

# Master-Arbeit

zur Erlangung des Titels «Master of Advanced Studies in  
Psychology of Career Counseling and Human Resources Management  
MASP-CC&HRM» der Universitäten Bern, Freiburg und Zürich

**Titel**                    **Prüfung von Skalen zur Ermittlung der Sozialen Validität und  
Prozeduralen Gerechtigkeit anhand des Multichecks**

Eingereicht von **Marcel Widmer**  
lic. phil., Psychologe  
Metzgergasse 1  
3400 Burgdorf

**Peter Eberle**  
lic. phil., Psychologe  
Hochfeldstrasse 24  
3012 Bern

Eingereicht zur Beurteilung gemäss den Masterarbeitsrichtlinien für das NABB-10 bei den  
nachstehenden Experten:

Experte I  
**Martin Kleinmann**  
Prof. Dr. phil., Psychologe  
Universität Zürich  
Binzmühlestrasse 14/12  
8050 Zürich

Experte II  
**Jörg Renz**  
Dr. phil., Psychologe  
BIZ Bern-Mittelland  
Bremgartenstrasse 37  
3001 Bern

## **Master-Arbeit**

zur Erlangung des Titels «Master of Advanced Studies in Psychology of Career Counseling and Human Resources Management MASP-CC&HRM» der Universitäten Bern, Freiburg und Zürich

**Titel**                    **Prüfung von Skalen zur Ermittlung der Sozialen Validität und Prozeduralen Gerechtigkeit anhand des Multichecks**

Eingereicht von **Marcel Widmer**

lic. phil., Psychologe  
Metzgergasse 1  
3400 Burgdorf

**Peter Eberle**

lic. phil., Psychologe  
Hochfeldstrasse 24  
3012 Bern

**Bestätigung**            Wir bestätigen ehrenwörtlich, dass wir die vorliegende Master-Arbeit selbständig und ohne unerlaubte Hilfe erarbeitet haben und sie an keiner anderen Institution eingereicht haben. Die bei den Experten und beim NABB eingereichten Exemplare der Master-Arbeit sind text- und layout-identisch abgefasst.

Ort, Datum und  
Unterschrift  
der Autoren

---

Die Experten bestätigen nachstehend mit ihren Unterschriften, dass die vorliegende Master-Arbeit in Umfang und Inhalt/Qualität den im NABB-Curriculum vom 12.05.2003/18.09.2003 mit Stand am 05.04.2005 und den in den Master-Arbeitsrichtlinien gem. Art. 9.2. des NABB-Curriculums genannten Anforderungen genügt und damit das Erreichen der formulierten Kompetenzziele durch die Verfasser hinreichend nachgewiesen ist. Die ausführliche Bewertung der vorliegenden Master-Arbeit ist in einem zusätzlichen Dokument festgehalten.

Ort, Datum und  
Unterschrift  
der Experten

---

### **Postanschrift des Veranstalters**

[die Universität Freiburg ist leading-house für das NABB-10]

Postgraduale Weiterbildung NABB zum «Master of Advanced Studies in Psychology of Career Counseling and Human Resources Management MASP-CC&HRM» der Universitäten Bern, Freiburg und Zürich, p. A. Universität Freiburg, Departement für Psychologie, Abteilung Personal- und Organisationspsychologie, Rue de Faucigny 2, CH-1700 Freiburg  
nabb@unifr.ch, www.nabb.ch

NABB-10/BST/090824

## Inhaltsverzeichnis:

1. Einleitung und Ziele der Studie	06
2. Theorie	09
2.1 Gerechtigkeit	09
2.1.1 Prozedurale Gerechtigkeit	10
2.2 Soziale Validität	14
3. Hypothesen	18
3.1 Reliabilität	18
3.2 Konvergente Validität	18
3.3 Diskriminante Validität	19
4. Methode	21
4.1 Fragebogen	21
4.2 Datenerhebung	23
4.3 Stichprobe	24
4.4 Datenauswertung	25
5. Ergebnisse	28
5.1 Reliabilität	28
5.2 Konvergente Validität	31
5.3 Diskriminante Validität	33
5.4 Mittelwertsvergleiche	34
6. Diskussion	37
6.1 Validität/Konstruktvalidität	37
6.2 Mittelwertsvergleiche	40
6.3 Empfehlungen für die Firma Multicheck	42
7. Anhänge	44
A) Literaturverzeichnis	44
B) Fragebogen zur Sozialen Validität nach Klingner (2003)	47
C) Fragebogen zur Prozeduralen Gerechtigkeit nach Müller et al. (2007) resp. Bauer et al. (2001)	48
D) Exkurs: Was ist subjektive Fairness?	50
E) Exkurs: Welche Attribute des einen Instrumentes zur Akzeptanz klärt die des anderen auf?	51
F) Vergleich der 4 Facetten der sozialen Validität des Multichecks mit anderen Verfahren	52

Kurzfassung der Masterarbeit (Praxisprojekt und schriftliche Master-Arbeit) aus der postgradualen Weiterbildung NABB zum «Master of Advanced Studies in Psychology of Career Counseling and Human Resources Management MASP-CC&HRM» der Universitäten Bern, Freiburg und Zürich.

**Problemstellung** Berufseignungstests spielen vermehrt eine Rolle in der Selektion von Auszubildenden durch Wirtschaftsbetriebe und Berufsbildungsinstitutionen. Der Multicheck, ein bekanntes Verfahren, wurde in einer vorhergehenden Studie von einem der Autoren hinsichtlich der Konstrukt- und Vorhersagevalidität untersucht (Widmer, 2006). Unbekannt ist dagegen die Akzeptanz und Fairness dieser Verfahren. Im Rahmen dieser Arbeit wird die Akzeptanz und Fairness durch zwei psychologische Konstrukte operationalisiert und mittels entsprechender Fragebögen (Fragebogen zur sozialen Validität von Klingner, 2003, und die Selection Procedural Justice Skale SPJS von Baur, 2001) am Beispiel des Multicheck untersucht.

