Die Übungen sind somit sehr realitätsbezogen (high fidelity). Zur Präsentation der Aufgaben ist auch der Einsatz von technischen Hilfsmitteln wie Video oder Computer möglich (ITFACG, 2009).

*Beobachter.* Jeder Teilnehmer wird von mehreren Assessoren beobachtet und beurteilt. Die Gruppe der Beobachter sollte heterogen im Hinblick auf Geschlecht, Alter, ethnische Herkunft, Position, berufliche Erfahrung und Ausbildung zusammengestellt werden. Das Teilnehmer-Beobachter-Verhältnis hängt unter anderem von der Komplexität der Übung und der Anzahl der Dimensionen ab (ITFACG, 2009).

*Beobachtertraining.* Vor dem Einsatz in einem AC werden die Beobachter entsprechend geschult. Zu den Trainingsinhalten gehören Kenntnisse der Dimensionen, Übungen, Beobachtung, Aufnahme, Klassifikation und Bewertung des relevanten Verhaltens sowie möglicher Fehlerquellen. Zusätzlich können in Abhängigkeit von den Zielen weitere Aspekte vermittelt werden. Nach Absolvieren des Trainings sollten die Beobachter neben den bereits

genannten Informationen Wissen über die Organisation und ihre speziellen Beurteilungsrichtlinien, die Zielposition, den Zusammenhang der Dimensionen mit der Übung sowie der Arbeitsstelle und Arten der Datenintegration erlangt haben. Darüber hinaus demonstrierten die Assessoren ihre Fähigkeiten in der Beobachtung und Beurteilung, dem Feedback und als Rollenspieler. Die Dauer ist abhängig von dem Trainingsdesign, dem Leiter, den Beobachtern und dem Konzept des AC. Das Gremium verweist explizit darauf, dass kein Zusammenhang zwischen Qualität und Dauer besteht. Nach Abschluss des Trainings sind die Beobachter in ihren erworbenen Fähigkeiten zu evaluieren. Als weiterer Aspekt wird die Zeit zwischen Training und Einsatz in einem AC genannt. Ist dieser Zeitraum zu groß, sollte eine Nachschulung gefordert werden, um die Aktualität zu wahren (ITFACG, 2009).

*Verhaltensaufnahme.* Das gezeigte Verhalten der Teilnehmer muss von den Beobachtern im Moment der Darbietung aufgezeichnet werden. Nützlich sind Checklisten oder Skalen, aber auch handschriftliche Notizen ermöglichen eine zeitgleiche Dokumentation. Die Verhaltenssequenzen können auch mittels technischer Hilfsmittel wie Video für eine Analyse zu späteren Zeitpunkten festgehalten werden (ITFACG, 2009).

*Datenintegration.* Die Erstellung des Gesamturteils für einen Kandidaten kann entweder durch statistische Aggregation der einzelnen Beurteilungen oder durch eine Beobachterkonferenz erfolgen. Wird eine Diskussion abgehalten, sollten die Assessoren keine irrelevanten Informationen nennen (ITFACG, 2009).

### **3.1.1.2 Grundsatzklärung**

Das AC sollte wenn möglich in das Human Resource System eingegliedert werden. Aus diesem Grund empfiehlt die ITFACG (2009) zunächst eine Grundsatzklärung zu erstellen. Neben der Zielsetzung sind Angaben zu den Teilnehmern, deren Auswahl und Beurteilung, zu den Beobachtermerkmalen und deren Qualifikation und zur Eignung der Entwickler und Berater darin zu finden. Darüber hinaus enthält die Grundsatzklärung Informationen zur Validierung, Datennutzung und zum rechtlichen Rahmen (ITFACG, 2009).

### **3.1.1.3 Informiertheit und Rechte der Teilnehmer**

Vor Durchführung des Verfahrens werden den Teilnehmern die Ziele des AC genannt. Ebenso können die Dimensionen und Aufgaben bekannt gegeben werden. Daneben gehören Auswahlkriterien, Materialien, Ergebnisnutzung und Feedbackangaben sowie der Zugang zu den erhobenen Daten und Kontaktpersonen zur Informiertheit. Angaben über die Beobachter, wie Schulung, Rolle und Zusammensetzung und die Möglichkeit der nochmaligen Teilnahme

sollten außerdem eröffnet werden. Darüber hinaus werden die Punkte Entwicklungsmechanismen, Abgleichung der Ergebnisse mit Organisationsaspekten wie Kultur oder Strategie und Wahlmöglichkeiten genannt (ITFACG, 2009).

Bezüglich der erhobenen Daten haben die Kandidaten verschiedene Rechte. Dazu gehören das Erhalten eines Feedbacks inklusive des Ergebniszustandekommens und die daraus abgeleiteten Empfehlungen. Ist der Teilnehmer ein Unternehmensmitarbeiter hat er sogar ein Recht auf eine formale Zusammenfassung. Auch Informationen über die Datenspeicherung und –nutzung, sowie die Verwendung der Resultate zu anderen als den ursprünglichen Zwecken sind den Kandidaten darzulegen (ITFACG, 2009).

#### **3.1.1.4 Evaluation**

Jedes AC sollte validiert werden, denn Ergebnisse aus Studien über ähnliche Verfahren geben keine Garantie, dass auch diese Konstruktion valide Daten liefert. Grundlage sind die Ableitung der Verhaltensdimensionen sowie die Beziehung zwischen Übung und Dimension. Die Reliabilitätsprüfung im Rahmen eines Pretest ist ebenso wichtig (ITFACG, 2009).

#### **3.1.1.5 Interkulturelle Anwendung**

Das AC kann für die Anwendung in einem Land entwickelt sein oder auch für die Durchführung über geographische Grenzen hinweg. Wichtig dabei ist die Berücksichtigung von Kontextfaktoren wie Gemeinsamkeiten in den Lebensumständen der Länder, Unterschiede in nationalen Richtlinien, Gemeinsamkeiten der arbeitsbezogenen und erfolgsrelevanten Fähigkeiten, die Vergleichbarkeit mit der Organisation, für die das AC entwickelt wurde, der Grad der (De-)Zentralisierung, die Übernahme von Mitarbeitern und ob statistische Vergleiche für die interkulturelle Ergebnisinterpretation notwendig sind. Es gibt dabei im AC kulturspezifische und allgemeingültige Aspekte. So orientieren sich die Auswahl der berufsrelevanten Kriterien, Beobachtungsdimensionen und Übungen ebenso wie das Beobachtertraining und der Feedbackprozess an der jeweiligen Kultur. Dagegen sind die Beobachtung, Klassifikation und Beurteilung des Verhaltens, das Beobachtertraining, die Integration der Bewertungen essentielle Bestandteile, die sich kulturübergreifend in dem Prozess finden (ITFACG, 2009).

#### **3.1.1.6 Abgrenzung des Assessment Center**

Die ITFACG (2009) gibt auch Beispiele welche Aktivitäten nicht als AC zu bezeichnen sind. Zum einen werden explizit Verhaltensbeurteilungen genannt, die nicht auf einer

tatsächlichen Simulation beruhen. Diese können zwar Bestandteil eines AC sein, müssen dann aber mit verhaltenssimulierenden Verfahren kombiniert werden. Der alleinige Gebrauch schließt die Definition als AC aus. Weiterhin werden Panelinterviews, Paper-Pencil-Tests, Ein-Assessor-Bewertungen oder die Anwendung lediglich einer Methode als Ausschlusskriterien genannt (ITFACG, 2009).

### **3.1.2 Standards des Arbeitskreis Assessment Center e.V.**

Ein Bündnis von Experten aus der Dienstleistungs- und Unternehmensbranche hat in den 70er Jahren des vergangenen Jahrhunderts den AkAC gegründet. Etwa 50 Personen arbeiten an der Verbesserung von Methoden der Personalauswahl und -entwicklung. Erstmals wurden die Richtlinien im Jahre 1992 verabschiedet. Mittlerweile liegen sie in der zweiten überarbeiteten Version von 2004 vor (AkAC, 2004).

Es werden insgesamt neun Standards thematisiert, welche eng an die Prozesse der Entwicklung, Durchführung und Nachbereitung des AC angelehnt sind. Neben dem Nutzen und den Umsetzungshinweisen werden auch Verstöße genannt. Dies stellt ein Alleinstellungsmerkmal dieser Empfehlungen dar (AkAC, 2004).

*Standard 1: Auftragsklärung und Vernetzung.* Zunächst müssen Rahmenbedingungen und Ziele sowie die Konsequenzen für die Teilnehmer ermittelt werden. Erst wenn der Nutzen verbalisiert wird, kann abgesehen werden, ob dieses Verfahren der richtige Weg ist und ob auf Widerstände bei Stelleninhabern und Bewerber gestoßen wird. Im Mittelpunkt steht die Frage nach dem Grund für die Einführung der Methode, deren Implementierung in bestehende Strukturen und die daraus entstehenden Konsequenzen. Als Verstöße gegen diesen Standard werden Intransparenz und falsche Vermarktung genannt (AkAC, 2004).

*Standard 2: Arbeits- und Anforderungsanalyse.* Eine Arbeits- und Anforderungsanalyse bildet die Basis für ein Anforderungsprofil, das mit Verhaltensbeispielen untermauert wird. Es ist wichtig, dass sich die Analyse auf die zu besetzende Stelle bezieht. Ein Verstoß besteht in dem Verzicht auf eine Arbeits- und Anforderungsanalyse (AkAC, 2004).

*Standard 3: Übungs-konstruktion.* Das AC besteht aus Arbeitssimulationen. Die Aufgabenauswahl bildet möglichst nah die spätere Tätigkeit ab und bezieht sich nur auf die Anforderungen aus dem Profil. Jede Anforderung soll mindestens zweimal beobachtet werden können. Insgesamt empfiehlt sich der Einsatz von mindestens drei verschiedenen Simulationsübungen. Aufgaben ohne Simulationscharakter sollen nur verwendet werden, wenn eine Anforderung anders nicht ausreichend erfasst werden kann. Alle so konstruierten Übungen müssen auf ihre Tauglichkeit hin überprüft werden (AkAC, 2004).

*Standard 4: Beobachtung und Bewertung.* Das Verhalten der Teilnehmer wird in jeder Aufgabe von mindestens zwei Beobachtern systematisch beobachtet und protokolliert. Hierfür empfiehlt sich der Einsatz eines Beobachtungssystems, das die relevanten Dimensionen pro Übung enthält. Dabei sollten es nicht mehr als fünf, möglichst drei Anforderungen pro Aufgabe sein. Die Beobachter lassen sich nicht von der Darbietung anderer Kandidaten in ihrem Urteil beeinflussen. Die Einzelurteile werden nach dem AC zu einem Gesamtergebnis integriert. Für alle Teilnehmer wird die gleiche Methode angewandt (AkAC, 2004).

*Standard 5: Beobachterausswahl und –vorbereitung.* Die Beobachter durchlaufen eine Schulung, sodass sie Kenntnisse über Ziele, Anforderungen, Übungen, Beobachtung und Beurteilung, Fehlerquellen, Feedback und Beobachterkonferenz gewinnen. Bei der Auswahl ist auf Repräsentativität des Unternehmens sowie Höherstellung zu achten. Ein Nachtraining erfolgt, wenn ein Beobachter längere Zeit nicht an einem AC teilnahm, die Zielgruppe wechselt oder grundlegende Änderungen in dem Verfahren auftreten (AkAC, 2004).

*Standard 6: Vorauswahl und Vorbereitung der potenziellen Teilnehmer.* Um die Akzeptanz des AC zu gewährleisten, müssen im Vorhinein die Auswahlkriterien für die Teilnahme daran sowie des Ablaufs und der Ziele bekannt sein. Es kann auch hilfreich sein den Kandidaten Informationen über die unterschiedlichen Übungen zu geben, sodass Neulinge keine Nachteile gegenüber AC-erprobten Personen haben. Schon für die Vorauswahl empfiehlt es sich standardisierte Verfahren anzuwenden (AkAC, 2004).

*Standard 7: Vorbereitung und Durchführung.* Neben dem Veranstaltungsort gehört auch die Erstellung eines Ablaufplans zu der Vorbereitung. Die Reihenfolge der Aufgaben ist soweit möglich für alle Teilnehmer gleich, sodass es nicht zu Urteilsverzerrungen auf Grund von Reihenfolgeeffekten kommt. Vor jeder Aufgabe sind die Kandidaten über die Kriterien zu informieren. Ein Moderator übernimmt die Organisation und Ankündigungen am Veranstaltungstag (AkAC, 2004).

*Standard 8: Feedback und Folgemaßnahmen.* Nach dem AC erhält jeder Kandidat ein individuelles, auf den Verhaltensbeobachtungen beruhendes Feedback, sodass er seine Stärken und Schwächen erkennt und daran arbeiten kann. Daran anknüpfende Entscheidungen werden mitgeteilt. Grundlage bilden Vertraulichkeit und Datenschutz (AkAC, 2004).

*Standard 9: Evaluation.* Damit die Ziele des AC auch tatsächlich erreicht werden, bedarf es einer regelmäßigen Überprüfung der Qualität. Probleme bei der Durchführung, fehlerhafte Entscheidungen oder Benachteiligung bestimmter Personengruppen können so aufgedeckt und beseitigt werden. Besonders wichtig ist die Erhebung der internen Struktur, der Vorhersagevalidität sowie der Akzeptanz und Fairness (AkAC, 2004).

### 3.1.3 Privatstandards am Beispiel Kanning

Einen Versuch als Einzelperson auf Grundlage empirischer Befunde Richtlinien abzuleiten hat Kanning (2004) unternommen. Seine insgesamt 35 Standards ordnet er den drei Bereichen Entwicklung, Durchführung und Auswertung zu (Kanning, 2004). Die hohe Anzahl ergibt sich daraus, dass er jeden Unterpunkt einzeln benennt ohne übergeordnete Kategorien zu bilden. Aus diesem Grund werden die jeweiligen Standards nicht explizit einzeln genannt, sondern für jede Phase wird eine kurze Zusammenfassung erstellt.

Der Entwicklungsprozess ist gekennzeichnet durch insgesamt acht Empfehlungen. Jedes AC wird für den spezifischen Anwendungsfall konstruiert. Dazu gehören die Anforderungsanalyse, Aufgabenauswahl und Bewertungskriterien. Alle Aufgaben werden vorab getestet und müssen beobachtbare Verhaltensdimensionen enthalten. Besonders bei selbstentworfenen Tests sind die Gütekriterien Objektivität, Reliabilität und Validität zu beachten. Die Beobachtungsunterlagen durchlaufen ebenso einen Pretest. Jede Dimension muss ausreichend abgedeckt sein. Für die Beurteilung werden einheitliche und für alle Personen gleich geltende Kriterien festgelegt, die am besten auf verhaltensverankerten Ratingskalen oder Checklisten eingeschätzt werden. Er empfiehlt den Einsatz einer Anforderungs-Methoden-Matrix (Kanning, 2004).

Die Phase der Durchführung folgt 22 Standards. Ein Moderator und entsprechend geschultes Personal kommen zum Einsatz. Es soll darauf geachtet werden, dass Teilnehmer auf Grund von Gruppenzusammensetzung, Vorwissen, Austausch untereinander oder aufgabenspezifischer Hinweise keine Vor- oder Nachteile haben. Zentral ist die Informiertheit der Kandidaten bezüglich Ablauf und Beurteilungskriterien. Die geschulten Beobachter beurteilen maximal zwei Personen auf maximal drei Dimensionen während einer Aufgabe. Gleichzeitig muss jede Dimension in mindestens zwei Übungen beobachtbar sein sowie jeder Teilnehmer muss von mindestens zwei Beobachtern pro Aufgabe beurteilt werden. Die Beobachter legen ein neutrales Verhalten an den Tag, dabei ist der Kontakt auf die Übungen zu beschränken. Wichtig ist sowohl die Rotation der Beobachter, damit ein Kandidat von möglichst vielen Personen bewertet wird, als auch der Teilnehmer bei der Reihenfolge während einer Einzelübung und bei der Gruppenzusammensetzung (Kanning, 2004).

Die letzten vier Empfehlungen ordnet Kanning (2004) dem Auswertungsprozess zu. Hierzu gehört neben der Beobachterkonferenz auch die Evaluation mit möglichen Verbesserungsvorschlägen. Außerdem erhält jeder Kandidat ein Feedback, in dem die Beurteilung mit konkreten Beispielen untermauert wird (Kanning, 2004).

### 3.2 Vergleich der Standards

Auf den ersten Blick klingen die drei ausgewählten Empfehlungsgruppen sehr ähnlich, doch bei genauerer Analyse lassen sich einige Unterschiede herauslesen. In Tabelle 1 wurden die Richtlinien von dem AkAC (2004) und von Kanning (2004) mit den internationalen Standards der ITFACG (2009) verglichen. Dabei wurde überprüft, ob der Aspekt in einer ähnlichen Form an beliebiger Stelle enthalten ist, da ein 1:1 Vergleich auf Grund des unterschiedlichen Aufbaus nicht möglich ist. Es wurden die Richtlinien der Task Force als Vergleichsmaßstab gewählt, da diese sich durch eine internationale Zusammenarbeit ausweisen, welche ihr eine höhere wissenschaftliche und praktische Bedeutsamkeit beschert. Zu jedem Unterpunkt werden relevante wissenschaftliche Befunde referiert, die die Wichtigkeit der Empfehlung untermauern soll.

Tabelle 1

#### *Vergleich der deutschen Richtlinien mit den internationalen Empfehlungen*

International Task Force on Assessment Center Guideline (2009)	Arbeitskreis Assessment Center e.V (2004)	Kanning (2004)
Anforderungsanalyse	o	o
Verhaltensklassifikation	✓	✓
Beurteilungstechniken	✓	✓
Multiple Beurteilungen	✓	✓
Simulationen	✓	✓
Beobachter	o	o
Beobachtertraining	✓	o
Verhaltensaufnahme	✓	✓
Datenintegration	o	o
Grundsatzklärung	✓	x
Informiertheit und Rechte der Teilnehmer	✓	✓
Evaluation	✓	✓
Technologieeinsatz	x	x
Interkulturalität	x	x
Abgrenzung	o	x

*Anmerkungen.* ✓: eindeutig enthalten, o: teilweise enthalten, x: nicht enthalten

Die Analyse der Anforderungen gilt in allen drei referierten Standards als essentiell zur Ableitung der später zu beobachtenden Verhaltensdimensionen, dabei darf der Bezug zur besetzenden Stelle laut den gemeinschaftlich erarbeitenden Richtlinien nicht fehlen, wobei in den Privatstandards dieser Aspekt fehlt (AkAC, 2004; ITFACG, 2009, Kanning, 2004).