**Fragestellungen** Die Fragestellungen beziehen sich auf drei Themen:

- 1) Es soll überprüft werden, inwiefern die Reliabilität der eingesetzten Verfahren den Werten aus anderen Ländern und in anderen Settings entspricht.
- 2) Es ist von Interesse, wie die konvergente und divergente Validität der zwei eingesetzten Verfahren ausfallen.
- 3) Es soll untersucht werden, wie die Teilnehmenden den Multicheck als Auswahlverfahren im Vergleich mit anderen Eignungsabklärungsverfahren beurteilen und welche Massnahmen dies zum Positiven ändern könnten.

**Resultate** Die Reliabilitäten der eingesetzten Tests entsprechen den Erwartungen und sind als gut zu bezeichnen, wenn die Tests jeweils als Ganzes untersucht werden. Nur wenig zufrieden stellend sind die Teilskalen der Tests, was möglicherweise an deren Operationalisierung liegen könnte.

Die Resultate zur Konvergenz und Divergenz der Verfahren zur Messung der Akzeptanz bestätigen die Ergebnisse zur Reliabilität: Die Übereinstimmung der Verfahren als Ganzes entspricht den Erwartungen und die hohen Korrelationen (grösser als .60) ergeben sich auch im Zusammenhang mit der subjektiven Beurteilung durch die Teilnehmenden des Multichecks. Allerdings resultieren unklare Abgrenzungen zwischen Teilskalen (Facetten) innerhalb der Instrumente.

Die Beurteilung der Akzeptanz des Multicheck entspricht insofern den Erwartungen, als dass der Eignungstest im Vergleich mit einem Assessment Center in den meisten Aspekten der Prozeduralen Gerechtigkeit schlechter bewertet wird. Hinsichtlich wichtiger Aspekte der sozialen Validität wird der Multicheck allerdings besser beurteilt, als es im Vergleich mit Intelligenztests hypothetisch angenommen wurde. Es wird diskutiert, dass diese Ergebnisse sowohl durch die Unterschiedlichkeit der Verfahren als auch durch die Besonderheit der Studie resp. der Stichprobe zustande kommen.

**Methodik** Die verwendeten Konstrukte zur Akzeptanzforschung stammen entweder aus dem deutschsprachigen Raum (Soziale Validität, vgl. Schuler & Stehle, 1983) oder dem angloamerikanischen Raum (Gilliland, 1993).

Die dazugehörigen Fragebögen (vgl. Klingner, 2003, und Baur, 2001) wurden an die Rahmenbedingungen der Studie angepasst und in eine internetbasierte Version übersetzt.

Befragt wurden zum grössten Teil 14- bis 16-jährige Jugendliche, die vor kurzer Zeit den Multicheck absolviert und die entsprechenden Resultate erhalten hatten.

Für die Reliabilität wurde Cronbachs Alpha bestimmt, für die Konstruktvalidität wurden Produkt-Moment- sowie Spearman-Rho-Korrelationen und für die Mittelwertsunterschiede (unabhängige) t-Tests berechnet. Die Auswertungen erfolgten mit SPSS Version 14.

**theoretischer Hintergrund** Die Untersuchung basiert auf einer früheren Untersuchung zum Berufseignungstest Multicheck (Widmer, 2006). Sie nimmt damit Bezug zur vielfältigen Literatur, die sich mit der Theorie und den Gütekriterien von Intelligenz- und Eignungstests beschäftigt, und hier im Speziellen auch zur Akzeptanz- und Fairnessforschung.

Die Grundlagen zur Prozeduralen Gerechtigkeit finden sich bei Gilliland (1993). Gilliland propagiert dabei ein Modell, welches sich auf die aktuelle Gerechtigkeitsforschung stützt und Bezug auf die organisationalen Gerechtigkeitswahrnehmung einer Person nimmt. Zwei Facetten, die prozedurale und distributive Gerechtigkeit, bestimmen diese Wahrnehmung.

Die Soziale Validität wurde von Schuler und Stehle (1983) als Mass für die Akzeptabilität von Eignungstests und –verfahren für Bewerbende diskutiert. Es setzt sich aus vier Komponenten zusammen, namentlich Information, Transparenz, Partizipation und Urteilskommunikation.

Bei den genannten Konstrukten der Akzeptanzforschung ist es gemein, dass sie den Fokus auf die Beurteilungen von eignungsdiagnostischen Situationen durch die Bewerbenden legen. Auf diesem Hintergrund definieren die Autoren das übergeordnete Ziel dieser Studie, Akzeptanz zu erfassen, so dass Massnahmen für Beteiligte abgeleitet werden können, um damit Situationen in der Personalwirtschaft sozial verträglicher zu machen.

<b>Verfasser</b>	Peter Eberle lic. phil., Psychologe Hochfeldstrasse 24 3012 Bern peter.eberle@powersurf.li	Marcel Widmer lic. phil., Psychologe Metzgergasse 1 3400 Burgdorf marcel.widmer2@gmx.ch
------------------	--	---

Die Verfasser/innen stehen für Auskünfte, Anfragen und Bemerkungen zum Inhalt der vorliegenden Publikation zur Verfügung. Die Master-Arbeit selbst kann bei den Verfasser/innen gegen eine Unkostenpauschale von Fr. 20.– für Porto und Verpackung ausgeliehen werden.

### Copyright

© NABB – Kopieren mit Quellenangabe gestattet

## 1. Einleitung und Ziele der Studie

Die Berufseignungsdiagnostik hat in den vergangenen Jahren erheblich an Bedeutung gewonnen. Es kommen vermehrt neue Verfahren aus dem psychologisch-wissenschaftlichen Bereich auf den Markt, welche vor allem im Bereich der Berufsberatung verwendet werden. Gleichzeitig vermissen viele Leute mit wirtschaftlichem Hintergrund, welche im Personalbereich von Firmen tätig sind, entsprechende Instrumente mit geringeren Interpretationsanforderungen für die Abklärung der Eignung ihrer Bewerber. Deshalb versuchen, und dies insbesondere im Auszubildenden-Bereich, privatwirtschaftliche Unternehmungen oder Verbände mit eigenen Instrumenten die Nachfrage zu decken (vgl. Grandjean, 2007). Die wohl bekanntesten Verfahren heissen Multicheck und Basic-Check. Sie decken einen grossen Teil der genannten Bedürfnisse für berufsabhängige beziehungsweise -unabhängigen Tests ab. Obwohl sich die Tests dieser Anbieter in einigen Merkmalen unterscheiden, stellen sie doch eine augenscheinlich gültige Kombination von Potential- und Schulleistungstests basierend auf Theorien der Intelligenz und Schulfähigkeit dar. Weitere Informationen zum Testverfahren vergleiche man beispielsweise bei Widmer (2006).

Aufgrund der Möglichkeit, Daten für wissenschaftliche Zwecke zu nutzen, haben sich die Autoren dieser Studien entschieden, das Instrument der Firma Multicheck einzusetzen. Die Firma Multicheck ist ein Personaldienstleistungsverfahren mit Sitz in Konolfingen (Kanton Bern) und besteht seit 1998. Zweck der Gründung vor gut zehn Jahren war es, mit praxisbezogenen und berufsspezifischen Eignungstests Ausbildungs- und Personalabteilungen in Fragen der Personalselektion, personenbezogenen Standortbestimmung und Personalentwicklung zu unterstützen. Innerhalb der genannten Geschäftsfelder konnte sich Multicheck vor allem als Anbieter von Berufseignungstests für Jugendliche profilieren. Mittlerweile entwickelt die Firma aber auch Eignungs- und Persönlichkeitstests für andere Zielgruppen (Berufstätige, Fach- und Führungskräfte). Aktuell beabsichtigt die Firma eine Expansion ins Ausland, primär in den deutschsprachigen Raum.

Obwohl viele dieser Verfahren keine wissenschaftliche Theorie als Grundlage haben, gibt es neuerdings vermehrt Untersuchungen zur „technischen Validität“ (Schuler & Stehle, 1983) derselben.

In einer empirischen Untersuchungen zum basic-check untersuchten Jungo und Zihlmann (2002) die Reliabilität und Konstruktvalidität des basic-checks. Auch zum Multicheck wurden bereits zwei wissenschaftliche Arbeiten verfasst (Widmer, 2006; Grandjean, 2007). In einer Studie zum Multicheck untersuchte Widmer (2006) mehrere Aspekte der

Konstrukt- und Vorhersagevalidität des Multicheck. In Bezug auf letztere Untersuchung lässt sich exemplarisch zeigen, dass trotz geringen wissenschaftlichen Hintergrunds Verfahren relativ valide sein können. So konnte Widmer (2006) zeigen, dass ein Eignungstest wie der Multicheck in Bezug auf die Vorhersage von schulischen Leistungen in der Ausbildung inkrementelle Validität über Schulnoten aus der abgebenden Schule aufweist. Allerdings gibt es auch Grenzen in der Nützlichkeit solcher Eignungstests, denn wie es sich in der gerade erwähnten Studie gezeigt hat, sind diese Verfahren nur begrenzt für andere Kriterien wie Ausbildungsbeurteilungen durch Vorgesetzte und Berufsbildende gültig.

Trotz der ersten Ergebnisse zu diesen Verfahren ist die Grundlage empirischer Untersuchungen solcher Verfahren spärlich. Insbesondere wenn es um nicht-technische Aspekte geht, sind den Autoren keine Untersuchungen bekannt, die Bezug auf solche nicht-wissenschaftliche Verfahren zu Selektionszwecken nehmen. Möglicherweise liegt dies daran, dass diese Verfahren von nicht-psychologischen Fachleuten respektive Verbänden erstellt worden sind; denn ganz im Gegensatz dazu sind die meisten Tests mit psychologischem Hintergrund für die Berufsberatung sehr wohl auf diese Aspekte untersucht worden (vgl. Arnold & Helbling, 2007). So wurde zwar über Sinn und Zweck dieser Verfahren in Zeitungsartikeln des Öfteren diskutiert, auch weil zum Teil für die Eltern der betroffenen Jugendlichen die finanziellen Kosten zu hoch sind. Allerdings wurde die subjektive Wahrnehmung der Tests und der Facetten, die zur Akzeptanz eines solchen Verfahrens gehören, selten oder praktisch nie untersucht. Gründe dafür sind vielfältig, ein wichtiger Grund könnte aber darin liegen, dass diese Berufseignungstests von einer Institution durchgeführt werden, die gar nichts oder nur wenig mit dem Bewerbungsprozess zu tun hat; somit muss keine Ausbildungsorganisation besorgt sein, dass ihr Image aufgrund einer unakzeptablen Selektionssituation Schaden nimmt.