Während die ITFACG (2009) verschiedene Methoden zur Informationssammlung nennt, geht der AkAC (2004) immerhin auf Rahmenbedingungen ein. Kanning (2004) subsumiert lediglich die Frage der relevanten Merkmalsaspekte unter der Anforderungsanalyse, die auch zur Übungskonstruktion herangezogen werden soll. Gerade die Anforderungsanalyse und die Umsetzung der Anforderungen in die Übungen sind für die Inhaltsvalidität sehr wichtig (Fisseni & Fennekels, 1995, Höft, 2007). Wie soll die am besten geeignete Person für die Position ausgewählt werden, wenn die Anforderungen der zu besetzenden Stelle unbekannt sind? Ein AC ohne Analyse der notwendigen Kompetenzen entbehrt sich einer fundierten Entscheidung (Schuler, 2006). Sie bildet eine Grundlage für ordnungsgemäße Entscheidungen, sodass auch die Kandidaten wissen, dass in dem darauf aufbauenden AC keine willkürlichen, eventuell sogar benachteiligten Elemente erfasst werden, was auch entscheidend für die Augenscheinvalidität ist (Sarges, 2001). Um dies zu garantieren ist deshalb eine komplexe Analyse durch Kombination verschiedener Methoden vorzuziehen (Schippmann, Hughes & Prien, 1987).

Die Möglichkeit der Kompetenzmodellierung findet sich nur in den internationalen Empfehlungen (ITFACG, 2009). Es werden dabei nicht konkrete Anforderungen, sondern ganze Kompetenzen wie Sozialkompetenz aufgestellt (Eck, Jöri & Vogt 2010; Marcus, 2011). Diese Begriffe sind oft sehr weit gefasst und zudem diffus (Lievens & Thornton, 2005), können aber in einer schnelllebigen Gesellschaft besser sein, da zukünftig relevante Kompetenzen im Fokus stehen (Rodriguez, Patel, Bright, Gregory & Gowing, 2002).

Alle drei Standards fordern die Operationalisierung der Anforderungen in beobachtbares Verhalten und deren Zuteilung zu verschiedenen Kategorien (AkAC, 2004; ITFACG, 2009, Kanning, 2004). Jede Kategorie sollte präzise Beschreibungen über das hierunter subsumierte Verhalten aufweisen. Sind die Dimensionen zu vage formuliert, können die Beobachter diese unterschiedlich interpretieren und es besteht die Gefahr, dass die Assessoren die Teilnehmer auf Grund ihres bisherigen Eindrucks beurteilen (Kleinmann, 2003).

Zur Darstellung empfehlen alle drei Richtlinien eine Anforderungs-Übungs-Matrix (AkAC, 2004; ITFACG, 2009, Kanning, 2004). Ziel ist es festzulegen, in welcher Übung welche Verhaltensdimensionen im Vordergrund stehen. So weiß der Beobachter, auf welche Aspekte er sich zu konzentrieren hat. Studien belegen, dass sich eine zu hohe Anzahl von zu beobachteten Dimensionen während einer Übung negativ auf die Konstruktvalidität auswirkt (Gaugler & Thornton, 1989; Woehr & Arthur, 2003). Meriac, Hoffman, Woehr und Fleisher (2008) zeigten, dass die beobachteten Dimensionen in einem AC einen hohen Anteil der Varianz neben der Persönlichkeit und kognitiven Fähigkeiten erklären. In einer Metaanalyse

ermittelten Arthur, Day, McNelly und Edens (2003), dass die in AC beobachteten Dimensionen sich zu sieben übergeordneten Anforderungen zusammenfassen lassen: Kommunikationsfähigkeit, Problemlösen, Organisation und Planung, Rücksichtnahme auf andere, Einflussnahme auf andere, Antriebsstärke und Belastbarkeit. Sackett und Dreher (1982) fanden allerdings, dass Assessoren eher übungsspezifische statt dimensionsspezifische Urteile bilden, was sich negativ auf die Konstruktvalidität auswirkt.

Übereinstimmend findet sich in den drei Manuals die Empfehlung zum Einsatz mehrerer Beurteilungsmethoden, allerdings lehnt der deutsche Arbeitskreis nicht simulierte Verfahren ab, außer eine Anforderung ist auf anderem Wege nicht ausreichend erfassbar (AkAC, 2004; ITFACG, 2009, Kanning, 2004). Der Einsatz verschiedener Verfahren ist eine der Voraussetzungen der ITFACG (2009) um von einem AC zu sprechen. Es können jegliche Methoden eingesetzt werden, die dazu dienen die relevanten Informationen für die spätere Entscheidung zu sammeln. Je größer die Aufgabenvariabilität ist, desto höher ist auch die prognostische Validität, da mehr Aspekte abgedeckt werden (Becker, Höft, Holzenkamp & Spinath, 2011; Gaugler et al., 1987). Außerdem zeigte sich, dass sich die inkrementelle Validität nur dann erhöht, wenn auch andere Verfahren neben Simulationen eingesetzt werden (Schuler, 2008). Wichtig ist die Unterscheidung zwischen typischer und maximaler Leistung. Die Arbeitssimulation misst die maximale Leistung, wohingegen Persönlichkeitstests die typische Leistung erfassen. Zusätzlich können weitere Beurteilungsquellen eingesetzt werden, um ein umfassenderes Bild der Kandidaten zu erhalten. So konnte Clapham (1998) zeigen, dass die Beobachterbeurteilung und die Selbsteinschätzung nur gering miteinander korrelieren ( $r = .18$ ). Gaugler et al. (1987) ermittelten, dass Peerbeurteilungen die Validität erhöhen.

Laut der ITFACG (2009) sind Simulationen Übungen, die eine Darstellung der arbeitsbezogenen Verhaltensweisen ermöglichen. Die Anzahl ist bei der ITFACG (2009) von der Komplexität der Arbeitsstelle abhängig, wohingegen der AkAC (2004) eine Mindestanzahl von drei vorgibt. Kanning (2004) gibt den Hinweis auf eine ausreichende Abdeckung der Anforderungen in den Aufgaben. Bei der Anzahl der Dimensionen pro Übung weist die ITFACG (2009) darauf hin, eine Aufgabe nicht zu komplex zu gestalten, da der Beobachter sonst in Probleme geraten kann. Der AkAC (2004) empfiehlt hier eine Maximalzahl von fünf, besser drei. Kanning (2004) nennt bei Gruppenübungen ein Maximum von drei bei maximal zwei Kandidaten. Studien belegen, dass zu viele Aspekte zu fehlerhaften Beurteilungen führen (Gaugler & Thornton, 1989; Lievens & Conway, 2001). Allerdings empfiehlt es sich die Anzahl immer im Kontext festzulegen. Es gibt Inhalte die leichter zu beobachten sind und auch die Übung spielt eine Rolle (Obermann, 2013).

Die Heterogenität der Beobachter findet sich so nur bei der ITFACG (2009). Der AkAC (2004) spricht immerhin von Repräsentativität für das Unternehmen, aber geht lediglich auf berufsbezogene und nicht auf weitere demografische Merkmale ein. Laut Kanning (2004) werden die Beobachter nach ihren Fähigkeiten und Neigungen ausgewählt. Gerade für die Akzeptanz aus Bewerbersicht sind die Beobachter wichtig (Obermann, 2013). So könnte ein afroamerikanischer Abgelehnter diese Absage besonders in den USA vor Gericht damit begründen, dass der Beobachterpool lediglich aus weißen Personen bestand. Untersuchungen belegen, dass die ethnische Zugehörigkeit zu unterschiedlichen Bewertungen führt (Bernardin, Konopaske & Hagan, 2012; Dean, Roth & Bobko, 2008). In verschiedenen Studien konnte nachgewiesen werden, dass der Beobachter den Zusammenhang zwischen der Performance und der Bewertung moderiert, was sich auf die Validität auswirkt (Damitz, Manzey, Kleinmann & Severin, 2003; Gaugler et al., 1987). So werden die Dimensionen besser gemessen, wenn Psychologen als Assessoren fungieren (Lievens & Conway, 2001).

Das Teilnehmer-Beobachter-Verhältnis wird nicht wie bei der ITFACG (2009) anhand der vorliegenden Rahmenbedingungen bestimmt, sondern der AkAC (2004) und Kanning (2004) verweisen auf mindestens zwei Beobachter je Kandidat pro Verhaltensübung. Ein Verhältnis von 2:1 kann angemessen sein, allerdings ist es bei sehr kommunikativen und interaktiven Übungen wie dem Rollenspiel möglich, dass zwei Beobachter zu wenig erfassen können (Obermann, 2013). Der AkAC (2004) und Kanning (2004) definieren die Wichtigkeit der Beobachterrotation, die von der ITFACG (2009) nicht genannt wird.

Das Beobachtertraining ist ebenfalls in allen drei Richtlinien enthalten (AkAC, 2004; ITFACG, 2009; Kanning, 2004). Während Kanning (2004) keine Details zu den Schulungsinhalten erwähnt, finden sich in den anderen beiden Manuals sehr spezifische Angaben (AkAC, 2004; ITFACG, 2009). Verschiedene Studien konnten zeigen, dass sich ein Training positiv auf die Reliabilität, Akkuratheit und Validität der Beurteilung auswirkt (Schleicher, Mayes, Day & Riggio, 2002; Woehr, Arthur & Meriac, 2007).

Sowohl die ITFACG (2009) als auch der AkAC (2004) und Kanning (2004) empfehlen eine unmittelbare Verhaltensaufnahme während bzw. direkt nach der Übung. Werden die Beobachtungen nicht schriftlich festgehalten, kann die spätere Erinnerung an das Beobachtete durch Verzerrungen überlagert sein (Spinath & Becker, 2011). Kolk, Born, van der Flier und Olman (2002) untersuchten den kognitiven Aufwand der Assessoren bei der gleichzeitigen und nach der Übung stattfindenden Niederschrift. Es zeigten sich keine signifikanten Unterschiede. Eine Studie belegte, dass sich verschiedene Methoden wie handschriftliche Notizen oder Checklisten in ihrer Einschätzung kaum unterscheiden, wobei bei freien

Formulierungen die Beurteiler-Übereinstimmung eine größere Variabilität aufweist (Hennessy, Mabey & Warr, 1998).

Die ITFACG (2009) sieht für die Integration der Einzelbeurteilungen eine statistische Aggregation oder eine Beobachterkonferenz vor. Die anderen beiden berufen sich lediglich auf eine Konferenz (AkAC, 2004; Kanning, 2004). Dabei eignet sich jede Methode für jeweils andere Zwecke besser. So ist die Aggregation für die Personalselektion und die Beobachterkonferenz durch ein spezifischeres Feedback für die Personalentwicklung besser geeignet (Höft & Funke, 2006).

Vergleichbar zur Grundsatzklärung bei der ITFACG (2009) thematisiert der AkAC (2004) die Festlegung von Rahmenbedingungen und Zielsetzungen. Kanning (2004) spricht diesen Punkt nicht an. Delmestri und Walgenbach (2009) kommen auf Basis einer Untersuchung zu dem Schluss, dass die Institutionalisierung des AC als gute Human Resource Management Praxis die Einführung des Verfahrens in dem Unternehmen stärkt.

Übereinstimmend bei allen findet sich die Wichtigkeit der Informiertheit und Rechte der Teilnehmer (AkAC, 2004; ITFACG, 2009; Kanning, 2004). Viele Studien haben sich mit dem Thema befasst, ob sich zum Beispiel die Kenntnis der Anforderungen negativ auf die Darstellung auswirkt. Zusammenfassend lässt sich festhalten, dass gerade diese Transparenz das Verfahren zu einem der beliebtesten Methoden bei den Bewerbern macht. Sie wissen, warum sie die einzelnen Aufgaben zu meistern haben und welchen Bezug diese zu der späteren Tätigkeit haben (Sarges, 2001). Untersuchungen zeigten, dass Vorbereitungstipps oder Informationen zu den Übungen unerwünschten Effekten wie Unsicherheit entgegen wirken und so die Konsistenz im gezeigten Verhalten erhöhen (Kolk, Born & van der Flier, 2003; Marggraf-Micheel, Höft & Bonnist, 2006).

Die Evaluation kommt überall zur Ansprache (AkAC, 2004; ITFACG, 2009; Kanning, 2004). Die Überprüfung und Einhaltung von wissenschaftlichen Standards und Gütekriterien ist wichtig um sich an das definierte Anforderungsprofil möglichst nahe anzunähern. Es macht keinen Sinn diesen Aufwand zu betreiben, wenn beispielsweise die Beobachter nach Aussehen und nicht nach gezeigtem Verhalten bewerten. Ein AC ist sehr teuer und zeitraubend, sodass man die Durchführung anhand guter Gütekriterien, worunter auch die Akzeptanz durch die Teilnehmer fällt, rechtfertigen kann (Obermann, 2013). Kersting (2010) konnte in einer Untersuchung in Deutschland zeigen, dass vor allem die Messqualität und die Augenscheinvalidität eine wichtige Rolle für das Akzeptanzurteil spielen. Mit Hilfe der Evaluation können auch Fehler und Verbesserungen ermittelt werden, um die Methode noch effektiver zu gestalten. Gerade die Kriteriumsvalidität, insbesondere in der Personalauswahl

die prognostische Validität, wird als sehr wichtig erachtet. Jedoch erreicht die Validität in einer neueren Metaanalyse mit  $\rho = .26$  nicht die höheren Werte aus früheren Untersuchungen mit  $\rho = .37$  (Gaugler, Rosenthal, Thornton & Bentson, 1987; Hardison & Sackett, 2007; Schmitt & Hunter, 1998).

Der adäquate Einsatz von Technologien und Medien wie Computer oder Video findet sich lediglich in den internationalen Standards (ITFACG, 2004), dabei verweisen Thornton und Gibbons (2009) darauf, dass durch den Einsatz eines Computers im AC nicht nur die Zeit, sondern auch die Kosten reduziert werden können. So sind zwar zunächst die Anschaffungskosten hoch, doch wenn man beispielsweise web-basierte Übungen von zu Hause durchführen kann, können Personen unabhängig von ihrem Aufenthaltsort teilnehmen ohne enorme Reisekosten zu verursachen (Thornton & Gibbons, 2009).

Aspekte der interkulturellen Anwendung sind ausschließlich in den Standards der ITFACG (2009) enthalten. Krause und Thornton (2007) ermitteln in einer Befragung, dass 22% der untersuchten AC-Maßnahmen kulturabhängig sind. Lanik und Gibbons (2011) setzten sich mit dem Beobachtertraining in multi-kulturellen AC auseinander. Sie kommen zu dem Schluss, dass die Schulung kulturspezifische Inhalte aufweisen muss, sodass ein Assessor eine Person nicht auf Grund ihrer kulturellen Besonderheiten anders bewertet. Ehret und Stumpf (2006) weisen darauf hin, dass man bei dem AC immer kulturelle Aspekte beachten muss wie beispielsweise Kollektivismus und Individualismus bei den Einzel- und Gruppenübungen, um für alle Teilnehmer die Fairness zu wahren. Schöning (1996) zeigte, dass es kulturelle Unterschiede in der Kommunikation gibt, die bei der Übungskonstruktion beachtet werden sollten.

Der Abgrenzung des AC von anderen Verfahren ist bei der ITFACG (2009) ein Unterabschnitt gewidmet. Ähnliche Punkte finden sich beim AkAC (2004) durch die Nennung von Verstößen der Standards, wobei diese auch spezifische Fehler gegen ein bestimmtes Kriterium enthalten und sich somit nicht bloß auf die Verfahrensabgrenzung beziehen. Bei Kanning (2004) sucht man nach solchen Inhalten vergeblich. Bei der Entwicklung oder Überarbeitung eines AC kann dies jedoch sehr hilfreich sein. Gegenüber der Geschäftsleitung kann der Konstrukteur auf die Inhalte dieser Richtlinien verweisen und somit die eigene Festlegung und die Ablehnung anderer Maßnahmen untermauern.

Zusammenfassend lässt sich feststellen, dass in der Grundstruktur eines AC zwischen allen erwähnten Standards ein gewisser Konsens besteht. Die Abweichungen in den drei Manuals betreffen allerdings auch sehr wichtige Aspekte wie Ergebnisse aus empirischen Untersuchungen zeigen. So ermittelte der Vergleich einige Inhalte wie den

Technologieeinsatz oder die Interkulturalität, die sich nur in dem internationalen Manual der ITFACG (2009) finden. Insgesamt weisen die Richtlinien des AkAC (2004) und von Kanning (2004) durch ihren Aufbau den Charakter einer Schritt-für-Schritt-Anleitung auf, wohingegen die ITFACG (2009) durch ein umfassenderes Verständnis gekennzeichnet ist.

#### **4. Umsetzung in der Praxis**

In den anschließenden Unterkapiteln werden empirische Befunde zur Umsetzung der Guidelines (ITFACG, 2009) referiert. Wie der Vergleich in Kapitel 3 zeigt, sind diese von den vorgestellten Standards die umfangreichsten. Natürlich sind die Ergebnisse auf die anderen Empfehlungen übertragbar, jedoch würde eine Orientierung an Kanning (2004) zu einem Informationsverlust führen. Die Studien beziehen sich nicht nur auf die Anwendung des AC zur Personalauswahl, sondern auch auf die Personalentwicklung, da die Daten mit Ausnahme weniger Studien nicht getrennt für verschiedene Ziele erfasst wurden. Für eine bessere Vergleichbarkeit werden in der Regel die Gesamtergebnisse genannt, außer der Wert ergibt sich durch sehr unterschiedliche Einzelresultate, also wenn hoch signifikante Unterschiede mit  $p < .001$  vorliegen. In manchen Fällen werden getrennte Werte berichtet, weil die Autoren kein Gesamtergebnis ermittelten oder auf Besonderheiten eingehen. An einigen Stellen sind Tabellen eingefügt. Darin sind der Übersichtlichkeit halber die Studien mit den größten Datenmengen enthalten. Die Angaben aus weiteren Studien werden dann ausschließlich im Text referiert.

##### **4.1 Überblick der Studien**

Zunächst erfolgt in Tabelle 2 ein Überblick über die verwendeten Studien. Neben den Ländern wird auch die Stichprobengröße angegeben. Dies erleichtert den Vergleich der in den folgenden Unterkapiteln referierten Ergebnisse und der Übersicht halber werden diese Daten nicht bei jeder Nennung wiederholt. Zu beachten ist, dass sich einzelne Kennwerte auf eine geringere Anzahl von Befragten beziehen können, was aber aus Platzgründen nicht einzeln aufgeführt werden kann. Alle Ergebnisse beruhen auf standardisierten Interviews per Telefon oder Online-Fragebögen. Es wird auf weitere Studien Bezug genommen, die nicht in Tabelle 2 zu finden sind. Grund hierfür ist, dass keine eigene Datenerhebung erfolgte, sondern die Ergebnisse anderer Untersuchungen einem Vergleich unterzogen wurden.