Die bisherigen Ausführungen machen klar, dass die Akzeptanz- und Fairness-Beurteilung von Eignungstests von Seiten der getesteten Personen ein wichtiges Thema ist. Leider zeigt sich auch, dass in der Praxis die Evaluation von Verfahren hinsichtlich dieses Gütekriteriums nur selten ein Thema ist. Mit dieser Studie soll gezeigt werden, dass es geeignete und ökonomische Fragebogen gibt, um diesen Aspekt in der psychodiagnostischen Praxis zu erheben. Dabei wird der Fragebogen zur sozialen Validität von Klingner (2003) gemäss der Theorie von Schuler und Stehle (1983) in Kombination mit einem anderen Verfahren, der Selection Procedurale Justice Scale von Bauer, Truxillo, Sanchez, Ferrara, Craig und Campion (2001) auf dem Hintergrund der

Theorie zur prozeduralen Gerechtigkeit (Gilliland, 1993) untersucht. Die Studie soll dabei nicht nur einen wissenschaftlichen Nutzen erbringen, sondern auch den in der Praxis tätigen Personen und Organisationen nützen. Hierzu wird mit den eingesetzten Verfahren die Fairness und Akzeptanz des Multicheck gemessen, dem meistverwendeten Eignungstest in der schweizerischen Arbeitswelt, um entsprechende Verbesserungsempfehlungen zu machen. Zusammengefasst lauten die Ziele folgendermassen:

- 1) Es soll ein deutschsprachiges Instrument zur Untersuchung der Akzeptanz von Eignungstests für Auszubildende auf die Anwendbarkeit und die psychometrische Güte hin überprüft werden.
- 2) Es soll überprüft werden, inwiefern der in der Schweiz am meisten verwendete Eignungstest für Auszubildende von den entsprechenden Absolvierenden auf die Akzeptanz hin beurteilt wird.
- 3) Es soll aufgezeigt werden, wo Veränderungsbedarf in der Akzeptanz des Multicheck besteht und wie die Massnahmen zur Verbesserung aussehen könnten.

Peter Eberle hat vorwiegend die Theorie und die Methodik verfasst, sowie sich um das Layout gekümmert. Marcel Widmer hat die Beiträge zu den Ergebnissen und der Diskussion verfasst. Gemeinsam widmeten sie sich der Einleitung und den Hypothesen.

An dieser Stelle bedanken wir uns bei der Firma Multicheck für die Bereitstellung ihrer Daten zu diesen Studienzwecken.

## 2. Theorie

### 2.1 Gerechtigkeit

In der Psychologie wird das hypothetische Konstrukt der Gerechtigkeit als subjektive Bewertung einer Situation gesehen. Das Interesse der Gerechtigkeitsforschung lag ursprünglich in der gerechten Verteilung von Ressourcen. Dabei wurde die Gerechtigkeit als Vergleichsprozess zwischen zwei oder mehreren Personen definiert. Die Equity-Theorie, welche vorwiegend durch Studien von Adams (1965) begründet und von Walster, Berscheid und Walster (1973) weiterentwickelt wurde, beschreibt in der Psychologie die Verteilungsgerechtigkeit. Gemäss Thibaut und Walker (1975) bezog sich die psychologische Forschung nicht nur auf die Verteilung der Gerechtigkeit, sondern sie untersuchte auch das subjektive Erleben der Gerechtigkeit in andauernden Prozessen sowie in Interaktionen mit Autoritäten. Subsumiert wird diese Forschungsrichtung unter dem Ausdruck der prozeduralen Gerechtigkeit (Tyler & Lind, 1992). Es zeigte sich jedoch in neueren Untersuchungen, dass diese beiden Typen der Gerechtigkeit nicht genügen, um das komplexe Konstrukt der Gerechtigkeit adäquat zu beschreiben. Neben den genannten Konstrukten spielt auch der gerechte Umgang miteinander eine wichtige Rolle (Lupfer, Weeks, Doan & Houston, 2000). Ausserdem argumentieren Van den Bos, Vermunt und Wilke (1997), dass die prozedurale Gerechtigkeit und die Verteilungsgerechtigkeit eng miteinander verknüpft sind, obwohl sie in den bisherigen Forschungen als zwei unabhängige Faktoren betrachtet wurden. Gemäss vorliegenden Informationen erfolgte eine Beurteilung der Gerechtigkeit entweder aufgrund der distributiven Gerechtigkeit oder der prozeduralen Gerechtigkeit (Van den Bos et al., 1997). Zusätzlich spielt auch die Reihenfolge der gegebenen Informationen eine entscheidende Rolle. Werden als erstes Angaben zur distributiven Gerechtigkeit und erst in zweiten Schritt Angaben zur prozeduralen Gerechtigkeit gemacht, so werden die Angaben zur Verteilungsgerechtigkeit für das Gerechtigkeitsurteil stärker gewichtet. Wird die Reihenfolge umgekehrt, so werden die Daten zur prozeduralen Gerechtigkeit höher bewertet (Van den Bos et al., 1997).

Auch im organisationalen Kontext gibt es Literatur zur Prozeduralen Gerechtigkeit. Müller, Janssen und Jarzina (2007) kommen zum Schluss, dass die Bewertung der Prozeduralen Gerechtigkeit in starkem Masse die Einstellungen und das Verhalten von Personen gegenüber Organisationen beeinflusst. Dies gilt auch für die Bewertung von Personalauswahlverfahren. Personen, die ein Auswahlverfahren als fair einschätzen, bewerten die betreffende Organisation auch als attraktiveren Arbeitgeber. Die Ergebnisse

der Studie demonstrieren in Abhängigkeit von der Auswahlentscheidung unterschiedliche moderierende Einflüsse von Vorhersageunsicherheit auf den Zusammenhang zwischen prozeduraler Fairnessbewertung und organisationaler Attraktivität.

In der Literatur herrscht keine Übereinstimmung über die vielfältigen Dimensionen der Gerechtigkeit. Colquitt, Conlon, Wesson, Porter und Ng (2001) konnten in einer Metaanalyse bestätigen, dass sich das Konstrukt der organisationalen Gerechtigkeit aus distributiven, prozeduralen, informational sowie interpersonalen Bereichen zusammensetzt.

In einer anderen Metaanalyse von Cohen-Charash und Spector (2001) wird von drei Dimensionen der Gerechtigkeit ausgegangen. Gemäss ihrer Metanalyse ergeben sich distributive, prozedurale sowie interaktionale Gerechtigkeit. Die interpersonale und informale Gerechtigkeit fassen diese Autoren unter dem Begriff der interaktionalen Gerechtigkeit zusammen.

Da sich die Autoren der vorliegenden Studie vorwiegend mit der prozeduralen Gerechtigkeit befassen, wird diese als nächstes genauer betrachtet.

### 2.1.1 Prozedurale Gerechtigkeit

Die psychologischen Forschungen zur Verfahrensgerechtigkeit - auch prozedurale Gerechtigkeit genannt - beschäftigen sich mit der subjektiv wahrgenommenen, gerechten Behandlung während eines laufenden, sozialen Prozesses (Tyler & Lind, 1992). Bei der Gerechtigkeitsbeurteilung spielt das Resultat des Verfahrens eine weniger wichtige Rolle als bei Untersuchungen zur Verteilungsgerechtigkeit, welche sich mit der Verteilung von Ressourcen beschäftigen. Bei der Gerechtigkeitsbewertung spielt gemäss Thibaut und Walker (1975) die Kontrolle über das Verfahren und ihre Verteilung auf die involvierten Parteien und den Entscheidungsträger eine massgebende Rolle. Die Autoren unterscheiden zwischen der Entscheidungs- und Prozesskontrolle. Die Entscheidungskontrolle legt fest, was es zu entscheiden gibt. Eine hohe Prozesskontrolle bedeutet, das Verfahren steuern und festlegen, in welche Reihenfolge und in welche Komplexität die Argumente der Konfliktparteien gehören. Gemäss Theorie sind die Konfliktparteien jeweils äusserst bemüht, möglichst viel Kontrolle zu gewinnen. Gründe dafür sind einerseits, adäquat von der Autorität wahrgenommen zu werden, andererseits um ein unangenehmes Urteil für sich selbst abzuwenden. Gewisse Forscher postulierten, dass die Konfliktparteien die Entscheidungskontrolle der Prozesskontrolle vorziehen (u.a. Leventhal, 1976; Vidmar, 1990). Mittels empirischer Prüfung dieser Hypothese konnte aber das Gegenteil festgestellt werden (Lind, Lissak & Conlon, 1983). Die

Entscheidungskontrolle und die Prozesskontrolle wurden in dieser Studie experimentell manipuliert. Die Ergebnisse zeigten, dass das Verfahren nur in der Bedingung der hohen Prozesskontrolle (unabhängig von der Höhe der Entscheidungskontrolle) als gerecht beurteilt. Dies bedeutet, dass die Kontrolle über die Entscheidung selbst zweitrangig ist, wichtiger ist, wie man von Entscheidungsträgern behandelt wird.

Es gibt auch Theorien zur prozeduralen Gerechtigkeit, welche andere Faktoren als ausschlaggebend für die Bewertung der Gerechtigkeit des Verfahrens betrachten. Leventhal (1980) stellt fest, dass die Entscheidung von folgenden sechs Faktoren beeinflusst wird:

1. Konsistenz der verwendeten Regeln über die Zeit und über verschiedene Autoritäten
2. Unterdrückung von einseitigen Urteilen und Ausserachtlassen persönlicher, egoistischer Interessen der Autorität
3. Genauigkeit der Entscheidung unter Berücksichtigung aller Informationen und Standpunkte
4. Korrekturmöglichkeit bzw. Modifikation des Ergebnisses bzw. der Entscheidung
5. Repräsentativität der Entscheidung unter Einbeziehung aller Konfliktparteien und Personen
6. ethische Verträglichkeit der Entscheidung

Keine weitere Theorie, jedoch eine operationalisierte Form zur prozeduralen Gerechtigkeit stellt Gilliland (1993) vor; diese wird in Kapitel 4.1 (zum Fragebogen) genauer erläutert.

Tyler und Lind (1992) stellen fest, dass die bisherigen Theorien der prozeduralen Gerechtigkeit nur eine kurzfristige Beziehung zwischen der entscheidenden Autorität und den Konfliktparteien zulassen. Das Modell der Autoren geht davon aus, dass Menschen bestrebt sind, längerfristige Beziehungen zu Institutionen und Autoritäten zu pflegen. Die eingesetzten Verfahren sind Ausdruck gesellschaftlich vorherrschender Werte. Erst wenn Beziehungen zu Autoritäten positiv erlebt werden und das Gefühl entsteht, ein vollwertiges Mitglied der Gesellschaft zu sein, wird ein Verfahren als gerecht betrachtet. Sobald die Beziehung zu Autoritäten negativ erlebt wird, werde auch das Verfahren als ungerecht eingestuft.

Zur Bewertung des Verfahrens werden von Tyler (1989) die Faktoren Status, Neutralität und Vertrauen berücksichtigt.