Tabelle 2

*Überblick der verwendeten Studien zur Umsetzung des Assessment Center in der Praxis*

Studie	<i>n</i>	Land
Hughes, Riley, Shalfrooshan, Gibbons & Thornton (2012)	543	international (53 Nationen)
Povah (2011)	412	international (43 Nationen)
Kudisch, Avis, Thibodeaux & Fallon (2001) <sup>a</sup>	115	international
Thornton & Krause (2009)	144	Amerika; Europa; Asien (18 Nationen)
Krause & Thornton (2009)	97	Nordamerika n=52; Westeuropa n=45
Spychalski, Quiñones, Gaugler & Pohley (1997)	215	USA
Eurich, Krause, Cigularov & Thornton (2009)	54	USA
Lowry (1996)	105	USA
Krause, Meyer zu Kniendorf & Gebert (2001)	141	Deutschland, Österreich, Schweiz
Höft & Obermann (2008)	200	Deutschland, Österreich, Schweiz
Obermann, Höft & Becker (2012) <sup>b</sup>	125	Deutschland, Österreich, Schweiz
Kanning, Pöttker & Gelléri (2007)	97	Deutschland
Boyle, Fullerton & Wood (1995)	75	Großbritannien
Lievens und Goemaere (1999)	23	Belgien
Krause, Rossberger, Dowdeswell, Venter & Joubert (2011)	43	Südafrika

*Anmerkungen.* *n* bezieht sich auf Organisationen, die zum Erhebungszeitpunkt das Assessment Center einsetzten, nicht auf die Anzahl der ursprünglich befragten Unternehmen.

<sup>a</sup> zitiert nach Thornton und Krause (2009)

<sup>b</sup> zitiert nach Obermann (2013)

## 4.2. Anforderungsanalyse

Die Anforderungsanalyse bildet die Basis für alle weiteren Schritte. Die Übersicht in Tabelle 3 offenbart, dass sich dieser Schritt auch in den Unternehmen etabliert hat. Durchschnittlich in 90% der Unternehmen wird zuvor eine Anforderungsanalyse durchgeführt, wobei die Werte in internationalen und amerikanischen Untersuchungen höher sind als in Südafrika (Eurich, Krause, Cigularov & Thornton, 2009; Hughes, Riley, Shalfrooshan, Gibbons & Thornton, 2012; Krause, Rossberger, Dowdeswell, Venter und Joubert, 2011; Krause & Thornton, 2009; Spsychalski, Quiñones, Gaugler & Pohley, 1997; Thornton & Krause, 2009). Eine niedrigere Zahl mit 64% findet sich bei Höft und Obermann (2008) für den deutschsprachigen Raum, wobei Krause, Meyer zu Kniendorf & Gebert (2001) für die gleiche Region einen Anteil von 88% und Kanning, Pöttker und Gelléri (2007) sogar von 97% nur für Deutschland benennen. Der Unterschied ergibt sich durch die explizite Frage nach empirischen Anforderungsanalysen in der Studie von Höft und Obermann (2008). Lowry (1996) nennt mit 80% einen niedrigeren Wert für die USA. Jedoch beschränkt sich

seine Studie auf öffentliche Institutionen. Untersuchungen zu den Ergebnissen von Spychalski et al. (1997) zeigen, dass sich die Angabe zu der Anforderungsanalyse weder signifikant von dem Wert von Eurich et al. (2009),  $\chi^2(1, N = 269) = 0.33, p > .05$ , noch von dem Resultat von einer Teilstichprobe ( $n = 75$ ) von Krause et al. (2001),  $\chi^2(1, N = 290) = 0.44; p > .001$ , unterscheidet (Eurich et al., 2009; Krause & Gebert, 2003a). Povah (2011) stellte in einer internationalen Untersuchung fest, dass nur bei 55% der Befragten eine geeignete Methode der Anforderungsanalyse durchgeführt wird.

Tabelle 3

*Anwendung der Anforderungsanalyse und deren verschiedenen Methoden in Prozentangaben an der Stichprobengröße n*

	Hughes et al. (2012)	Thornton & Krause (2009)	Krause & Thornton (2009)		Spychalski et al. (1997)	Eurich et al. (2009)	Krause et al. (2011)
Land	int.	int.	WE	NA	USA	USA	Südafrika
<i>n</i>	419	144	45	52	215	54	43
Anforderungsanalyse	97	93	89	95	93	90	84
Stellenbeschreibung	54	69	41	76	82	70	47
Stelleninhaber:							
Interview	30	62	39	56	79	54	23
Fragebogen	13	21	14	41	61	40	16
Beobachtung	15	15	7	31	43	30	5
Supervisor:							
Interview	40	45	59	54	80	47	26
Fragebogen	12	16	16	29	62	26	16
CIT	23	24	36	27	45	28	2
Kompetenzmodelle:							44
exisierend	52	44	39	20	—	16	—
neu	32	26	43	20	—	—	—
Workshop	—	19	36	15	—	14	9

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa, NA: Nordamerika; CIT: Critical Incident Technique (Flanagan, 1954)

Wie die Ergebnisse in Tabelle 3 darlegen, verwenden viele Organisationen für die Ableitung der Anforderungen verschiedene Verfahren. Stellenbeschreibungen finden sich in mehr als 70% der amerikanischen Unternehmen (Eurich et al., 2009; Krause & Thornton, 2009; Spychalski et al., 1997) und liegen somit über den Ergebnissen aus internationalen Studien (Hughes et al., 2012; Thornton & Krause, 2009). Jedoch berichtet Povah (2011) mit

89% einen viel höheren Wert auf internationaler Basis, wobei er anmerkt, dass länderübergreifend die Stellenbeschreibung die höchsten Werte erzielt. In Südafrika, Westeuropa und dem deutschsprachigen Gebiet wird die Stellenbeschreibung in rund der Hälfte der Unternehmen eingesetzt (Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009). Ein Vergleich von Spychalski et al. (1997) und Krause et al. (2001) ermittelt einen signifikanten Unterschied,  $\chi^2(1, N = 290) = 30.17, p < .001$  (Krause & Gebert, 2003a), wohingegen Eurich et al. (2009) keinen bedeutsamen Unterschied in den beiden US-amerikanischen Studien fanden,  $\chi^2(1, N = 269) = 3.50, p > .05$ .

Den Stelleninhaber als Datenquelle verwenden viele Unternehmen. Die beliebtesten Methoden sind Interviews, gefolgt von Fragebögen und Beobachtungen (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). Die Daten schwanken dabei sehr von Studie zu Studie. Für das Interview werden Werte zwischen einem Viertel und mehr als drei Viertel der Unternehmen gefunden (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Povah, 2011; Spychalski et al., 1997; Thornton & Krause, 2009). Es lassen sich selbst für Studien aus denselben geografischen Regionen keine einheitlichen Befunde referieren. Für den Fragebogen und die Beobachtung sind die Angaben ebenso sehr verschieden in den unterschiedlichen Studien, wobei die höchsten Werte für Nordamerika berichtet werden. Zusammenfassend finden sich Anteile zwischen 13% und 61% für den Fragebogen und zwischen 5% und 43% für die Beobachtung (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). Eine Testung zeigt, dass sich die Ergebnisse der Studien aus den USA signifikant für Interviews,  $\chi^2(1, N = 269) = 14.43, p < .001$  und Fragebögen,  $\chi^2(1, N = 269) = 7.17, p < .01$ , für die Beobachtung allerdings nicht,  $\chi^2(1, N = 269) = 3.11, p > .05$ , unterscheiden (Eurich et al., 2009). Für Deutschland liegen keine ausführlichen Daten vor. Höft und Obermann (2008) geben an, dass etwa jedes zweite Unternehmen den Stelleninhaber als Datenquelle verwendet und bei Krause et al. (2001) liegt dieser Anteil bei lediglich einem Drittel.

Für den Supervisor ergeben sich in den deutschen Studien Angaben um die 70% (Höft & Obermann, 2008; Krause et al., 2001). In etwa der Hälfte der Unternehmen wird ein Interview mit einem Supervisor durchgeführt (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009; Thornton & Krause, 2009). Lediglich Spychalski et al. (1997) berichten mit 80% einen höheren und Krause et al. (2011) mit 26% einen geringeren Anteil. Nach Zielen getrennt betrachtet ergibt sich bei Thornton und Krause (2009) für die Personalauswahl ein Wert von 67% und für die Personalentwicklung ein Resultat von 30%. Für den Fragebogen

finden sich studienübergreifend Werte um die 15% (Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009) mit Ausnahme der nordamerikanischen Stichproben mit einem Viertel (Eurich et al., 2009; Krause & Thornton, 2009) und einem größeren Anteil von 62% bei Spychalski et al. (1997). Signifikante Unterschiede für die USA bei der Verwendung von Interviews,  $\chi^2(1, N = 269) = 25.01, p < .001$  und Fragebögen,  $\chi^2(1, N = 269) = 22.49, p < .001$  mit einem Supervisor wurden ermittelt (Eurich et al., 2009).

Die Critical Incident Technique (Flanagan, 1954) kommt bei einem Viertel der befragten Unternehmen zum Einsatz (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009; Thornton & Krause, 2009). Ein etwas höherer Anteil von etwa einem Drittel finden Krause und Thornton (2009) in Westeuropa und Povah (2011) weltweit. Eher selten wird das Verfahren in Südafrika eingesetzt (Krause et al., 2011). Wieder nennen Spychalski et al. (1997) mit 45% den höchsten Wert aller Studien, welcher sich signifikant von neueren Ergebnissen unterscheidet,  $\chi^2(1, N = 269) = 5.34, p < .05$  (Eurich et al., 2009).

Eine weitere beliebte Methode zur Ableitung der Anforderung sind Kompetenzmodelle. In Südafrika und Deutschland verwendet etwa jedes zweite Unternehmen diese Variante (Höft & Obermann, 2008; Krause et al., 2011). In anderen Studien wurde nochmals zwischen bestehenden und neu entwickelten Kompetenzmodellen unterschieden. Dabei werden in internationalen und europäischen Studien mit mehr als einem Drittel bis zu der Hälfte für existierende und mehr als einem Viertel bis ein Drittel für neu entwickelte Kompetenzmodelle höhere Werte gefunden als in den nordamerikanischen Stichproben mit maximal 20% für beide Varianten (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009; Thornton & Krause, 2009). Getrennt nach Zielen ergibt sich bei bestehenden Kompetenzmodellen für die Selektion ein Anteil von nur 20%, wohingegen die Personalentwicklung einen Wert von 66% erreicht (Thornton & Krause, 2009).

Workshops kommen besonders in den europäischen Unternehmen zur Ableitung der Anforderungen zum Einsatz. Während die Werte für andere Regionen und auch in internationalen Studien unter 20% liegen, sind es für Westeuropa ein Drittel der Unternehmen und für den deutschsprachigen Raum sogar etwa die Hälfte der Organisationen (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Thornton & Krause, 2009).

Höft und Obermann (2008) untersuchten Unterschiede in den Methoden der Anforderungsanalyse in den Studien von 2001 und 2008. Die Befragung von Stelleninhabern stellt das einzig signifikante Ergebnis,  $\chi^2(1, N = 249) = 13.18, p < .001$ , in den verschiedenen Methoden dar.

### 4.3 Verhaltensklassifikation

Weltweit werden in den meisten AC zwischen sechs und zehn Dimensionen erfasst (Eurich et al., 2009; Hughes et al., 2012; Kanning et al., 2007; Krause, 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Dabei unterscheiden sich die Ergebnisse für den deutschsprachigen Raum von 2008 ( $M$  11,49,  $SD$  4,03) und 2001 ( $M$  8,47,  $SD$  3,03, Krause et al., 2001) signifikant voneinander,  $t(247) = 6,90$ ,  $p < .001$  (Höft & Obermann, 2008). Pro Übung werden laut Hughes et al. (2012) in 78% der Unternehmen maximal 5 Dimensionen beobachtet, was sich mit anderen Studien deckt (Höft & Obermann, 2008; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). In Süd- und Mittelamerika werden in 30% aller Unternehmen mehr als 7 Kompetenzen beobachtet, was ein substantiell höherer Wert als in anderen Region mit Werten zwischen 3% und 16% ist (Hughes et al., 2012).

Laut Kudisch, Avis, Thibodeaux and Fallon (2001, zitiert nach Thornton & Krause, 2009) sind die populärsten Dimensionen, die von Arthur et al. (2003) ermittelten Anforderungen mit Ausnahme der Belastbarkeit. Die Ergebnisse von Hughes et al. (2012) sind mit anderen Studien nur schwer vergleichbar, da einige Kompetenzen weiter gefasst sind. Deshalb werden diese Resultate nur bei ähnlichem Verständnis genannt.

Mit Ausnahme von Südafrika wird die Kommunikationsfähigkeit in über 90% der Unternehmen erfasst (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Rücksichtnahme auf andere ermittelt jedes zweite nordamerikanische Unternehmen (Eurich et al., 2009; Krause & Thornton, 2009) und ein Drittel der Organisationen in Südafrika (Krause, 2011). Die Werte von internationalen Studien liegen mit 77% weltweit und 80% für Westeuropa höher (Krause & Thornton, 2009; Thornton & Krause, 2009). Während zwei Drittel der Unternehmen weltweit und in Westeuropa die Antriebsstärke ihrer Kandidaten erfasst, sind es in Südafrika und Nordamerika nur ein Drittel (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Bei der Einflussnahme sind die Werte für Südafrika und Nordamerika mit etwa zwei Dritteln ebenso etwas geringer als in internationalen Untersuchungen mit etwa 75% (Eurich et al., 2009, Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Thornton und Krause (2009) ermitteln in 65% der Auswahlprogramme und 93% der Entwicklungscenter diese Dimension. Für das Organisieren und Planen finden sich etwas höhere Zahlen mit etwa 80% für Nordamerika als in anderen Regionen mit etwa zwei Dritteln, wobei der Anteil international bei 60% liegt (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause). Hughes et al. (2012) stellten fest, dass im kontinentalen Vergleich der Unterschied in dieser Dimension das

einzig signifikante Ergebnis ist. In Ozeanien mit 39% und Europa mit 45% wird die Dimension Organisieren und Planen erheblich seltener erfasst als auf anderen Kontinenten mit Werten über 60%. Allerdings gehen in diese Berechnung nur Unternehmen ein, die angaben die Dimension in einem substantiellen Ausmaß zu beobachten (Hughes et al., 2012). Für unterschiedliche Ziele unterscheidet sich die Erhebung mit 80% für die Selektion und 42% für die Entwicklung (Thornton & Krause, 2009). Das Problemlösen wird in mehr als zwei Drittel der Unternehmen erfasst, wobei die Daten für Nordamerika und Westeuropa höher liegen (Eurich et al., 2009, Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Ebenfalls unterscheiden sich die Werte in Abhängigkeit von der Zielsetzung mit 88% für die Auswahl und 51% für die Entwicklung (Thornton & Krause, 2009).

Im deutschsprachigen Raum liegen die Werte ähnlich denen aus internationalen Studien (Höft & Obermann, 2008, Krause et al., 2001). Ansonsten werden andere Dimensionen wie Durchsetzung, Kooperation, Konfliktfähigkeit, Führungskompetenz, Zielorientierung, Engagement und Entscheidungsfreude mit Werten bis zu 90% genannt, wobei sich keine signifikanten Unterschiede in den beiden Studien ergeben (Höft & Obermann, 2008; Krause et al., 2001). In der aktuellsten Studie von 2012 sind die Kommunikationsfähigkeit, Durchsetzungskraft und Analysefähigkeit die beliebtesten Methoden (Obermann, Höft & Becker, 2012; zitiert nach Obermann, 2013). Ebenso Povah (2011) erfragte noch weitere Dimensionen wie Führungskompetenz und Teamarbeit, die weltweit die am häufigsten erfassten Dimensionen sind. Handelt es sich um ein AC zur Selektion erfassen 83% die Teamfähigkeit, 79% die Führungskompetenz und 73% Organisieren und Planen. Dagegen werden im Falle der Entwicklung bei 80% die Führungskompetenz, bei 74% die Teamfähigkeit und bei 69% das Problemlösen erhoben (Povah, 2011).

Boyle, Fullerton und Wood (1995) ermitteln, dass diese Globaldimensionen ihre Tücken haben, denn für die einzelnen Übungen werden nur bei 41% spezifische Beispiele gegeben. Eurich et al. (2009) stellten fest, dass der Bezug der Dimensionen zum Job nur in 63% der Fälle sehr gut ist. Ebenso finden Krause et al. (2001) heraus, dass 43% der Unternehmen, welche eine Anforderungsanalyse machen, ihre Dimensionen auch daraus ableiten. Lievens und Goemaere (1999) erhalten in einer belgischen Stichprobe, dass ein Drittel der befragten Firmen die Dimensionen aus einer Anforderungsanalyse ableiten. Die anderen übernehmen Listen von externen Beratern oder lassen sie von der Unternehmensleitung erstellen. So kritisieren einige Befragte die Operationalisierung in beobachtbares Verhalten, denn es finden sich auf den Beobachtungslisten Dimensionen die so nicht messbar sind (Lievens & Goemaere, 1999).

#### 4.4 Multiple Beurteilungen insbesondere durch Simulationen

Mehr als die Hälfte der Firmen lassen die Teilnehmer vier bis fünf Übungen absolvieren (Eurich et al., 2009; Höft & Obermann, 2008, Kanning et al., 2007; Krause et al., 2001, 2011; Krause & Thornton, 2009; Lowry, 1996; Obermann et al., 2012; zitiert nach Obermann, 2013; Thornton & Krause, 2009). In der Studie von Hughes et al. (2012) ergibt sich mit 25% für drei, 27% für vier und 15% für fünf Simulationen eine geringfügig kleinere Anzahl. Krause und Gebert (2003b) untersuchten die Ergebnisse von Krause et al. (2001) im Hinblick auf die Unterscheidung zwischen internen ( $n = 82$ ) und externen ( $n = 55$ ) AC. Interne AC beziehen sich auf die Personalauswahl und –entwicklung von Organisationsmitgliedern, während das externe AC auf die Selektion neuer Personen gerichtet ist. Sie entdecken mit Hilfe eines Mann-Whitney-Test, dass in einem internen AC signifikant mehr Aufgaben eingesetzt werden als in einem externen AC,  $U = 1500.00$ ,  $p < .001$  (Krause & Gebert, 2003b).

Laut Thornton und Krause (2009) werden in 37% der Unternehmen drei Viertel der Simulationen für die Organisation entwickelt, wobei es für die Selektion 57% und für die Entwicklung 21% sind. In einer anderen Studie finden sie für Westeuropa einen Wert von 49% und für Nordamerika von 58% (Krause & Thornton, 2009). Krause et al. (2001) beklagen in ihrer deutschsprachigen Stichprobe die mangelnde Äquivalenz zwischen Übung und Anforderung, da nur die Hälfte der Unternehmen mehr als 50% der Übungen eigens für das AC konstruieren. Ebenfalls zeigt sich, dass mehr Übungen neu entwickelt werden, wenn es sich um ein internes AC handelt,  $U = 1412.50$ ,  $p < .001$  (Krause & Gebert, 2003b). In einer neueren Studie aus den gleichen Ländern erhalten Obermann et al. (2012, zitiert nach Obermann, 2013), dass 56% der Übungen spezifisch für die Position entwickelt werden.