Hinsichtlich des ersten Faktors stellt Festinger (1954) grundlegende Elemente der Sozialpsychologie dar, indem er beschreibt, dass Personen verschiedenen sozialen

Gruppen angehören. Um ihren Status in der Gruppe zu bestimmen, achten Personen ganz besonders auf Zeichen und Symbole, die Informationen über ihre Stellung in der Gruppe erschliessen. Definiert wird der Status dabei als Bewertung durch die Gruppenmitglieder. Zentral dabei ist, wie eine Person geschätzt wird und wie ihre Meinung in der Gruppe akzeptiert wird. Je positiver das Ergebnis ausfällt, desto mehr wird sich die Person als vollständiges Mitglied der Gruppe fühlen. Die Gruppenmitgliedschaft gibt auch Auskunft über die Person selbst. Beispielsweise werden eigene Werte und Einstellungen sowie das Verhalten mit Personen der Gruppe verglichen (Festinger, 1954). Um ihre eigene Identität positiv zu beeinflussen, sind die Mitglieder einer Gruppe stets bemüht, einen guten Status in der Gruppe einzunehmen. Die Gruppenmitglieder legen grossen Wert darauf gerecht behandelt zu werden, weil die dazu eingesetzten Verfahren zur Entscheidungsfindung wichtige Indikatoren für den eigenen Status in der Gruppe darstellen. Dabei spielen psychologische Verhaltensformen wie Freundlichkeit, Rechte eines Gruppenmitglieds, Respekt vor anderen Meinungen sowie ein würdevoller Umgang miteinander eine besonders wichtige Rolle. Das Gefühl der Zugehörigkeit und einen guten Status in der Gruppe inne zu haben wird bestärkt, falls jemand freundlich behandelt wird. Falls ein Mitglied unfreundlich oder despektierlich behandelt wird, impliziert dies, dass man nicht als vollwertiges Gruppenmitglied betrachtet wird. Eine Zurückweisung oder ein Gruppenausschluss löst negative Emotionen aus (Schachter, 1951).

Als zweiter Faktor zur Evaluation der prozeduralen Gerechtigkeit wird die Neutralität betrachtet. Diese zeigt sich darin, dass Fakten und nicht Meinungen zur Entscheidungsfindung herangezogen werden. Vorurteile der Autorität könnten zu einer diskriminierenden Entscheidung führen. Einzelne Gruppenvertreter sind der Meinung, dass die Autorität über ein anerkanntes Mitglied der Gruppe Bescheid wissen und zudem ihre echten Persönlichkeitseigenschaften kennen sollte. Vorurteile können somit ein Indikator sein, als nicht vollwertiges Mitglied der Gruppe akzeptiert zu sein (Tyler, 1989).

Tyler (1989) nennt die Vertrauenswürdigkeit der Autorität als dritten Faktor. Durch den Glauben an zukünftige Vorhaben der Autorität wird das Vertrauen verstärkt. Dabei gilt es zu unterscheiden zwischen dem tatsächlichen Verhalten und der Interpretation der Autorität. Beim tatsächlichen Verhalten wird unterschieden, ob Handlungsweisen ethisch vertretbar sind oder nicht. Bei der Interpretation kann beispielsweise festgestellt werden, ob die Schlussfolgerungen auf die zugrundeliegenden Motive verweisen. Heider (1958) beschreibt in seinen Attributionstheorien, dass vom Verhalten einer Person auf dahinterliegende, nicht beobachtbare Persönlichkeitseigenschaften geschlossen wird. Immer wieder tritt in der Praxis ein möglicher Widerspruch auf, welcher sich mit diesem

Ansatz lösen lässt. In einer Untersuchung fand Tyler (1989), dass obwohl Personen ein gesamtes Verfahren als fair und gerecht erlebten, obwohl sie einzelne, unehrliche Verhaltensweisen von Polizisten oder Richter anprangerten. Weil eine Interaktion in der Zukunft als wahrscheinlich angenommen wird, spielt die Vertrauenswürdigkeit dabei eine zentrale Rolle zwischen der Autorität und den Konfliktparteien. Die Vertrauenswürdigkeit dient als Ausgangspunkt, wie sich eine Autorität in Zukunft verhält. Es stellt sich die Frage, ob man gerecht behandelt wird oder nicht.

Experimentell konnten Lind, Kanfer und Earley (1990) die nicht-instrumentelle Funktion der Prozesskontrolle nachweisen. In einer Arbeitsgruppensituation konnten die Teilnehmenden entweder vor oder nach der Zielvereinbarung ihre Meinung über das festzulegende bzw. festgelegte Ziel äussern. Ihnen wurde in einer Nachentscheidungs-Bedingung mitgeteilt, der Versuchsleiter habe Interesse an ihrer Meinung über das Ziel, auch wenn dieses bereits feststand und nicht mehr geändert werden konnte. Die Alternativ-Bedingung bestand darin, dass die Teilnehmenden gebeten wurden, vor der Zielfestlegung ihre Meinung kundzutun. In Kontrollbedingung gab es gar keine Möglichkeit sich über das Ziel zu äussern. Die Ergebnisse zeigten deutlich, dass Teilnehmende, welche die Option bekamen ihre Meinung zu äussern, das Verfahren respektive die Zielvereinbarung als gerechter beurteilten als jene, die keine Möglichkeit zur Meinungsäußerung erhielt.

In einer anderen Studie untersuchte Tyler (1987) die Beurteilung der Gerechtigkeit bezüglich ihres Einflusses auf die Prozesskontrolle. Es wurden vier Hypothesen überprüft: Der Einfluss sollte abnehmen, wenn (1) die Autorität als voreingenommen gilt, (2) oder nicht im Sinne einer Mediation zwischen den sich streitenden Gruppen nach einer guten Lösung sucht, (3) oder die Meinungen der Konfliktparteien einholt, aber für die Entscheidung nicht berücksichtigt, (4) oder die Resultate sehr wichtig sind. Die Resultate zeigten, dass nur Bedingung drei einen Einfluss auf die Gerechtigkeitsbeurteilung hatte.