In Tabelle 4 sind die Prozentzahlen der verschiedenen Simulationen angegeben. Es lässt sich kein eindeutiges Muster erkennen, welche Übungen am häufigsten zum Einsatz kommen. Povah (2011) ermittelt, dass weltweit das Rollenspiel und die Präsentation die beliebtesten Methoden sind, was sich allerdings nicht mit allen Resultaten in Tabelle 4 deckt. Auffallend ist der doppelt so hohe Wert für die Gruppendiskussion bei Spychalski et al. (1997) als in anderen nordamerikanischen Studien (Eurich et al., 2009; Krause & Thornton, 2009). In internationalen Studien schwanken die Werte von 50% (Hughes et al., 2012; Povah, 2011) bis zu 78%, mit 62% für die Selektion und 91% für die Entwicklung (Thornton & Krause, 2009). Krause und Thornton (2009) ermitteln für Westeuropa 90%, was sich mit Studien aus dem deutschsprachigen Raum deckt (Kanning et al., 2007; Krause et al., 2001), wobei ein geringerer Anteil von Höft und Obermann (2008) mit 61% genannt wird. In Südafrika kommt die Gruppendiskussion mit einem Drittel am seltensten zum Einsatz (Krause et al., 2011).

Tabelle 4

*Arten von verwendeten Simulationsaufgaben in Prozentangaben an der Stichprobengröße n*

	Hughes et al. (2012)	Thornton & Krause (2009)	Krause & Thornton (2009)		Spychalski et al. (1997)	Eurich et al. (2009)	Krause et al. (2011)	Höft & Obermann (2008)	Kanning et al. (2007)
Land	int.	int.	WE	NA	USA	USA	ZA	D, A, CH	D
<i>n</i>	419	144	45	52	215	54	43	200	97
Gruppen- diskussion	51	78	90	45	100	43	37	61	87
Postkorb	42	59	35	55	82	57	54	45	47
Rollenspiel	70	61	88	78	54	76	49	73	76
Präsentation	59	79	92	58	46	64	51	89	76
Fakten erkennen	14	44	20	33	38	29	16	17	—
Fallstudie	72	65	78	38	—	—	26	70	11
Planspiel	17	37	40	48	—	48	—	—	37
Pretest	—	61	61	43	65	41	50	—	—

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa; NA: Nordamerika; ZA: Südafrika; D: Deutschland; A: Österreich; CH: Schweiz

In rund der Hälfte der Organisationen wird der Postkorb verwendet, mit etwas niedrigeren Angaben für Westeuropa mit rund einem Drittel (Eurich et al., 2009; Höft & Obermann, 2008; Hughes et al., 2012; Kanning et al., 2007; Krause et al., 2001, 2011; Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009). Wieder berichten Spychalski et al. (1997) mit 82% einen höheren Wert.

Das Rollenspiel kommt bei Spychalski et al. (1997) und Krause et al. (2011) in etwa jedem zweiten Unternehmen zum Einsatz, während andere Studien auf zwei Drittel bis drei Viertel kommen, wobei 88% der westeuropäischen Befragten auf das Rollenspiel setzen (Eurich et al., 2009; Hughes et al., 2012, Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009;). Hierbei finden sich Unterschiede für verschiedene Ziele mit 80% für die Auswahl und 44% für die Entwicklung (Thornton & Krause, 2009). Besonders auffallend sind die Ergebnisse aus dem deutschsprachigen Raum. Höft und Obermann (2008) und Kanning et al. (2007) erhalten einen Wert von über 70%, wobei Krause et al. (2001) das Rollenspiel nur bei 10% aller Unternehmen findet. Grund hierfür ist die Unterteilung in Rollenspiel und Zweiergespräch, welches in 82% der Fälle eingesetzt wird (Krause et al., 2001).

Für die Präsentation werden die höchsten Werte in Europa ausgemacht. Während in anderen Ländern etwa die Hälfte dieses Verfahren nutzen und internationale Studien diese

Ergebnisse belegen (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Povah, 2011; Spychalski et al., 1997), sind es im deutschsprachigen Gebiet mehr als drei Viertel aller Unternehmen (Höft & Obermann, 2008; Kanning et al., 2007; Krause et al., 2001). Lediglich Thornton und Krause (2009) finden mit 79% einen höheren Anteil in einer weltweiten Untersuchung. Insgesamt scheint laut Krause und Thornton (2009) in Europa diese Methode beliebter zu sein.

Fallstudien kommen in etwa zwei Dritteln aller Unternehmen weltweit zur Anwendung (Hughes et al., 2012, Povah, 2011; Thornton & Krause, 2009), mit höheren Angaben für Westeuropa mit drei Viertel und geringeren Werten für Nordamerika mit etwas mehr als einem Drittel (Krause und Thornton, 2009) und Südafrika mit einem Viertel (Krause et al., 2011). Die deutschen Studien berichten ähnliche Ergebnisse wie die internationalen Studien (Höft & Obermann, 2008; Krause et al., 2001) mit Ausnahme von Kanning et al. (2007), die lediglich einen Anteil von 11% erhalten.

Hughes et al. (2012) finden für das Fakten erkennen international mit 14% einen ähnlich niedrigen Wert wie Krause et al. (2011) für Südafrika, Höft und Obermann (2008) für Deutschland und Krause und Thornton (2009) für Westeuropa. In Nordamerika kommt diese Methode bei etwa einem Drittel der Unternehmen zum Einsatz (Eurich et al., 2009; Krause & Thornton, 2009; Spychalski et al., 1997). Den höchsten Wert ermitteln Thornton und Krause (2009) in einer internationalen Studie mit 44%. Im Gegensatz dazu nennt Povah (2011) den Einsatz in jedem dritten Unternehmen.

Das Planspiel erreicht Angaben zwischen 32% und 48% sowohl in nationalen als auch in internationalen Studien (Eurich et al., 2009; Kanning et al., 2007; Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009). Lediglich Hughes et al. (2012) berichten mit 17% einen geringeren Anteil.

Rund die Hälfte der Unternehmen führt einen Pretest mit den Übungen durch (Eurich et al., 2009; Krause et al., 2011; Spychalski et al., 1997; Thornton & Krause, 2009). Eurich et al. (2009) errechnen dabei wiederum signifikante Unterschiede für die beiden Studien aus den USA,  $\chi^2(1, N = 269) = 10.27, p < .01$  (Eurich et al., 2009).

Neben Simulationen kommen in den meisten AC noch weitere Methoden wie psychologische Tests und Interviews zum Einsatz. Höft und Obermann (2008) ermitteln für Deutschland, dass etwa ein Drittel der Unternehmen neben Simulationen auch Tests und Interviews verwendet, wohingegen sich etwa ein Viertel auf Simulationen beschränkt. Thornton und Krause (2009) kommen zu dem Resultat, dass 82% der Unternehmen neben Übungen auf weitere Verfahren zurückgreifen.

Tabelle 5

*Einsatz von weiteren Verfahren ohne Simulationscharakter in Prozent als Anteil an der Stichprobengröße n*

	Thornton & Krause (2009)	Krause & Thornton (2009)		Eurich et al. (2009)	Krause et al. (2011)
Land	int.	WE	NA	USA	Südafrika
<i>n</i>	144	45	52	54	43
Situationales Interview	38	48	48	48	23
Background Interview	34	48	40	35	16
Wissenstest	18	15	30	31	5
Fertigkeitstest	22	20	30	29	49
Persönlichkeitstest	45	25	20	26	54
Intelligenztest	16	20	10	10	7
Biographische Daten	14	23	10	17	14
Soziometrische Instrumente	6	5	5	7	2
Arbeitsprobe	11	10	15	17	7
Selbsteinschätzung	28	50	8	—	29
Peer-Rating	10	18	5	—	18

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa; NA: Nordamerika

Neben dem häufigen Einsatz von verschiedenen Arten des Interviews werden wie in Tabelle 5 dargelegt besonders Persönlichkeits-, Intelligenz-, Fertigkeitstests und Wissenstests eingesetzt. Arbeitsproben, soziometrische Instrumente und biographische Fragen kommen darüber hinaus auch noch zum Einsatz (Eurich et al., 2009; Höft & Obermann, 2008; Hughes et al., 2012; Kanning et al., 2007; Krause et al., 2001, 2011; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). Dabei werden Persönlichkeitstests signifikant häufiger eingesetzt, wenn das Ziel Entwicklung (63%) statt Auswahl (25%) ist (Thornton & Krause, 2009). Povah (2011) findet ebenso Unterschiede in Abhängigkeit von der Zielsetzung. So kommen in der Personalauswahl am häufigsten Interviews mit 84% und in der Personalentwicklung Persönlichkeitsfragebögen mit 78% zum Einsatz (Povah, 2011).

Krause und Gebert (2003a) ermitteln in ihrer Studie, dass sich die Werte aus den USA von 1997 und dem deutschsprachigen Raum von 2001 in allen eingesetzten Verfahren mit Ausnahme der Gruppendiskussion signifikant unterscheiden. Ebenso erhalten Eurich et al. (2009) signifikante Ergebnisse für die Daten aus den USA mit Ausnahme der Methode des Fakten Findens. Im deutschsprachigen Raum finden sich dagegen nur bei der Gruppendiskussion signifikante Unterschiede (Höft & Obermann, 2008). Obermann et al. (2012, zitiert nach Obermann, 2013) benennen als beliebteste Verfahren für den

deutschsprachigen Raum die Präsentation, das Rollenspiel und das Interview, wobei sie auch einen zunehmenden Testeinsatz mit einem Anstieg von 19% für 2001, auf 31% für 2008 und mehr als 40% für 2012 ausmachen.

Weitere Datenquellen sind die Selbst- und Peereinschätzungen. Die Selbsteinschätzung kommt in Westeuropa in jedem zweiten Unternehmen zum Einsatz (Krause & Thornton, 2009). Studien aus dem deutschsprachigen Raum bestärken diese Angabe (Höft & Obermann, 2008; Krause et al., 2001; Obermann et al., 2012; zitiert nach Obermann, 2013). Dieser Wert liegt damit über den internationalen Ergebnissen mit etwa einem Drittel von Thornton und Krause (2009). Für Südafrika (Krause et al., 2011) und die USA (Spychalski et al., 1997) werden ähnliche Zahlen angegeben. In einer anderen Studie finden Krause und Thornton (2009) für Nordamerika sogar nur einen Wert von 8%. Krause und Gebert (2003a) berechnen signifikante Unterschiede zwischen den Ergebnissen aus dem deutschsprachigen Raum und der USA für die Selbsteinschätzung,  $\chi^2(1, N = 290) = 14.62, p < .001$ , jedoch nicht bei der Beurteilung durch einen Peer,  $\chi^2(1, N = 290) = .34, p > .001$ . Insgesamt kommt die Einschätzung durch einen Peer seltener zum Einsatz. Nicht einmal ein Viertel der Unternehmen wendet diese Methode an (Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Spychalski et al., 1997; Thornton & Krause, 2009).

## 4.5 Beobachter

Länderübergreifend fungieren hauptsächlich drei Personengruppen als Beobachter in einem AC. Wie aus Tabelle 6 ersichtlich sind dies Vorgesetzte, Human Resource Experten und Psychologen. Bei Lievens und Goemaere (1999) werden in der Hälfte aller AC nur Vorgesetzte eingesetzt und zu jeweils einem Viertel nur externe Berater oder eine Mischung aus beiden Personengruppen.

Etwa in jedem zweiten Unternehmen treten Vorgesetzte als Beobachter auf (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009; Povah, 2011; Spychalski et al., 1997; Thornton & Krause, 2009). In Westeuropa und im deutschsprachigen Raum fungieren sie sogar fast in jedem AC, wohingegen in Südafrika nur 16% der Firmen einem Vorgesetzten diese Rolle überlassen (Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013). Ein Vergleich ermittelt signifikant unterschiedliche Werte für die USA und die deutschsprachigen Länder,  $\chi^2(1, N = 290) = 52.01, p < .001$  (Krause & Gebert, 2003a), dagegen unterscheiden sich die Daten aus den beiden USA-Studien nicht signifikant,  $\chi^2(1, N = 269) = 0.68, p > .05$  (Eurich et al., 2009).

Tabelle 6

*Beruflicher Hintergrund der Beobachter in Prozentangaben an der Stichprobengröße n*

	Hughes et al. (2012)	Thornton & Krause (2009)	Krause & Thornton (2009)		Eurich et al. (2009)	Krause et al. (2011)
Land	int.	int.	WE	NA	USA	Südafrika
<i>n</i>	393	144	45	52	54	43
Vorgesetzte	51	70	93	44	42	16
Interne HR Experten	53	42	75	37	42	23
Externe HR Experten	45	24	20	34	33	9
Interne Psychologen	15	22	43	12	14	28
Externe Psychologen	43	44	32	15	19	42
Supervisor	—	6	7	10	9	4
Betriebsrat	—	2	5	2	2	—

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa; NA: Nordamerika; HR: Human Resource

Mit Ausnahme der Studie von Spychalski et al. (1997) mit 2% finden sich hohe Werte für den Einsatz von Human Resource Experten, was einen signifikanten Unterschied zu den Stichproben aus den USA,  $\chi^2(1, N = 269) = 169.96, p < .001$ , und dem deutschsprachigen Raum darstellt,  $\chi^2(1, N = 290) = 184.01, p < .001$  (Eurich et al., 2009; Krause & Gebert, 2003a). So werden in der Hälfte der Unternehmen interne Personalmitarbeiter eingesetzt, mit etwas geringeren Werten für Nordamerika (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009, Thornton & Krause, 2009). In Südafrika liegt der Anteil bei nur einem Viertel (Krause et al., 2011), während es in Westeuropa und dem deutschsprachigen Raum sogar drei Viertel sind (Höft & Obermann, 2008; Krause et al., 2001; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013). Kanning et al. (2007) berichten als Ausreißer einen Wert von 90% für Deutschland. Die Anzahl externer Human Resource Experten ist durchweg geringer (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Povah (2011) findet insgesamt einen Anteil von zwei Drittel.

Während in den internationalen Studien, ebenso wie in Südafrika, Psychologen in etwa zwei Dritteln aller Unternehmen zum Einsatz kommen, hält in Nordamerika nur in etwa einem Drittel der Organisationen ein Psychologe Einzug ins AC (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Laut den Angaben von Spychalski et al. (1997) setzen sogar nur 6% Psychologen für diese Aufgabe ein. Ein Vergleich ergibt hierbei einen signifikanten Unterschied für die Studien aus

den USA,  $\chi^2(1, N = 269) = 31.52, p < .001$  (Eurich et al., 2009). Für den deutschsprachigen Raum finden sich mit 7% ebenso niedrige Angaben (Krause et al., 2001), was keinen signifikanten Unterschied zu Spychalski et al. (1997) ergibt,  $\chi^2(1, N = 290) = 1.30, p > .001$  (Krause & Gebert, 2003a). Insgesamt schwankt der Anteil von Psychologen weltweit stark, mit 7% in Ozeanien, 18% in Asien, einem Drittel in Europa und Amerika und 55% in Afrika (Povah, 2011). Ebenso für verschiedene Ziele ist der Einsatz von Psychologen sehr unterschiedlich. So agieren bei der Selektion bei 32% interne und bei 19% externe Psychologen, wohingegen bei der Entwicklung bei 13% interne und bei 68% externe Psychologen eingesetzt werden (Thornton & Krause, 2009).

Eher seltener übernimmt ein Supervisor oder der Betriebsrat diese Funktion. Die Kennwerte liegen hier studienübergreifend bei maximal 10% (Eurich et al., 2009; Krause et al., 2001, 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Thornton und Krause (2009) finden zusätzlich in der Hälfte der Unternehmen einen Gleichstellungsbeauftragten. Höft und Obermann (2008) ermitteln bei dem beruflichen Hintergrund der Beobachter keine signifikanten Unterschiede für die Studien von 2001 und 2008.

Lowry (1996) gibt an, dass das Teilnehmer-Beobachter-Verhältnis in Abhängigkeit von der Position des Beobachters bestimmt wird, indem die Anzahl der gleichzeitig zu beobachtenden Personen mit der Organisationsebene steigt. Andere Studien berichten Kennwerte über das Verhältnis ohne die Bedingungen zu erfragen (Boyle et al., 1995; Krause et al., 2011; Povah, 2011).

Tabelle 7

*Prozentangaben der berücksichtigten Merkmale bei der Beobachterausswahl als Anteil an der Stichprobengröße n*

	Thornton & Krause (2009)	Krause & Thornton (2009)		Eurich et al. (2009)	Krause et al. (2011)
Land	int.	WE	NA	USA	Südafrika
<i>n</i>	144	45	52	54	43
Ebene	74	89	54	56	9
Funktion	72	80	54	54	28
Geschlecht	18	20	29	28	9
Rasse	9	—	27	26	9
Ethnische Herkunft	6	—	17	16	7
Alter	6	11	7	5	2
Bildungsstand	17	18	24	4	28

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa; NA: Nordamerika

In Tabelle 7 sind die weiteren Aspekte neben der Position zur Beobachterausswahl dargestellt. Die Organisationsebene und die Funktion spielen dabei die größten Rollen. So wählen über 80% der Unternehmen in Westeuropa und die Hälfte der Organisationen im amerikanischen Raum nach diesen Kriterien aus (Eurich et al., 2009; Krause & Thornton, 2009). Eine internationale Studie von Thornton und Krause (2009) kommt auf einen Anteil von knapp drei Viertel aller Befragten. Lediglich in Südafrika finden sich Unterschiede in den beiden Kriterien. So entscheiden ein Viertel der Unternehmen nach der Funktion, aber nur 9% nach der Ebene (Krause et al., 2011). Das Alter spielt kaum eine Rolle. Nur für Westeuropa finden sich mit 11% zweistellige Angaben (Thornton & Krause, 2009). Das Geschlecht wird in etwa 20% aller Unternehmen berücksichtigt, mit einem geringeren Anteil in Südafrika und höheren Werten für Nordamerika (Eurich et al., 2009; Krause et al., 2001; Krause & Thornton, 2009; Thornton & Krause, 2009). Die ethnische Herkunft und Rasse werden hauptsächlich in Nordamerika bei der Beobachterausswahl berücksichtigt (Eurich et al., 2009; Krause & Thornton, 2009). Der Bildungsstand erreicht Werte um die 20% (Krause & Thornton, 2009; Thornton & Krause, 2009). Eurich et al. (2009) präsentiert mit 4% eine stark geringere Angabe. Für Südafrika wird mit 28% der höchste Wert genannt, wobei in den anderen Punkten die geringsten Werte für dieses Land zu erkennen sind (Krause et al., 2011).

#### **4.6 Beobachtertraining**

Bis zu 10% der befragten Unternehmen schulen ihre Beobachter nicht (Eurich et al., 2009; Höft & Obermann, 2008; Hughes et al.; 2012; Kanning et al., 2007; Krause & Thornton, 2009; Spychalski et al., 1997). Krause et al. (2001) finden für den deutschsprachigen Raum nur einen Trainingsanteil von 82% und in Belgien werden lediglich in jedem zweiten Unternehmen die Beobachter vorher ausgebildet (Lievens & Goemaere, 1999). In Süd- und Mittelamerika schult ein Drittel der Organisationen ihre Beobachter nicht (Hughes et al., 2012). Povah (2011) findet mit Ausnahme von Amerika in drei Viertel der Organisationen eine Beobachterschulung. In Amerika ist es nur die Hälfte der Unternehmen, was er aber auf die geringe Stichprobengröße zurückführt.

Die Trainingsdauer schwankt von einem halben Tag bis hin zu drei Tagen (Hughes et al., 2012). Dabei unterscheidet sich die Dauer bei Spychalski et al. (1997,  $M$  3.84,  $SD$  .74) und Krause et al. (2001,  $M$  1.72,  $SD$  .72) signifikant voneinander,  $t(288) = -2.66$ ,  $p < .001$  (Krause & Gebert, 2003a). Ein Nachtraining wird in zwei Drittel der Firmen in den USA durchgeführt (Spychalski et al., 1997), während Lievens und Goemaere (1999) diese Angabe nur in einem Unternehmen in Belgien finden.

Die Trainingsinhalte reichen von Informationen über die Organisation, Zielposition, Übungen, Dimensionen, Beobachtung und Beurteilung, Feedback oder Datenintegration bis hin zur Demonstration der eigenen Fertigkeiten und sind in den referierten Studien sehr heterogen. Nahezu alle befragten Organisationen vermitteln Kenntnisse über die zu beobachteten Dimensionen (Hughes et al., 2012, Spychalski et al., 1997). Die eingesetzten Übungen nehmen eine wichtige Rolle in mehr als drei Viertel der Unternehmen ein, wobei einige Studien sogar fast jede Organisation benennen (Eurich et al., 2009; Höft & Obermann, 2008; Hughes et al., 2012; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). In Südafrika werden nur in jedem zweiten Unternehmen Übungsinhalte an die Beobachter im Training vermittelt (Krause et al., 2011). Wissen über die Beobachtung und Beurteilung sowie Klassifikation erhalten die Trainingsteilnehmer in etwa 80% der Firmen, wiederum mit einer Ausnahme für Südafrika mit 49% (Hughes et al., 2012, Krause et al., 2011; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). In der Hälfte der Unternehmen werden Kenntnisse über die Organisation und Zielposition vermittelt, wobei der Anteil in Westeuropa bei nur einem Viertel liegt (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009). Thornton & Krause (2009) finden diese Informationen in einem Drittel der Firmen mit Unterschieden von 49% für die Selektion und 16% für die Entwicklung. In mehr als der Hälfte der Institutionen erfolgt die Darstellung der Fähigkeiten im Trainingsrahmen (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009). Dabei kann es sich um das Geben von Feedback, die Beurteilung oder die Darbietung als Rollenspieler handeln. Spychalski et al. (1997) ermitteln in 88% der Organisationen eine Demonstration der eigenen Fertigkeiten.

Eine Evaluation der Assessoren erfolgt in jedem zweiten Unternehmen (Eurich, 2009; Hughes et al., 2012; Spychalski et al., 1997; Thornton & Krause, 2009). Dabei lassen sich regionale Unterschiede erkennen. So liegt der Anteil in Afrika bei etwa 74%, in Asien und Süd- und Mittelamerika bei jeweils 64%, wohingegen in Nordamerika 47%, in Ozeanien 38% und in Europa 34% ihre Beobachter zertifizieren (Hughes et al., 2012). Dies deckt sich mit anderen Studien (Krause et al., 2011; Krause & Thornton, 2009).

#### **4.7 Verhaltensaufnahme**

Zur Verhaltensaufnahme werden verschiedene Verhaltensbeschreibungen, Skalen und Checklisten, aber auch freie Notizen eingesetzt. Handschriftliche Notizen nutzt jede zweite Organisation in Westeuropa, Nordamerika und Südafrika (Krause et al., 2011; Krause & Thornton, 2009). Checklisten kommen länderübergreifend in rund der Hälfte der befragten

Unternehmen zum Einsatz (Eurich et al., 2009; Hughes et al., 2012, Krause et al., 2001, 2011; Krause & Thornton, 2009; Spychalski et al., 1997; Thornton & Krause, 2009). Kanning et al. (2007) berichten einen geringeren Anteil von einem Drittel. Die Angaben unterscheiden sich weder für die US-Studien signifikant,  $\chi^2(1, N = 269) = 0.92, p > .05$ , noch bei Krause und Gebert (2003a),  $\chi^2(1, N = 290) = 5.11, p > .001$ .

In jedem vierten bis fünften Unternehmen werden realistische Verhaltensbeschreibungen eingesetzt (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009). In der internationalen Studie von Hughes et al. (2012) wird ein Wert von 80% berichtet. Im deutschsprachigen Raum ist es mehr als jede zweite Organisation (Krause et al., 2001) und Spychalski et al. (1997) kommen auf fast jede Institution. Krause und Gebert (2003a) errechnen, dass sich diese Werte signifikant unterscheiden,  $\chi^2(1, N = 290) = 40.72, p < .001$ .

Als Skalen werden zum Beispiel Verhaltensbeobachtungsskalen, graphische Ratingskalen oder verhaltensverankernde Einstufungsskalen eingesetzt. Den höchsten Wert für graphische Ratingskalen findet sich in der britischen Stichprobe mit 39% (Boyle et al., 1995), den geringsten in Südafrika mit 9% (Krause et al., 2011). In den anderen Studien wird ein Anteil von 20% genannt (Eurich et al., 2009; Krause & Thornton, 2009). Verhaltensverankernde Einstufungsverfahren nutzen ein Viertel der Unternehmen in den USA (Eurich et al., 2009) und die Hälfte der Firmen in Südafrika (Krause et al., 2011). International sind es ebenso etwa die Hälfte der Organisationen (Hughes et al., 2012). Der Einsatz von Beobachtungsskalen schwankt von einem Viertel für die USA (Spychalski et al., 1997), über zwei Drittel für Großbritannien (Boyle et al., 1995) bis zu drei Viertel für den deutschsprachigen Raum (Kanning et al., 2007; Krause et al., 2001). Ein Chi<sup>2</sup>-Test zeigt signifikante Unterschiede zwischen den USA und den deutschsprachigen Ländern,  $\chi^2(1, N = 290) = 66.71, p < .001$  (Krause & Gebert, 2003a). Neben Boyle et al. (1995) bemerken auch Lievens und Goemaere (1999) und Krause et al. (2001), dass Videos eher selten zur Anwendung kommen.

#### **4.8 Datenintegration**

Die Studien enthalten als Methoden zur Integration der Einzelurteile die statistische Aggregation, die Beobachterkonferenz oder eine Kombination aus beiden Möglichkeiten.

Die Aggregation kommt am seltensten zum Einsatz. In Studien mit Kennwerten zu den drei Varianten finden sich Werte unter 10% (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009) mit Ausnahme bei Kudisch et al. (2001, zitiert nach Thornton & Krause, 2009) mit etwas höheren 15%. Ansonsten erreichen die Angaben bis zu 20%, wenn die Kombination nicht mit einbezogen wurde (Boyle

et al., 1995; Höft & Obermann, 2008; Povah, 2011; Spsychalski et al., 1997), wiederum mit einer Ausnahme bei Krause et al. (2001) mit 7%. Der Unterschied ist weder für die USA,  $\chi^2(1, N = 269) = 1.68, p > .05$  (Eurich et al., 2009), noch bei Krause und Gebert (2003a) signifikant,  $\chi^2(1, N = 290) = 3.99, p > .001$ , aber bei dem deutschsprachigen Vergleich,  $\chi^2(1, N = 207) = 14.40, p < .001$  (Höft & Obermann, 2008).

Durchschnittlich die höchsten Angaben erhält die Beobachterkonferenz, wobei auch diese Werte zwischen 32% für Südafrika (Krause et al., 2011) und 86% für den deutschsprachigen Raum (Krause et al., 2001) schwanken. In internationalen Studien sind die Ergebnisse mit etwa der Hälfte bei Hughes et al. (2012) und Thornton und Krause (2009) abweichend von 65% bei Kudisch et al. (2001, zitiert nach Thornton & Krause, 2009), sowie 75% bei Povah (2001). Dabei unterscheidet sich die Angabe bei Spsychalski et al. mit 84% signifikant von den 45% aus einer neueren Studie,  $\chi^2(1, N = 269) = 37.60, p < .001$  (Eurich et al., 2009), allerdings nicht von den deutschsprachigen Ländern,  $\chi^2(1, N = 290) = 2.86, p > .001$  (Krause & Gebert, 2003a). Für den deutschsprachigen Raum sind die Unterschiede ebenso nicht signifikant,  $\chi^2(1, N = 207) = 4.77, p > .001$  (Höft & Obermann, 2008).

Eine Kombination der beiden Methoden kommt in etwa der Hälfte der Unternehmen zur Anwendung (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009) mit einem geringeren Anteil von einem Fünftel in der internationalen Studie von Kudisch et al. (2001, zitiert nach Thornton & Krause, 2009).

In den Studien aus dem deutschsprachigen Raum und den USA wird als weitere Möglichkeit die Abstimmung mit Mehrheitsentscheid genannt, allerdings mit Werten unter 10%, wobei weder für die USA,  $\chi^2(1, N = 269) = 0.67, p > .05$ , noch für den amerikanischen und deutschen Vergleich,  $\chi^2(1, N = 290) = 0.18, p > .001$ , noch für den deutschsprachigen Raum,  $\chi^2(1, N = 207) = 2.52, p > .001$ , signifikante Unterschiede berechnet werden (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001; Spsychalski et al., 1997).

## 4.9 Grundsatzklärung

Die Grundsatzklärung beinhaltet unter anderem die Aspekte Ziele, Teilnehmerauswahl und beteiligte Personen. Hughes et al. (2012) erhalten mit 62% einen höheren Wert für das Ziel Selektion in Ozeanien als in anderen Kontinenten, besonders Afrika mit 17% und Süd- und Mittelamerika mit 12%. In diesen Ländern werden mit 45% für Afrika und 59% für Süd- und Mittelamerika mehr AC mit dem Ziel der Personalentwicklung veranstaltet. Povah (2011) ermittelt in mehr als jedem zweiten Unternehmen ein AC mit dem Ziel der Selektion oder der Entwicklung. Etwas seltener werden die Potentiale diagnostiziert. In Europa spielt die

Auswahl eine größere Rolle, wohingegen in Nordamerika die meisten AC zum Zwecke der Personalentwicklung durchgeführt werden (Povah, 2011). Eurich et al. (2009) nennen für die USA einen Anteil von 54% für Beförderungen, 27% für die Entwicklung und 19% für die Auswahl. Spychalski et al. (1997) kommen mit Ausnahme der Selektion mit jedem zweiten Unternehmen auf ähnliche Resultate. Ein Chi<sup>2</sup>-Test bestätigt einen hoch signifikanten Unterschied für die Selektion,  $\chi^2(1, N = 269) = 17.63, p < .001$  und leicht signifikante Differenzen für die Beförderung,  $\chi^2(1, N = 269) = 1.02, p < .05$ , sowie für die Entwicklung,  $\chi^2(1, N = 269) = 5.90, p < .05$  (Eurich et al., 2009). Krause und Gebert (2003a) erhalten für die Entwicklung mit 8% signifikante Unterschiede,  $\chi^2(1, N = 290) = 25.08, p < .001$ , und für die Personalauswahl mit 40%,  $\chi^2(1, N = 290) = 2.33, p > .001$  und die Beförderung mit 52%,  $\chi^2(1, N = 290) = 0.79, p > .001$ , keine bedeutsamen Abweichungen zu den Ergebnissen von Spychalski et al. (1997).

Die Teilnehmerauswahl erfolgt auf Empfehlung eines Supervisors, durch eine Selbstnominierung oder durch eine Personalbeurteilung. Hierbei erhält der Supervisor mit 50% die höchsten Werte (Krause et al., 2011; Spychalski et al., 1997; Thornton & Krause, 2009). Die Ergebnisse unterscheiden sich weder für die USA noch für den deutschsprachigen Raum und die USA in einem der drei Möglichkeiten signifikant (Eurich et al., 2009; Krause & Gebert, 2003a). Eine Aufgliederung nach den Zielen des AC zeigt, dass bei der Personalentwicklung 87% auf eine Empfehlung des Supervisors, 17% auf Personalbeurteilungen und 16% auf die Selbstnominierung zur Teilnehmerauswahl zurückgreifen. Ist das Ziel hingegen die Selektion ergibt sich ein anderes Bild. Mit 45% macht der Supervisor zwar noch den größten Anteil aus, allerdings direkt gefolgt von der Selbstnominierung mit 40% und der Personalbeurteilung mit 32% (Thornton & Krause, 2009).

#### **4.10 Informiertheit und Rechte der Teilnehmer**

Ein Vergleich der Studien zeigt, dass die Informiertheit der Teilnehmer sehr unterschiedlich ist. Positiv erwähnen verschiedene Autoren in ihren Studien den geringen Anteil mit unter 10% der Unternehmen, die keine Informationen an ihre Teilnehmer geben (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001; Spychalski et al., 1997).

Länderübergreifend werden in mehr als zwei Drittel der Fälle die Ziele des AC bekannt gegeben (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause). Im deutschsprachigen Raum ist es sogar in fast 90% der Organisationen üblich, wobei es in neueren Studien nur noch etwa 80% der Unternehmen sind (Höft & Obermann, 2008; Krause et al., 2001; Obermann et al., 2012; zitiert nach Obermann, 2013). Spychalski et

al. (1997) präsentieren mit 7% einen wesentlich niedrigeren Anteil. In 20-50% der Unternehmen erhalten die Kandidaten Vorbereitungstipps (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Obermann et al., 2012; zitiert nach Obermann, 2013; Thornton & Krause, 2009), Spychalski et al. (1997) erhalten mit 6% den niedrigsten Wert. Mit Ausnahme von Südafrika geben etwa die Hälfte der Unternehmen Informationen über die Beobachterrolle (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Thornton und Krause, 2009). Wiederum nennen Spychalski et al. (1997) mit 6% einen wesentlich geringeren Wert. Ein bis zwei Drittel der Firmen informieren über die Übungsarten (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Spychalski et al., 1997; Thornton & Krause, 2009). In den europäischen Ländern machen zwischen der Hälfte und zwei Drittel der Organisationen Angaben zu den erhobenen Daten, wohingegen die Werte für Nordamerika zwischen 8% und 55% schwanken und für Südafrika bei einem Fünftel liegen (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2001; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Spychalski et al., 1997). In Deutschland finden sich mit 80% hohe Werte zum Thema Ablauf (Höft & Obermann, 2008; Krause et al., 2001; Obermann et al., 2012; zitiert nach Obermann, 2013). Bei Spychalski et al. (1997) wird das Vorgehen dagegen nur in einem Drittel der Unternehmen offen kommuniziert. Weitere Aspekte wie Anforderungen, beobachtetes Verhalten, Auswahlkriterien, Empfehlungen, Ergebnisse und Feedbackart tauchen nur in wenigen Studien auf. Zusammenfassend lässt sich sagen, dass diese Punkte in jedem zweiten bis dritten Unternehmen offen gelegt werden, wobei das Feedback sogar Werte bis zu 80% für Westeuropa erreicht (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Thornton und Krause (2009) zeigten, dass bei allen Informationen mit Ausnahme der Dimensionen die Werte bei Entwicklungszentren höher sind als bei Auswahlprogrammen. Die Informiertheit unterscheidet sich für den deutschsprachigen Raum in keinem der erhobenen Aspekte (Höft & Obermann, 2008), dagegen sind die Resultate für die USA und den deutschsprachigen Ländern mit Ausnahme der Anforderungen und des beobachteten Verhaltens signifikant unterschiedlich (Krause & Gebert, 2003a). Ebenso sind die Ergebnisse für die USA signifikant unterschiedlich (Eurich et al., 2009).

Wie in Tabelle 8 zu erkennen ist, finden sich in fast allen Studien sehr hohe Werte in der Feedbackrate. Länderübergreifend liegt der Anteil bei über 90% (Höft & Obermann, 2009; Kanning et al., 2007; Krause et al., 2001; Krause & Thornton, 2009; Spychalski et al., 1997;

Thornton & Krause, 2009). Lowry (1996) ermittelt, dass 72 Firmen ein Feedback geben und immerhin 15 darauf verzichten. Krause und Gebert (2003a) errechnen keine signifikanten Unterschiede für den deutschsprachigen Raum und die USA,  $\chi^2(1, N = 290) = 2.49, p > .001$ .

Tabelle 8

*Varianten und Zeitpunkt des Feedbacks in Prozent als Anteil an der Stichprobengröße n*

	Thornton & Krause (2009)	Krause & Thornton (2009)		Eurich et al. (2009)	Krause et al. (2011)	Höft & Obermann (2008)	Krause et al. (2001)
Land	int.	WE	NA	USA	Südafrika	D, A, Ch	D, A, Ch
<i>n</i>	144	45	52	54	43	200	141
Verbal	25	28	32	21	18	34	46
Schriftlich	—	5	19	18	3	2	1
Kombination	61	64	46	54	79	64	50
kein	4	3	3	—	—	1	3
<b>Zeitpunkt</b>							
direkt	37	54	25	22	7	57 <sup>a</sup>	67
< eine Woche	18	15	17	19	36	29 <sup>a</sup>	21
> eine Woche	46	31	58	59	57	15 <sup>a</sup>	12

*Anmerkungen.* —: keine Daten vorhanden; int.: international; WE: Westeuropa; Na: Nordamerika; D: Deutschland; A: Österreich; Ch: Schweiz

<sup>a</sup> bezieht sich auf n=193

Ein Fünftel bis ein Drittel der Firmen bevorzugt ein Gespräch, wobei im deutschsprachigen Raum diese Variante in einer älteren Studie von fast der Hälfte der Unternehmen genannt wurde (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Thornton & Krause, 2009). Spychalski et al. (1997) erhalten mit 71% einen wesentlich höheren Wert, allerdings berichten die Autoren auch mit 61% die höchste Angabe in der schriftlichen Rückmeldung. Ähnliche Werte ermittelt auch Povah (2001) mit 84% für die verbale und 57% für die schriftliche Variante, wobei diese Möglichkeit bei internen Mitarbeitern häufiger verwendet wird. In beiden Studien wurde die Kombination nicht erfragt. Ansonsten erreichen nur die amerikanischen Stichproben einen nennenswerten Anteil von knapp 20% in der schriftlichen Variante (Eurich et al., 2009; Krause & Thornton, 2009). Die Resultate aus den USA unterscheiden sich sowohl beim schriftlichen,  $\chi^2(1, N = 269) = 31.13, p < .001$ , als auch beim mündlichen Feedback,  $\chi^2(1, N = 269) = 46.79, p < .001$  (Eurich et al., 2009). Ebenfalls im Vergleich zu den deutschsprachigen Ländern sind die Werte aus den

USA für das verbale Feedback signifikant unterschiedlich,  $\chi^2(1, N = 290) = 23.19, p < .001$ , allerdings nicht für die schriftliche Rückmeldung,  $\chi^2(1, N = 290) = .30, p > .001$  (Krause & Gebert, 2003a). Vorherrschend findet sich in den Studien die Bevorzugung einer Kombination aus schriftlichem und mündlichem Feedback. Jedes zweite bis dritte Unternehmen greift auf eine Kombination zurück, wobei es in Südafrika sogar 79% der Firmen sind (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton, 2009; Obermann et al., 2012; zitiert nach Obermann, 2013; Thornton & Krause, 2009). In der Feedbackart gibt es wiederum keine signifikanten Unterschiede für die deutschsprachigen Studien (Höft & Obermann, 2008).