Es kann zusammenfassend festgestellt werden, dass die Möglichkeit, seine persönliche Meinung zu äussern, ein zentraler Faktor für die Gerechtigkeitsbeurteilung ist. Die Studien legen nahe, dass es nicht relevant ist, ob eine solche Äusserung vor oder nach der Entscheidung von Autoritäten geschieht.

In einer Untersuchung von Rasinski und Tyler (1987) konnten die Autoren belegen, dass die subjektiv wahrgenommene Verfahrensgerechtigkeit einen Einfluss auf Wahlverhalten hat. Die wahrgenommene Gerechtigkeit korrelierte dabei signifikant mit dem Wahlverhalten. Ein Präsidentschaftskandidat wurde umso eher gewählt, je höher seine Gerechtigkeitsorientierung eingeschätzt wurde. In einer Folgeuntersuchung wurde der

Zusammenhang zwischen Gerechtigkeitsorientierung und politischer Teilnahme in der US Stadt Chicago bei 400 Bürger erhoben. Dabei wurde das politische Engagement durch das Tragen von Wahlkampf-Aufklebern, die Teilnahme an Wahlveranstaltungen, die persönliche Kontaktaufnahme mit Politikern sowie die Teilnahme an regionalen Wahlveranstaltungen operationalisiert. Die Resultate der Untersuchung zeigten einen signifikanten Einfluss der prozedurale Gerechtigkeit auf die politischen Interessen. Demgegenüber hatte die distributive Gerechtigkeit keinen Einfluss auf die politische Anteilnahme.

## 2.2 Soziale Validität

Schuler und Stehle (1983) stellen das Unbehagen an der psychologischen Eignungsdiagnostik in den Vordergrund. Sie legen dar, dass die Eignungsdiagnostik im Öffentlichkeitsbild und in der Praxis einen nicht unbestrittenen Ruf hat. Aus eigener Erfahrung und die ihrer Kollegen stellen sie fest, dass selbst Psychologen an der Relevanz und der Fairness ihrer Tests zweifeln. Sie monieren, dass insbesondere der wichtigste und wohl kritikfähigste Punkt das Unbehagen der Testabsolventen in der Testsituation ist. Dabei charakterisieren verschiedene Punkte dieses Unbehagen: Die Autonomie wird verletzt, die Situation ist durch ein Macht-Gefälle charakterisiert, die Verhaltensmöglichkeiten werden eingeschränkt, Bewertungsangst und Belastung herrschen. Zudem bewegt die fehlende Kenntnis der Prinzipien des diagnostischen Schlusses die Probanden dazu, ein „impressions management“ zu betreiben. Insgesamt ist auch nicht klar, inwiefern die Persönlichkeit geschützt oder verletzt wird. Zu einem ähnlichen Ergebnis kommen auch Fruhner, Schuler, Funke und Moser (1991). Bei der Bewertung von Personalauswahlverfahren stellten sie fest, dass beispielsweise Vorstellungsgespräche in allen Aspekten der Sozialen Validität (Ergebniskommunikation, Transparenz, Information und Kontrollierbarkeit) besser beurteilt wurden als psychologische Tests.

Allerdings erwähnen Schuler und Stehle (1983) auch Faktoren, die diese ungleiche Situation auch positiv beeinflussen können: So sprechen Freiwilligkeit der Testung, Einsicht in die Nützlichkeit des Tests, Klarheit über die Rolle der Diagnostik und des Diagnostikers, Informiertheit und Erwartungen des Klienten für eine angenehme Testsituation. Es ist aber auch klar, dass diese Faktoren nicht immer und in gleichem Masse in jeder Testsituation oder Personalauswahl erfüllt werden können. Deshalb liegt es gemäss Schuler und Stehle (1983) an der Psychologie, diese Faktoren zu identifizieren

und sie so zu verändern, dass das Unbehagen reduziert wird. Anders als „technisch-empirische“ Aspekte sollen diese Faktoren das Erleben der Personen in den Testsituationen untersuchen, um Verbesserungsmöglichkeiten zu identifizieren. Den Bereich der resultierenden Prinzipien bezeichnen Schuler und Stehle (1983) als Soziale Validität. Die soziale Validität soll die eignungsdiagnostische Situation „zu einer akzeptablen sozialen Situation“ machen und der „technisch-empirischen“ Validität zur Seite gestellt werden. Sie soll betonen, dass Diagnostik allemal in einem sozialen Kontext stattfindet und dass die Fortentwicklung psychologischer Methoden auch eine Anpassung an die Veränderung sozialer Realitäten zu sein hat.

Schuler und Stehle (1983) unterstellen der sozialen Validität vier Komponenten, namentlich Information, Partizipation, Transparenz und Kommunikation. Information wird als „Berücksichtigung sozialpsychologischer Anforderungen“ und sozialer Fertigkeiten definiert, wie sie in Anforderungsanalysen bestimmt werden kann. Wichtig ist die Berücksichtigung relevanter Informationen über die Organisation und deren Klima sowie den Führungs- und Verhaltensstil der Mitarbeiter. Während der beruflichen Grundbildung ist darunter aber vor allem der Aspekt des Kennenlernens von beruflichen Tätigkeiten und Richtungen zu verstehen.