Der Zeitpunkt des Feedbacks variiert wiederum von Land zu Land. So hat der Teilnehmer in Westeuropa (Krause & Thornton, 2009), besonders im deutschsprachigen Raum (Höft & Obermann, 2008; Krause et al., 2001), die größte Wahrscheinlichkeit direkt im Anschluss ein Feedback zu erhalten, wobei Obermann et al. (2012, zitiert nach Obermann, 2013) feststellen, dass das Feedback mittlerweile später, eher bis zu einer Woche danach gegeben wird und nur noch 37% der Firmen das Feedback direkt im Anschluss geben. Rund ein Drittel der Unternehmen in Südafrika geben innerhalb einer Woche eine Rückmeldung (Krause et al., 2011), was in allen gesichteten Studien der höchste Wert ist. Im Durchschnitt fallen etwa 20% der Unternehmen in diese Zeitspanne (Eurich et al., 2009; Krause et al., 2001; Krause & Thornton, 2009; Thornton & Krause, 2009). In Nordamerika und Südafrika muss der Kandidat in fast der Hälfte der Organisationen mehr als eine Woche warten, was sich auch mit internationalen Ergebnissen deckt (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). In der aktuellen weltweiten Studie von Hughes et al. (2012) zeigt sich ein ähnliches Muster. So geben nur 18% das Feedback direkt, während 30% sich bis zu sieben Tage Zeit lassen und 39% ihre Kandidaten länger als eine Woche warten lassen. Wiederum finden Höft und Obermann (2008) für keinen der drei Zeitintervalle Unterschiede für die Untersuchungen aus 2001 und 2008.

#### **4.11 Evaluation**

Die Daten aus verschiedenen Studien offenbaren, dass nicht alle Anwender die Brauchbarkeit des AC überprüfen. In Tabelle 9 fallen direkt die durchweg niedrigen Zahlen der Evaluation auf. Durchschnittlich führt lediglich jedes zweite Unternehmen eine Evaluation durch (Krause et al., 2001, Krause & Thornton, 2009; Lowry, 1996; Povah, 2011; Spsychalski, et al., 1997; Thornton & Krause, 2009). Die höheren Angaben von etwa 70% sind ebenso weit vom Optimum entfernt (Eurich et al., 2009; Höft & Obermann, 2008; Krause et

al., 2011; Obermann et al., 2012; zitiert nach Obermann, 2013). Dabei sind die Ergebnisse von 2001 mit 40% und 2008 mit 77% aus dem deutschsprachigen Gebiet signifikant unterschiedlich,  $\chi^2(1, N = 335) = 47,58, p < .001$  (Höft & Obermann, 2008). Hughes et al. (2012) finden in 14% der Unternehmen keine Evaluation vor, wobei Nordamerika mit 6% den geringsten und Asien mit 21% den höchsten Anteil ausweist. Krause et al. (2001) bemängeln, dass einem Viertel der Unternehmen die Begriffe Validität und Reliabilität unbekannt sind.

Tabelle 9

*Prozentangaben der Evaluation und erhobenen Gütekriterien an der Stichprobengröße n*

	Thornton & Krause (2009)	Krause & Thornton (2009)		Spychalski et al. (1997)	Eurich et al. (2009)	Krause et al. (2011)
Land	international	WE	NA	USA	USA	Südafrika
<i>n</i>	144	45	52	215	54	43
Evaluation	34	—	—	66	—	71
systematisch	50	51	46	—	50	—
Objektivität	37	32	52	—	—	75
Reliabilität	66	36	71	52	52	70
Inhaltsvalidität	67	63	79	88	72	65
Konstruktvalidität	52	40	40	20	36	75
Kriteriumsvalidität	—	32	21	—	—	—
konkurrente Validität	—	—	—	—	28	45
prädiktive Validität	—	—	—	36	38	75

*Anmerkungen.* —: keine Daten vorhanden; WE: Westeuropa; NA: Nordamerika

Lediglich drei Studien referieren Kennwerte der Objektivität. Dabei wird die Objektivität in rund einem Drittel der Unternehmen weltweit sowie in Westeuropa erhoben, wobei die Daten für Nordamerika mit der Hälfte und drei Viertel der Organisationen für Südafrika höher sind (Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009).

Die Reliabilität wird in zwei Drittel der Unternehmen weltweit überprüft, mit geringeren Werten für Westeuropa mit einem Drittel und uneinheitlichen Angaben mit 52% bis 71% für Nordamerika (Eurich et al., 2009; Krause et al., 2011; Krause & Thornton, 2009; Spsychalski et al., 1997; Thornton & Krause, 2009). Lievens und Goemaere (1999) sprechen sogar davon, dass in Belgien die Prüfung der Validität und Reliabilität eher die Ausnahme stellt.

Für die Validität findet sich bei Kudisch et al. (2001, zitiert nach Thornton & Krause, 2009) ein Wert von 72%. Wenn man Tabelle 9 betrachtet, lässt sich feststellen, dass besonders die Inhaltsvalidität erfasst wird. Spsychalski et al. (1997) erhalten einen Wert nahe 90%, wohingegen Eurich et al. (2009) auf 72% und Krause und Thornton (2009) auf 79%

kommen. Eine genauere Analyse von Eurich et al. (2009) zeigt signifikante Unterschiede in den Ergebnissen von 1997 und 2009,  $\chi^2(1, N = 269) = 8.22, p < .01$ . In Südafrika und Westeuropa liegen die Angaben mit zwei Dritteln etwas niedriger (Krause et al., 2011; Krause & Thornton, 2009). Diese Zahlen stimmen in etwa mit den Ergebnissen der internationalen Studie von Thornton und Krause (2009) überein. Ein sehr uneinheitliches Bild findet sich wiederum bei der Konstruktvalidität. Krause et al. (2011) berichten von einem Anteil von 75% in Südafrika, wohingegen Lowry (1996) gerade einmal auf 6% der befragten Unternehmen kommt. Insgesamt legt jedes zweite bis dritte Unternehmen weltweit Wert auf die Bestimmung der Konstruktvalidität (Thornton & Krause, 2009). Sychalski et al. (1997) berichten wiederum mit 20% einen abweichenden Wert, was einen signifikanten Unterschied für die USA-Studien ergibt,  $X(1, N = 269) = 100.55, p < .001$  (Eurich et al., 2009). Für die Kriteriumsvalidität werden Gesamtwerte von 32% für Westeuropa und 21% für Nordamerika berichtet (Krause & Thornton, 2009). Lowry (1996) erhält sogar nur einen Anteil von 10%. In den anderen Studien werden getrennte Resultate für die konkurrente und prognostische Validität genannt. Wieder einmal springt die Angabe mit 75% für die Vorhersagevalidität aus Südafrika ins Auge (Krause et al., 2011). Die Werte von Sychalski et al. (1997) und Eurich et al. (2009) sind mit einem Drittel viel niedriger. Hughes et al. (2012) kommen für Afrika auf einen Wert von 21%, der etwas niedriger als für Nordamerika mit 25% ist. In Asien, Ozeanien und Europa werden in etwa 15%, in Süd- und Zentralamerika in 7% der Unternehmen die prognostische Validität untersucht (Hughes et al., 2012). Für die konkurrente Validität berichten Eurich et al. (2009) einen Anteil von einem Viertel und Krause et al. (2011) von fast der Hälfte. Diese Angaben decken sich teilweise mit denen aus einer internationalen Studie von Hughes et al. (2012). Insgesamt liegt der Anteil bei 13%. Nordamerika mit 28%, Süd- und Mittelamerika mit 27% und Afrika mit 24% machen hierbei den größten Anteil aus. In den anderen Kontinenten liegt der Anteil bei etwa 10% (Hughes et al., 2012). Bei Höft und Obermann (2008) lässt sich die Information finden, dass von den 77% Firmen, die eine Evaluation durchführen, 29% die Korrelation zu Berufserfolgskriterien berechnen.

#### **4.12 Technologieeinsatz**

Die neuesten technologischen Errungenschaften lassen sich nach kürzester Zeit kaum noch aus dem alltäglichen Leben weg denken. In zwei internationalen Studien geben rund 60% aller Befragten an, technische Hilfsmittel wie Telefon, Computer oder Videokamera in das AC einzubinden (Hughes et al., 2012, Povah, 2011). Kudisch et al. (2001, zitiert nach Thornton & Krause, 2009) stellen fest, dass der Computer selten zum Einsatz kommt.

Tabelle 10

*Der Einsatz von Computern in einem Assessment Center in Prozentangaben an der Stichprobengröße  $n$*

Land	Thornton & Krause (2009)	Krause & Thornton (2009)		Eurich et al. (2009)
	international	WE	NA	USA
$n$	144	45	52	54
Computereinsatz	43	58	56	61
Berichte erstellen	32	46	31	37
Aufgabenpräsentation	19	22	31	32
Auswertung	21	38	18	22
Verhaltensklassifikation	10	11	10	12
Verhaltenschecklisten	7	14	8	9
Webübungen	8	7	8	10

*Anmerkungen.* WE: Westeuropa; NA: Nordamerika

In Tabelle 10 ist der Einsatz von Computern aufgelistet. Insgesamt wird der Computer in etwa jedem zweiten Unternehmen verwendet (Eurich et al., 2009; Krause & Thornton, 2009; Thornton & Krause, 2009). An erster Stelle des Einsatznutzens steht die Erstellung von Berichten, gefolgt von der Aufgabenpräsentation, Auswertung, Verhaltensklassifikation und Checklisten. Die Werte für Nordamerika sind mit Ausnahme der Aufgabenpräsentation fast identisch mit den Ergebnissen aus einer internationalen Untersuchung, wohingegen in Westeuropa häufiger Berichte erstellt und Auswertungen vorgenommen werden (Eurich et al., 2009; Krause & Thornton, 2009; Thornton & Krause, 2009). In etwa 10% der Unternehmen kommt der Computer zur Bearbeitung von Übungen zum Einsatz (Eurich et al., 2009, Krause & Thornton, 2009; Thornton & Krause, 2009). Lievens und Goemaere (1999) nennen als Beispiel hierfür den PC-Postkorb. Povah (2001) ermittelt hierbei andere Ergebnisse. In rund der Hälfte der Firmen in Europa, Amerika und Ozeanien, in einem Drittel der asiatischen Firmen und fast in jeder afrikanischen Organisation werden Computer-Aufgaben eingesetzt, die den größten Anteil an der Computernutzung stellen (Povah, 2011)

### 4.13 Interkulturalität

Die Mehrheit der Anwender fertigt das AC für die Nutzung in einem Land an. So geben 94% der Konstrukteure an, dass das AC speziell für das Anwendungsland entwickelt wurde (Thornton & Krause, 2009). Hughes et al. (2012) ermitteln, dass 77% der Unternehmen das AC in nur einem Staat anwenden. Der Rest weitet den Einsatz auf mehr als ein Land aus,

wobei 10% das Verfahren in geographisch ähnlichen Regionen und 13% unabhängig von geographischen Gegebenheiten anwenden (Hughes et al., 2012).

Die Autoren untersuchten in diesem Zusammenhang, inwiefern eine Anpassung an andere Anwendungsländer erfolgt. Insgesamt drei Viertel der Institutionen nehmen kulturelle Adaptionen vor. Die Hälfte der Unternehmen überprüft die Übungsinhalte in Bezug auf die kulturelle Angemessenheit und adaptieren die Dimensionen und deren Definitionen an die jeweilige Kultur. Das Feedback wird in 42% und die Simulationen in 34% der Unternehmen angepasst. Ein Viertel der Organisationen überarbeitet die Verhaltensbeispiele für die Beobachtung und 37% beziehen kulturelle Unterschiede in das Beobachtertraining ein (Hughes et al., 2012).

#### 4.14 Entwicklung

Wie Tabelle 11 zeigt, ist die Entwicklung des AC in den verschiedenen Ländern mehrheitlich homogen. Mit Ausnahme der Studie von Krause et al. (2001), in der 57% das Verfahren adaptieren, sind die Prozentangaben derer, die eine Eigenentwicklung wählen, in den europäischen und den nordamerikanischen Ländern höher (Eurich et al., 2009; Höft & Obermann, 2008; Krause & Thornton, 2009). Die Ergebnisse der deutschsprachigen Studien unterscheiden sich für die Adaption,  $\chi^2(1, N = 340) = 19.11, p < .001$  und die Eigenentwicklung,  $\chi^2(1, N = 340) = 19.68, p < .001$ , signifikant (Höft & Obermann, 2008). Lediglich in Südafrika wählt man eher ein bestehendes AC, wobei jedes zweite Unternehmen eine Adaption vornimmt und 41% auf ein Standard-AC zurückgreifen (Krause et al., 2011). Verglichen mit den Angaben aus Westeuropa (Krause & Thornton, 2009) oder dem deutschsprachigen Raum (Höft & Obermann 2008, Krause et al., 2001) mit ca. 5% für das Standard-AC zeigt wie stark dieser Unterschied ist. Allerdings wählt auch etwa ein Fünftel der Unternehmen in Übersee diese Variante (Eurich et al., 2009, Krause & Thornton 2009). Obermann et al. (2012, zitiert nach Obermann, 2013) erhalten für den deutschsprachigen Raum, dass mehr Unternehmen auf ein Standard-AC zurückgreifen als in älteren Studien aus dem Gebiet. In einer Studie aus Belgien entwickeln acht von 23 Unternehmen das AC selbst. Die anderen 15 beauftragten hierfür eine Beratungsagentur (Lievens & Goemaere, 1999). In der Studie von Thornton und Krause (2009) geben fast drei Viertel der Unternehmen an, dass das AC für die angedachte Entscheidung entwickelt wird.

Bei der Revision des Verfahrens zeigt sich länderübergreifend ein einheitliches Bild. Die meisten AC werden jährlich oder alle zwei bis drei Jahre überarbeitet (Eurich et al., 2009; Höft & Obermann, 2008; Krause et al., 2001, 2011; Krause & Thornton; 2009). Bis zu 10 %

der Unternehmen in Westeuropa und Nordamerika entscheidet sich gegen eine Anpassung an Veränderungen (Eurich et al., 2009; Krause & Thornton, 2009).

Tabelle 11

*Entwicklung und Revision des Assessment Center in Prozent an der Stichprobengröße n*

	Krause & Thornton		Eurich et al.	Krause et al.	Höft & Obermann	Krause et al.
	(2009)		(2009)	(2011)	(2008)	(2001)
Land	WE	NA	USA	Südafrika	D, A, Ch	D, A, Ch
<i>n</i>	45	52	54	43	200	141
Eigenentwicklung	72	56	54	9	63	38
Adaption	23	26	24	50	34	57
Standard-AC	5	18	22	41	3	4
<b>Revision</b>						
Nie	7	10	7	—	—	—
Jährlich	33	45	39	22	—	—
2-3 Jahre	57	30	37	56	—	—
4-6 Jahre	3	15	17	19	—	—

*Anmerkungen.* —: keine Daten vorhanden; WE: Westeuropa; NA: Nordamerika; D: Deutschland, A: Österreich; Ch: Schweiz

## 5. Diskussion

Zunächst werden die Ergebnisse zur Umsetzung der Standards in der Praxis noch einmal kurz zusammengefasst. Anschließend werden die verschiedenen Studien kritisch beleuchtet und Einschränkungen diskutiert. Den Abschluss bildet ein Ausblick und Empfehlungen zur weiteren Forschung sowie zur Verbesserung der Standards.

### 5.1 Zusammenfassung der Ergebnisse

Insgesamt lässt sich festhalten, dass viele Empfehlungen der ITFACG (2009) in der Praxis Einzug gefunden haben. Allerdings werden in einigen Unternehmen diese Empfehlungen zumindest zum Teil rigoros ignoriert. Dabei finden sich weltweit Mängel in der Umsetzung. Thornton und Krause (2009) kommen zu dem Schluss, dass bei einem AC zum Zwecke der Personalauswahl 89% der befragten Unternehmen den Richtlinien der ITFACG (2009) folgen, wohingegen bei dem Ziel Personalentwicklung dies nur in 55% der Fällen zutreffend ist. Krause und Gebert (2003b) ermitteln einen höheren Aufwand in der Anwendung bei internen AC als bei externen AC.

Manche Inhalte wurden lediglich von wenigen Studien untersucht. Die Verwendung einer Anforderungs-Verfahrens-Matrix findet sich nur bei Krause et al. (2011), welche in 93% der Unternehmen in Südafrika zum Einsatz kommt. Die Interkulturalität wird ebenfalls lediglich von zwei internationalen Studien (Hughes et al., 2012; Thornton & Krause, 2009) thematisiert.

Während die Arbeits- und Anforderungsanalyse als Grundlage in mehr als 90% der Unternehmen zum Einsatz kommt, findet eine Evaluation des AC nur in rund der Hälfte der Organisationen statt (Eurich et al., 2009; Hughes et al., 2012; Kanning et al., 2007; Krause et al., 2001, 2011; Krause & Thornton, 2009; Lowry, 1996; Povah, 2011; Spychalski et al., 1997; Thornton & Krause, 2009). Die Frage, die sich hierdurch ergibt, ist wie die Unternehmen von der Korrektheit der Anforderungen überzeugt sein können, wenn die Inhaltsvalidität nicht erhoben wurde. Ein weiterer Mangel ist, dass in der Praxis das AC nicht wie empfohlen entsprechend der Ziele gestaltet wird. Bei der Übernahme eines Standard-AC ohne Adaption an die eigenen Bedürfnisse kann nicht garantiert werden, dass das Verfahren brauchbare Resultate liefert. Boltz, Kanning und Hüttemann (2009) entdeckten, dass die prognostische Validität umso höher ist, je mehr Standards in einem AC eingehalten werden. Ebenso sind die berücksichtigten Beobachtermerkmale und der Feedbackzeitpunkt zu bemängeln. Die eingesetzten Assessoren wirken sich auf das Akzeptanzurteil der Teilnehmer aus. Deshalb sollten die Unternehmen stärker auf demografische Merkmale bei der Beobachterausswahl achten. Das Feedback wird in vielen Organisationen mehr als eine Woche später gegeben (Eurich et al., 2009; Hughes et al., 2012; Krause et al., 2011; Krause & Thornton, 2009; Thornton & Krause, 2009). Je größer das Zeitintervall zwischen Darbietung und Rückmeldung ist, desto schlechter kann das Urteil nachvollzogen werden. Grund hierfür sind Erinnerungsverzerrungen.

Ein interessanter Aspekt aus den Studien ist der seltene Einsatz von technischen Mitteln (Eurich et al., 2009; Hughes et al., 2012; Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009). Das tägliche Leben von den Menschen weltweit wird durch Laptops, Smartphones und Tablets bestimmt. Kaum noch vorstellbar, dass jemand seine aktuelle Aktivität oder Meinung nicht zu jeder Zeit via Facebook oder Twitter teilen kann. Trotzdem findet das Internet bisher kaum Einzug in das AC. Verwunderlich dabei ist gerade der Aspekt, dass immer mehr arbeitsbezogene Aufgaben mit Hilfe eines Computers und dem Internet ausgeführt werden. So sind Präsentationen für mögliche neue Kunden via Skype alltäglich und in der Mitarbeitergewinnung setzen sich immer mehr die Internet-Bewerbung und die Anwerbung von potenziellen Bewerbern über Soziale Netzwerke durch. Durch eine

Anpassung des AC an das veränderte Kommunikationsverhalten können die Organisationen diesen Trends folgen.

Länderspezifische Unterschiede zwischen den USA und dem deutschsprachigen Raum werden von Krause und Gebert (2003a) durch kulturelle Differenzen erklärt. Die Autoren weisen hierbei auf einen institutionalisierten Kollektivismus hin, der in den deutschsprachigen Ländern stärker ausgeprägt ist als in den Vereinigten Staaten (Krause & Gebert, 2003a). Andere Studien untersuchten zwar ebenfalls Unterschiede zwischen verschiedenen Ländern, geben allerdings lediglich die Ergebnisse an ohne Erklärungsversuche zu liefern (Hughes et al., 2012; Krause & Thornton, 2009; Povah, 2011).

## 5.2 Mängel der verwendeten Studien

Die referierten Studien beinhalten sehr unterschiedlich große Stichproben. Krause et al. (2001) bezeichnen ihre Studie selbst als repräsentativ, jedoch kann bei einer Anzahl von 281 befragten Unternehmen aus Deutschland, Österreich und der Schweiz, von denen 141 das AC verwenden, nicht von einer ausreichend großen Stichprobe gesprochen werden. Andere Studien beziehen sich sogar nur auf etwa 50 Unternehmen. Dabei scheint gerade für die USA diese Anzahl sehr gering zu sein. In einem so großen Land gibt es zahlreiche Firmen, weshalb eine Menge von  $N = 54$  bei Eurich et al. (2009) lediglich einen geringen Anteil an den tatsächlichen Gegebenheiten widerspiegeln kann. Ebenso sind in den internationalen Untersuchungen (Hughes et al., 2012; Krause & Thornton, 2009; Povah, 2011; Thornton & Krause, 2009) einige Länder innerhalb eines Kontinentes überrepräsentiert und andere Länder fehlen gänzlich. So wird in der Studie von Krause und Thornton (2009) ein Vergleich zwischen Nordamerika und Westeuropa unternommen. Bei Betrachtung der einzelnen Stichprobengrößen der Länder fällt auf, dass von  $n = 54$  für Nordamerika 47 Organisationen aus den USA stammen. Ebenso verhält es sich bei den Angaben aus Westeuropa. Insgesamt beziehen sich die Resultate auf sieben Länder, wobei 33 von insgesamt 45 Unternehmen aus dem deutschsprachigen Raum stammen (Krause & Thornton, 2009). Handelt es sich hierbei also nicht eher um einen Vergleich zwischen den USA und Deutschland, Österreich und der Schweiz? In der Studie von Hughes et al. (2012) haben die meisten der 53 Länder einen Anteil von 1-5% an der Gesamtstichprobe von  $N = 543$ , wobei Großbritannien mit 148 Unternehmen, was einem Anteil von 29% entspricht, viel stärker vertreten ist.

Weiterhin muss die teilweise unterschiedliche Verwendung der Begriffe bemängelt werden. Die Vergleichsstudie von Krause und Gebert (2003a) enthält Begriffe samt Prozentangaben, die in der Originalstudie von Krause et al. (2001) so nicht zu finden sind.

Bestes Beispiel hierfür ist die Angabe zu dem Stelleninhaber als Datenquelle für die Anforderungsanalyse. Krause und Gebert (2003a) subsumieren die 78% unter dem Interview, während sie für Fragebögen und Beobachtungen 0% bzw. 1% nennen. Dies legt den Schluss nahe, dass sie der besseren Vergleichbarkeit der Studien die Komponenten der ähnlichsten Kategorie von Spychalski et al. (1997) zuordneten, ungeachtet ob diese beiden Aspekte wirklich das gleiche bei den Unternehmen erfragten. Aus diesem Grund wurde auch auf die Nennung solcher Daten verzichtet. Ebenso wenden auch Höft und Obermann (2008) solche Methoden an. So finden sie signifikante Unterschiede für die Anwendung der Anforderungsanalyse für die Jahre 2001 und 2008. Allerdings erwähnen die Autoren selbst, dass die Daten aus der 2008er Studie nur empirische Analysen erhoben, während 2001 die Grundlage nicht berücksichtigt wurde (Höft & Obermann, 2008). In einer weiteren Untersuchung von Krause und Gerber (2003b) findet man ebenfalls nicht nachvollziehbare Werte. Die Autoren geben beispielsweise für die Feedbackart die Ergebnisse eines Chi<sup>2</sup>-Test an. Bei genauerer Betrachtung fällt auf, dass sich dieser Wert nicht auf die zuvor erwähnte Stichprobe von  $n = 137$  bezieht, jedoch wird die verwendete Stichprobengröße nicht angegeben. Berechnungen ergeben, dass die Autoren die Ergebnisse auf Basis von  $N = 141$  errechneten. Zu Beginn der Studie wurde allerdings erwähnt, dass vier Fälle wegen nicht eindeutiger Zuordnung ausgeschlossen wurden. Wie konnten diese Werte also als Anteil externer AC in die Berechnung eingehen? Die Unterschiede zwischen intern und extern sind in beiden Fällen hoch signifikant, allerdings wurden wegen diesen mathematischen Unkorrektheiten solche Ergebnisse nicht im empirischen Teil dieser Thesis erwähnt. Solch erreichte Resultate sind für den Leser nicht nachvollziehbar. Wiederum tauchen in dieser Studie Angaben auf, die in der Originalstudie von Krause et al. (2001) nicht zu finden sind.

Gerade die Studie von Kanning et al. (2007) erntete sehr viel Aufmerksamkeit, vor allem in negativer Hinsicht. Die Autoren vermischen darin die Standards der ITFACG (2000), des AkAC (2004) und von Kanning (2004) zu 15 mehr oder weniger neuen Richtlinien. Da für diese Arbeit lediglich die Ergebnisse interessieren und die Standards in vielen Punkten den in Kapitel 3 referierten entsprachen, wurde die Studie trotz Kritik einiger Kollegen wie Krause (2009), Kersting (2009), Höft (2009) oder Sarges (2009) an der Ableitung dieser neuen, vermischten Richtlinien miteinbezogen.

Ein weiterer interessanter Punkt ist das Zustandekommen der einzelnen Prozentangaben. In den verschiedenen Studien werden die absoluten oder relativen Häufigkeiten der Unternehmen angegeben, die einen bestimmten Aspekt eines Standards berücksichtigen. Da viele der Ergebnisse zwischen den Studien sehr heterogen sind, wäre eine Einzelauflistung der

Unternehmen interessant. Hierbei ist nicht gemeint, ob Organisation A eine Anforderungsanalyse durchführt, sondern die Verteilung der Antworten innerhalb einer Organisation. So könnten die Ergebnisse darauf untersucht werden, ob es Unternehmen gibt, die jegliche Standards ignorieren und andere, die genau darauf achten entsprechend der Richtlinien zu handeln. Die Prozentangaben liefern leider keine Information darüber, inwiefern die Einhaltung der Richtlinien innerhalb der verschiedenen Unternehmen verteilt ist. Doch die Ergebnisse können dadurch verzerrt werden, dass es Fragebögen gibt, in denen die Antworten alle verneint oder bejaht wurden.

Gerade hiermit ist ein weiterer Aspekt verbunden. Alle referierten Studien setzen zur Datenerhebung einen Online-Fragebogen oder ein standardisiertes Interview per Telefon ein. Zum einen kann auf diese Weise nicht die Korrektheit der Angaben garantiert werden und zum anderen können die Ergebnisse durch unterschiedliches Verständnis der Fragen und ihrer Antwortmöglichkeiten verzerrt sein. Ebenso ist eine Verzerrung durch Beantwortung im Sinne der sozialen Erwünschtheit nicht ausschließbar. Eine Möglichkeit wäre die Anfrage zur stillen Teilnahme an einem AC, um zumindest den Prozess der Durchführung zu beobachten. Weiterhin könnte man den Fragebogen nicht nur von einer Person aus dem Unternehmen ausfüllen lassen, sondern wenn möglich alle beteiligten Personen darum bitten. Der Forscher erhält auf diese Weise die Möglichkeit die Antworten miteinander zu vergleichen und erkennt, ob es starke Unterschiede in den Ergebnissen gibt. So kann eher sichergestellt werden, dass die in eine länderspezifische oder nationenübergreifende Studie einfließenden Organisationsdaten korrekt sind.

### **5.3 Resümee zu den Standards**

Gerade die Privatstandards wie von Kanning (2004) weisen, wie in Kapitel 3 dargelegt, erhebliche Defizite auf. Er hat sich sicherlich sorgfältig mit dem Thema auseinandergesetzt, allerdings sind sie zu unpräzise, knapp und lückenhaft dargelegt. Wenn mehrere Personen etwas erarbeiten, werden verschiedene Perspektiven ineinander integriert. In diesem Fall garantiert dies einen umfassenderen Blick auf das Thema (Höft, 2007). Ein Anwender des AC wird mit den Standards von Kanning (2004) ein unzureichendes Ergebnis erhalten, da einfach zu viele wichtige Details fehlen. Bei der Erarbeitung solcher Empfehlungen sollten also mehrere Experten zusammenarbeiten, sodass man am Ende ein umfangreiches Ergebnis erhält (Höft, 2007). Der Aufbau der Standards von dem AkAC (2004) hat gegenüber den anderen beiden Richtlinien (ITFACG, 2009; Kanning, 2004) den Vorteil, dass eine Person durch die Nennung von Verstößen bei jedem einzelnen Standard überprüfen kann, ob er Fehler gemacht

hat. Jede Formulierung kann bei einem Leser zu einem unterschiedlichen Verständnis führen. So achtet der Rezipient darauf die genannten Aspekte zu befolgen, jedoch können dadurch trotzdem Fehler in dem AC enthalten sein. Der AkAC (2004) versucht gerade durch die explizite Nennung von Verfehlungen diesen mangelbehafteten Anteil zu verringern.

In den Standards sollte die Anwendung verschiedener Möglichkeiten der Eignungsdiagnostik stärker hervorgehoben werden. Ein Teilnehmer mit dem besten Abschneiden im AC zeigt, dass er die Aufgaben der Stelle mit hoher Wahrscheinlichkeit bewältigen kann. Allerdings sagen die Ergebnisse nicht unbedingt etwas darüber aus, ob die Person auch in das Klima und die Kultur der Organisation passt. So kann für die spezifische Stelle eigenverantwortliches Arbeiten gefordert sein, aber zwischen den Abteilungen herrscht eine kooperative Verzweigung. Der Stelleninhaber müsste also in der Lage sein selbstständig zu arbeiten und auf der anderen Seite aber auch hilfsbereit und teamfähig zu sein. Zusätzlich kann die Performanz während des AC durch eine schlechte Tagesform verzerrt sein. Da Persönlichkeitstests die typische Leistung erfassen, können sie weitere hilfreiche Informationen liefern.

Ein wesentlicher Kritikpunkt an den Standards des AkAC (2004) und von Kanning (2004) ist die Vernachlässigung der Interkulturalität. Durch Globalisierung und Unternehmensverlagerungen stammen die Bewerber aus unterschiedlichen Ländern mit verschiedenen Kulturen. Wie in Kapitel 3 bereits angeführt, können Teilnehmer auf Grund kultureller Eigenheiten anders bewertet werden (Ehret & Stumpf, 2006; Lanik & Gibbons, 2011; Schöning, 1996). Diese Themen müssen auch in nationalen Richtlinien zur Ansprache kommen. Hat ein Unternehmen Zweigstellen im Ausland und werden Mitarbeiter aus diesem Land angeworben, führt die Nichtberücksichtigung von kulturellen Merkmalen zu einer Fehleinschätzung der Bewerber.

Der Anwender sollte darauf achten, welche Standards er zur Konstruktion zu Grunde legt. Hilfreich wäre vermutlich eine internationale Arbeitsgruppe wie die ITFACG (2009), welche grundlegende Empfehlungen erarbeitet. Nationale Arbeitskreise sollten diese dann als Grundlage nehmen und nur gegebenenfalls notwendige Anpassung an das eigene Land vornehmen, wie beispielsweise sprachliche oder kulturelle Besonderheiten. So kann dem Anwender eine vernünftige Hilfe zur Seite gelegt werden, ohne dass er alle bekannten Standards übergeht, weil er nicht weiß, welche er nun zur Konstruktion anwenden soll. Allerdings zeigt sich beim Vergleich der Standards, dass selbst in Expertenkreisen keine einheitliche Meinung besteht (Kersting, Häcker & Hornke, 2011). So lange dies nicht der Fall ist, stellt sich die Frage, weshalb bei der Einführung eines AC in einem Unternehmen

irgendwelche Richtlinien beachtet werden sollten. Der Verantwortliche könnte sich ebenso auf seine eigenen Erfahrungen verlassen, denn die uneinheitlichen Standards können ihn zu dem Entschluss kommen lassen, dass es relativ gleichgültig ist, was er nun bei der Konstruktion beachtet, denn eine Expertengruppe besteht auf einen bestimmten Aspekt und eine andere Gruppe lehnt diesen Punkt ab. Einzelne wissenschaftliche Befunde werden in den Manuals nicht angegeben, um nicht eine Richtung zu favorisieren (Höft, 2009) bzw. weil sie eher dem Kriterium Nützlichkeit als empirischer Fundierung folgen (Kersting, 2009), jedoch können sie in dem Unternehmen sehr sinnvoll sein. Wenn der Entwickler weiß, dass die Ableitung der Anforderungen aus einer fundierten Analyse bessere Resultate erzielt, wird er eher einsehen diesen Aufwand zu betreiben. Wissenschaft und Praxis gehen nicht immer Hand in Hand, ist es aber nicht Ziel der Wissenschaft die offenen Fragen aus der Praxis zu beantworten? Welchen Sinn sollte also die Untersuchung der prognostischen Validität des AC in Abhängigkeit seiner Entwicklung haben, wenn die Ergebnisse in der Praxis unberücksichtigt bleiben? Die Umsetzung solcher empirischen Befunde in der Praxis wird durch einheitliche Regeln wesentlich erleichtert. Dem Praktiker wird somit eine Hilfe an die Hand gegeben, ohne dass er zunächst alle empirischen Befunde einzeln zusammentragen muss oder nach dem Zufallsprinzip eine der vielen Richtlinienmanuskripte auswählt. Hilfreich sind auch Listen über typische Fehler in der Entwicklung, Durchführung und Auswertung von AC wie von Caldwell, Thornton und Gruys (2003) am Beispiel der Standards der ITFACG (2009) abgeleitet. Darin beschreiben die Autoren zehn Mängel, welche Auswirkungen auf die Aussagekraft der Ergebnisse des AC haben und referieren auch empirische Befunde zu den Auswirkungen dieser fehlerhaften Anwendung. Mitarbeiter der Personalabteilung können die Liste als Referenzpunkt zum Vergleich der eigenen Umsetzung mit der best practice nehmen. Eine ähnliche Quelle existiert mit der DIN-Screen (Kersting, 2006) für den gesamten Prozess der Eignungsbeurteilung.

#### **5.4 Einschränkungen dieses Reviews**

Es wurde versucht einen umfassenden Überblick über die Standards des AC von verschiedenen Personengruppen zu geben. Anschließend wurden Befunde zu deren Umsetzung in der Praxis referiert. Das AC ist eine heterogene Methode. Es werden viele verschiedene Begriffe in Abhängigkeit von der Zielsetzung zur Umschreibung des Verfahrens wie Potentialcenter oder Development Center verwendet. Diese Arbeit berücksichtigt lediglich Studien, die die Personalauswahl miteinbezogen haben. Es wurden keine eigenen empirischen Daten erhoben, sodass bei den Befunden nur auf veröffentlichte Informationen

zurückgegriffen werden konnte. Eventuelle Untersuchungen, die zu völlig abweichenden Ergebnissen kamen, wurden möglicherweise nie veröffentlicht, sodass sich dadurch ein verfälschtes Bild ergibt.

## **5.5 Fazit und Ausblick**

Die Arbeit referierte verschiedene Standards zum AC. Anschließend erfolgte ein Vergleich auch in Bezug auf wissenschaftliche Befunde zur jeweiligen Empfehlung, damit der Leser einen Überblick erhält, welcher Standard von welchem Arbeitskreis als wissenschaftlich fundierter einzuschätzen ist. Verschiedene Studien zu der Umsetzung in Organisationen weltweit wurden zusammengetragen. So konnte gezeigt werden, dass erhebliches Verbesserungspotential sowohl in der Umsetzung als auch in der Koexistenz verschiedener Standards besteht. Der Forscher kann die Unternehmen nicht zu einer Teilnahme an einer Studie zwingen. Eine Möglichkeit wäre die Praktiker von der Relevanz des Themas zu überzeugen und gleichzeitig ihnen die Möglichkeit zur Verbesserung durch einen Bewertungsbogen der Standards anzubieten. So könnte die Teilnahmebereitschaft eventuell gesteigert werden, in dem Sinne, dass der Anwender das Gefühl vermittelt bekommt, selbst etwas beitragen zu können. Voraussetzung hierfür wären allerdings nationenübergreifend einheitliche Standards, denn es bringt einer Untersuchungsgruppe wenig, wenn sie Auskünfte zu der Anwendung und Verbesserung von Standards einer anderen Arbeitsgruppe erhalten.

Eine solche Variante haben Psychologen von der Humboldt-Universität in Berlin verfolgt (Schölmerich, Nachtwei & Schermuly, 2011). Zunächst stellten sie eine Liste mit Qualitätskriterien für ein AC zusammen. Anschließend befragten sie Unternehmen zur Einschätzung der Sinnhaftigkeit und Praktikabilität, sowie Wissenschaftler zur Konsistenz und Sinnhaftigkeit der Kriterien. Aus diesen Ergebnissen entwickelten die Forscher ein webbasiertes Programm namens BACDI, mit welchem die Unternehmen ihr AC bezüglich der Qualität überprüfen können. Die Firmen können so erkennen, inwiefern das eigene AC dem wissenschaftlichen Stand entspricht und wie das AC im Vergleich zur durchschnittlichen Praxis in Deutschland einzuordnen ist (Schölmerich et al., 2011). Ein aktuelles Vorhaben ist die Erweiterung des BACDI-Projekts auf Nordamerika. Durch die Verwendung desselben Instruments haben die Forscher das Ziel einer Untersuchung von Unterschieden in der Anwendung des AC zwischen den deutschsprachigen Ländern und Nordamerika (Bairov, Nachtwei, von Bernstorff, Uedelhoven & Schermuly, 2013). Solche Ansätze können die Anwendung des AC in der Praxis positiv beeinflussen.

## Literaturverzeichnis

- Arbeitskreis Assessment Center e.V. (2004). *Standards der Assessment Center Technik*. Abgerufen von [http://www.arbeitskreis-ac.de/images/attachments/149\\_ac-standards\\_2004.pdf](http://www.arbeitskreis-ac.de/images/attachments/149_ac-standards_2004.pdf)
- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–154. doi: 10.1111/j.1744-6570.2003.tb00146.x
- Assessment Centre Study Group (2007). *Guidelines for assessment and development centres in South Africa*. Retrieved from [http://www.acsg.co.za/ac\\_information/guidelines/acdcguidelines.pdf](http://www.acsg.co.za/ac_information/guidelines/acdcguidelines.pdf)
- Bairov, H., Nachtwei, J., von Bernstorff, C., Uedelhoven, S. & Schermuly, C.C. (2013). Benchmark for assessment center diagnostics (BACDi) - A cross cultural comparison between the USA and the German speaking region. *Intercontinental Journal of Human Resource Research Review, 1*(9), 1-14. Retrieved from <http://icmrr.org/Archives.html>
- Becker, N., Höft, S., Holzenkamp, M. & Spinath, F.M. (2011). The predictive validity of assessment centers in German-speaking regions. A meta-analysis. *Journal of Personnel Psychology, 10*, 61-69. doi: 10.1027/1866-5888/a000031
- Bernardin, H. J., Konopaske, R., & Hagan, C. M. (2012). A comparison of adverse impact levels based on top-down, multisource, and assessment center data: Promoting diversity and reducing legal challenges. *Human Resource Management, 51*, 313-341. doi: 10.1002/hrm.21472
- Boltz, J., Kanning, U.P. & Hüttemann, T. (2009). Treffende Prognosen durch Beachtung von Standards. *Personalführung, 10*, 32-37. Abgerufen von <http://www.dgfp.de/wissen/personalwissen-direkt/dokument/83306/herunterladen>
- Boyle, S., Fullerton, J., & Wood, R. (1995). Do assessment/development centres use optimum evaluation procedures? A survey of practice in UK organizations. *International Journal of Selection and Assessment, 3*, 132-140. doi: 10.1111/j.1468-2389.1995.tb00017.x
- British Psychology Society (2003). *Design, implementation and evaluation of assessment and development centres - Best practice guidelines*. Retrieved from <http://www.psychtesting.org.uk/the-ptc/guidelinesandinformation.cfm>
- Caldwell, C., Thornton, G. C., III., & Gruys, M. L. (2003). Ten classic assessment center errors: Challenges to selection validity. *Public Personnel Management, 32*, 73-88. doi: 10.1177/009102600303200104

- Clapham, M. M. (1998). A comparison of assessor and self dimension ratings in an advanced management assessment center. *Journal of Occupational and Organizational Psychology, 71*, 193-203. doi: 10.1111/j.2044-8325.1998.tb00672.x
- Damitz, M. M., Manzey, D. D., Kleinmann, M. M., & Severin, K. K. (2003). Assessment center for pilot selection: Construct and criterion validity and the impact of assessor type. *Applied Psychology: An International Review, 52*, 193-212. doi: 10.1111/1464-0597.00131
- Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93*, 685-691. doi: 10.1037/0021-9010.93.3.685
- Delmestri, G., & Walgenbach, P. (2009). Interference among conflicting institutions and technical-economic conditions: the adoption of the assessment center in French, German, Italian, UK, and US multinational firms. *International Journal of Human Resource Management, 20*, 885-911. doi: 10.1080/09585190902770828
- Eck, C.D., Jöri, H. & Vogt, M. (2010). Assessment Center. Entwicklung und Anwendung (2., überarbeitete und erweiterte Aufl.). doi: 10.1007/978-3-642-12998-8
- Ehret, A., & Stumpf, S. (2006). Interkulturelle Assessment Center: Anforderungen und Gestaltungsempfehlungen. *Wirtschaftspsychologie, 8*(4), 20-30. Abgerufen von <http://www.psychologie-aktuell.com/index.php?id=wirtschaftspsychologie>
- Eurich, T., Krause, D., Cigularov, K., & Thornton, G. C., III. (2009). Assessment Centers: Current practices in the United States. *Journal of Business & Psychology, 24*, 387-407. doi: 10.1007/s10869-009-9123-3
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511. doi: 10.1037/0021-9010.72.3.493
- Gaugler, B. B., & Thornton, G. C., III. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*, 611-618. doi: 10.1037/0021-9010.74.4.611
- Fisseni, H.-J., Fennekels, G. P., (1995). *Das Assessment Center: Eine Einführung für Praktiker*. Göttingen: Hogrefe
- Flanagan, J. C. (1954). The critical incidents technique. *Psychological Bulletin, 51*, 327-358. doi: 10.1037/h0061470

- Hardison, C. M., & Sackett, P. R. (2007). Kriteriumsbezogene Validität des Assessment Centers: lebendig und wohlauf? In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 192–202). Göttingen, Germany: Hogrefe.
- Hell, B., Boramir, I., Schaar, H., & Schuler, H. (2006). Interne Personalauswahl und Personalentwicklung in deutschen Unternehmen. *Wirtschaftspsychologie*, 8(1), 2-22. Abgerufen von <http://www.psychologie-aktuell.com/index.php?id=wirtschaftspsychologie>
- Hennessy, J., Mabey, B., & Warr, P. (1998). Assessment centre observation procedures: An experimental comparison of traditional, checklist and coding methods. *International Journal of Selection and Assessment*, 6, 222-231. doi: 10.1111/1468-2389.00093
- Höft, S. (2007). Assessment Center. In H. Schuler & K. Sonntag (Hrsg.) *Handbuch der Arbeits- und Organisationspsychologie* [Adobe Digital Edition version] (S. 475-482). Abgerufen von <http://www.ciando.com/>
- Höft, S. (2009). Gutes AC, schlechtes AC? Ein kritischer Kommentar zu den wissenschaftlichen Qualitätskriterien für Assessment Center von Kanning et al. (2007). *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 74-77. doi: 10.1026/0932-4089.53.2.74
- Höft & Funke (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.) (2., überarbeitete und erweiterte Aufl.), *Lehrbuch der Personalpsychologie* [Adobe Digital Edition version] (S. 145-188). Abgerufen unter <http://www.ciando.com/>
- Höft, S. & Obermann, C. (2010). Der Praxiseinsatz von Assessment Centern im deutschsprachigen Raum: Eine zeitliche Verlaufsanalyse basierend auf den Anwenderbefragungen des Arbeitskreises Assessment Center e.V. von 2001 und 2008. *Wirtschaftspsychologie*, 12 (2), 5-16. Abgerufen von <http://www.psychologie-aktuell.com/index.php?id=wirtschaftspsychologie>
- Hughes, D., Riley, P., Shalfrooshan, A., Gibbons, A. & Thornton, G. C. III. (2012). *A Global Survey of Assessment Centre Practices*. Retrieved from <http://www.adc.uk.com/resources/research/findings/>
- International Task Force on Assessment Center Guidelines (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management*, 29, 315-331. doi:10.1177/009102600002900302

- International Task Force on Assessment Center Guidelines (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment*, 17, 243-253. doi: 10.1111/j.1468-2389.2009.00467.x
- Kanning, U. P. (2004). *Standards der Personaldiagnostik. Ein Lehrbuch für die Praxis*. Göttingen: Hogrefe.
- Kanning, U.P., Pöttker, J. & Gelléri, P. (2007). Assessment Center Praxis in deutschen Großunternehmen. Ein Vergleich zwischen wissenschaftlichem Anspruch und Realität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 51, 155-167. doi: 10.1026/0932-4089.51.4.155
- Kersting, M. (2006). „DIN Screen“ – Leitfaden zur Kontrolle und Optimierung der Qualität von Verfahren und deren Einsatz bei beruflichen Eignungsbeurteilungen. Lengerich: Pabst Science Publishers.
- Kersting, M. (2009). Qualität im Assessment Center. *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 70-74. doi: 10.1026/0932-4089.53.2.70
- Kersting, M. (2010). Akzeptanz von Assessment Centern: Was kommt an und worauf kommt es an? *Wirtschaftspsychologie*, 12(2), 58-65. Abgerufen von <http://www.psychologie-aktuell.com/index.php?id=wirtschaftspsychologie>
- Kersting, M., Häcker, H. O., & Hornke, L. F. (2011). Qualitätsstandards in der psychologischen Diagnostik. In L. F. Hornke, M. Amelang, M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder psychologischer Diagnostik* [Adobe Digital Edition version] (S. 1-86). Abgerufen unter <http://www.ciando.com/>
- Kleinmann, M. (2003). *Assessment-Center*. Göttingen: Hogrefe.
- Kolk N. J., Born, M. Ph., & van der Flier, H. (2003). The transparent assessment centre: The effects of revealing dimensions to candidates. *Applied Psychology: An International Review*, 52, 648-668. doi: 10.1111/1464-0597.00156
- Kolk, N. J., Born, M. Ph., van der Flier, H., & Olman, J. M. (2002). Assessment center procedures: Cognitive load during the observation phase. *International Journal Of Selection And Assessment*, 10, 271-278. doi: 10.1111/1468-2389.00217
- Krause, D. E. (2009). Ob Kritik konstruktiv wird, entscheidet der Kritisierte - Kritische Reflexion der AC-Standards. *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 77-79. doi: 10.1026/0932-4089.53.2.77

- Krause, D. E., & Gebert, D. (2003a). A comparison of assessment center practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment*, *11*, 297-312. doi: 10.1111/j.0965-075X.2003.00253.x
- Krause, D. E. & Gebert, D. (2003b). Unterschiede in der Durchführung zwischen internen und externen Assessment Centern – Eine empirische Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, *47*, 87-94. doi: 10.1026//0932-4089.47.2.87
- Krause, D. E., Meyer zu Kniendorf, C., & Gebert, D. (2001). Aktuelle Trends in der Assessment-Center-Anwendung. *Wirtschaftspsychologie aktuell*, *3*, 47-55. Abgerufen von <http://www.wirtschaftspsychologie-aktuell.de/archiv.html>
- Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, N., & Joubert, T. (2011). Assessment center practices in South Africa. *International Journal of Selection and Assessment*, *19*, 262-275. doi: 10.1111/j.1468-2389.2011.00555.x
- Krause, D.E. & Thornton, G.C., III. (2007). Kulturelle Vielfalt und Assessment-Center. Handlungsempfehlungen für Forschung und Praxis. *Zeitschrift Führung + Organisation*, *76*, 223-230. Abgerufen unter [http://www.zfo.de/index.php?mod=docDetail&docID=356\\_12&search=kulturelle vielfalt\\*](http://www.zfo.de/index.php?mod=docDetail&docID=356_12&search=kulturelle%20vielfalt*)
- Krause, D. E., & Thornton, G. C., III. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. *Applied Psychology: An International Review*, *58*, 557-585. doi: 10.1111/j.1464-0597.2008.00371.x
- Lanik, M., & Mitchell Gibbons, A. (2011). Guidelines for cross-cultural assessor training in multicultural assessment centers. *The Psychologist-Manager Journal*, *14*, 221-246. doi: 10.1080/10887156.2011.595970
- Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, *86*, 1202-1222. doi: 10.1037/0021-9010.86.6.1202
- Lievens, F., & Goemaere, H. (1999). A different look at assessment centers: Views of assessment center users. *International Journal of Selection and Assessment*, *7*, 215-219. doi: 10.1111/1468-2389.00122
- Lievens, F., & Thornton, G. C., III. (2005). Assessment centers: recent developments in practice and research. In A. Evers, O. Smit-Voskuil, & N. Anderson (Eds.), *Handbook of Selection* (pp. 243-264). Retrieved from <http://users.ugent.be/~flievens/Site/Research.html>

- Lowry, P. E. (1996). A survey of the assessment center process in the public sector. *Public Personnel Management*, 25(3), 307-321. doi:10.1177/009102609602500304
- Marcus, B. (2011). *Personalpsychologie*. doi: 10.1007/978-3-531-93093-0
- Marggraf-Micheel, C., Höft, S., & Bonnist, H. (2006). Practice und Coaching: Strategien zur Vorbereitung von Teilnehmern auf Assessment-Center-Verfahren. *Wirtschaftspsychologie*, 8(4), 31-39. Abgerufen von <http://www.psychologie-aktuell.com/index.php?id=wirtschaftspsychologie>
- Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93, 1042-1052. doi: 10.1037/0021-9010.93.5.1042
- Nerdinger, F. W., Blickle, G., & Schaper, N. (2011). *Arbeits- und Organisationspsychologie* (2. Aufl.). doi: 10.1007/978-3-642-16972-4
- Obermann, C. (2013). *Assessment-Center. Entwicklung - Durchführung - Trends* (5. überarbeitete und erweiterte Aufl.). doi: 10.1007/978-3-8349-3813-8
- Pendit, V. G. (2011) Assessment center adaptation and implementation in Indonesia. In N. Povah & G. C. Thornton III. (Eds.) *Assessment centres and global talent management* (pp. 363–374). Surrey: Gower Publishing Limited.
- Povah N (2011) A review of recent international surveys into assessment centre practices. In N. Povah & G. C. Thornton III. (Eds.) *Assessment centres and global talent management* (pp. 329–350). Surrey: Gower Publishing Limited.
- Rodriguez, D., Patel, R., Bright, A., Gregory, D., & Gowing, M. K. (2002). Developing competency models to promote integrated human-resource practices. *Human Resource Management*, 41, 309-324. doi:10.1002/hrm.10043
- Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401-410. doi:10.1037/0021-9010.67.4.401
- Sarges, W. (2001). Die Assessment Center-Methode. Herkunft, Kritik und Weiterentwicklungen. In W. Sarges (Hrsg.) (2., überarbeitete und erweiterte Aufl.), *Weiterentwicklungen der Assessment-Center-Methode* (S. 7-32). Abgerufen von <http://www.sarges-partner.de/management-diagnostik.php>
- Sarges, W. (2009). Warum Assessment Center häufig zu kurz greifen und zudem meist das Falsche zu messen versuchen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 79-82. doi:10.1026/0932-4089.53.2.79

- Schippmann, J. S., Hughes, G. L., & Prien, E. P. (1987). The use of structured multi-domain job analysis for the construction of assessment center methods and procedures. *Journal of Business and Psychology, 1*, 353-366. doi:10.1007/BF01018144
- Schleicher, D. J., Mayes, B. T., Day, D. V., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*, 735-746. doi: 10.1037/0021-9010.87.4.735
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274. doi:10.1037/0033-2909.124.2.262
- Schölmerich, F., Nachtwei, J. & Schermuly, C. C. (2011). Evaluating the quality of assessment centers used in employee selection – Development of a Benchmark for Assessment Diagnostics (BACDi). In: B. Krause, R. Beyer und G. Kaul (Hrsg.), *Empirische Evaluationsmethoden Band 15* (S. 5-16). Abgerufen von <http://www.bacdi.de/articleZeE.pdf>
- Schuler H. (2006) Stand und Perspektiven der Personalpsychologie. *Zeitschrift für Arbeits- und Organisationspsychologie, 4*, 176-188. doi:10.1026/0932-4089.50.4.176
- Schuler, H. (2007). Assessment Center als multiples Verfahren zur Potenzialanalyse: Einleitung und Überblick. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 3-36). Göttingen: Hogrefe.
- Schuler, H. (2008). Das Bewusstsein für mehr Qualität schärfen. *Personalwirtschaft, 35*(4), 18-20. Abgerufen von [http://archiv.personalwirtschaft.de/wkd\\_pw/cms/website.php?id=/de/jahrgang/2008/04/personalwirtschaft](http://archiv.personalwirtschaft.de/wkd_pw/cms/website.php?id=/de/jahrgang/2008/04/personalwirtschaft)
- Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie, 6*, 60-70. doi:10.1026/1617-6391.6.2.60
- Spinath, F. M., & Becker, N. (2011). Verhaltensbeobachtung. In L. F. Hornke, M. Amelang, M. Kersting (Hrsg.) , *Methoden der psychologischen Diagnostik* [Adobe Digital Edition version] (S. 325-369). Abgerufen von <http://www.ciando.com/>
- Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology, 50*, 71-90. doi: 10.1111/j.1744-6570.1997.tb00901.x

- Swiss Assessment (2007). *Qualitätsstandards von Swiss Assessment zur Entwicklung, Durchführung und Auswertung von Assessment Center*. Abgerufen von [http://www.swissassessment.ch/files/ac\\_standards\\_swiss\\_assessment.pdf](http://www.swissassessment.ch/files/ac_standards_swiss_assessment.pdf)
- Thornton, G. C., III., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, *19*, 169-187. doi: 10.1016/j.hrmr.2009.02.002
- Thornton, G. C., III., & Krause, D. E. (2009). Selection versus development assessment centers: An international survey of design, execution, and evaluation. *International Journal of Human Resource Management*, *20*, 478-498. doi: 10.1080/09585190802673536
- Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, *29*, 231–258. doi: 10.1177/014920630302900206
- Woehr, D. J., Arthur, W., Jr., & Meriac, J. (2007). Methodenfaktoren statt Fehlervarianz: eine Metaanalyse der Assessment Center-Konstruktvalidität. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 81-108). Göttingen: Hogrefe.

## **Eidesstattliche Erklärung**

Hiermit versichere ich, dass ich die vorliegende Abschlussarbeit mit dem Thema

### **Assessment Center in der Personalauswahl: Wunsch und Wirklichkeit**

ohne fremde Hilfe erstellt habe. Alle verwendeten Quellen wurden angegeben. Ich versichere, dass ich bisher keine Haus- oder Prüfungsarbeit mit gleichem oder ähnlichem Thema an der FernUniversität oder einer anderen Hochschule eingereicht habe.

Datum: \_\_\_\_\_ Unterschrift: \_\_\_\_\_