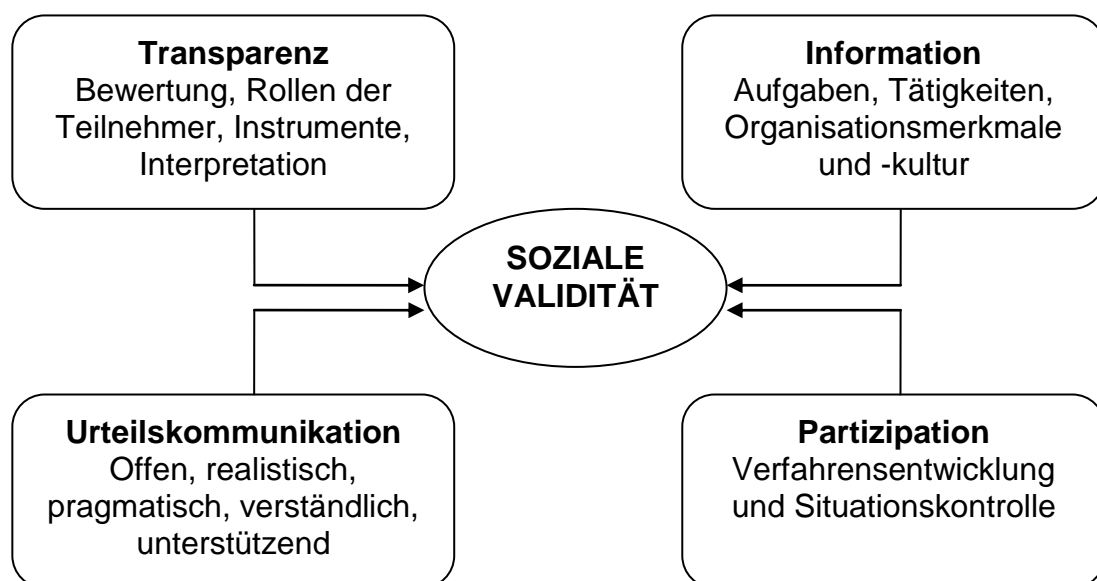


Abbildung 1: Konzept der Sozialen Validität (Schuler, 1990, S. 186)

Unter Partizipation verstehen die Autoren die Beteiligung der eignungsdiagnostisch untersuchten oder sonst wie tangierten sozialen Elemente. Der Einbezug der Zielgruppe für die Gestaltung des Auswahlverfahrens geschieht im gesamten Prozess von der Aufgabenanalyse bis zur Formulierung der Empfehlung. Schuler und Stehle (1983) erwähnen, dass die Partizipationsmöglichkeit unterschiedlich ist. So können die

betroffenen Personen natürlich fast oder gar nicht zur Mitgestaltung aktiviert werden, wenn sie direkt in der Testsituation sind.

Als sehr wichtiges Element bezeichnen Schuler und Stehle (1983) die Transparenz. Sie unterscheiden die Transparenz „als Erlebnisraum der diagnostischen Situation“ und die Transparenz der „vorgenommenen Bewertungen“. Ersteres bezeichnet die Bewertung der Relevanz der verwendeten Verfahren und der Durchführung des gesamten Auswahlverfahrens, wobei situative Elemente wie Beobachtungs- und Beurteilungsgelegenheiten die Wahrnehmung beeinflussen. Letzteres bezeichnet die Durchschaubarkeit der Auswertung und Urteilsaggregation über die Urteilkriterien und die Prinzipien des diagnostischen Schlusses. Eine Möglichkeit zur Selbstbeurteilung oder zum sozialen Vergleich beurteilen die Autoren als positiv für die Transparenz in einer eignungsdiagnostischen Situation. Aber auch die Schaffung einer angenehmen Atmosphäre kann hier sicher beispielhaft angegeben werden (vgl. Köchling & Körner, 1996).

Als letztes Element definieren die Autoren die Kommunikation im Sinne des wechselseitigen Austausches von Informationen von Bewerber und Organisation. Wechselseitig heisst, dass nicht nur Informationen aus der Testsituation über den Bewerber gesammelt werden, sondern dass auch Informationen über das Unternehmen zum Bewerber kommuniziert werden. Des Weiteren muss dieser Austausch in einer günstigen Form erfolgen; diese ist dann gegeben, wenn sich die Beurteilung der Ergebnisse auf die Schilderung von Verhaltensweisen beschränkt und Eigenschaftszuschreibungen auf der Verhaltensebene erfolgen. Ziel ist es, die Persönlichkeitssphäre durch die Wahrung von Distanz und Beschränkung mittels diagnostischer Bemühungen auf arbeitsrelevante Merkmale zu schützen.

Das Konzept der sozialen Validität und deren Untersuchung wurde auch kritisch begutachtet; so findet Kersting (1998), dass bereits der Name „soziale Validität“ aus Verständnisgründen in „soziale Akzeptanz“ unbenannt werden sollte. Zudem gibt er an, dass diese Validität nur ungenügend operationalisiert worden ist, weil das mehrdimensionale Konzept zuerst in ein Instrument umgesetzt und dann auf seine „technische“ Validität überprüft werden muss. In der Mehrzahl der empirischen Studien sei dies nicht vorgesehen gewesen. Zudem gelte es, differentialpsychologische Aspekte einzubeziehen. So könnte beispielsweise das Alter oder die Leistung in einem

Eignungstest einen erheblichen Einfluss auf die vermeintlich hohe oder geringe Akzeptanz einzelner Verfahren haben. Gemäss Kersting (1998) lassen sich auch verfahrensgruppenspezifische Schlussfolgerungen ableiten, die helfen, das Konzept der sozialen Validität in die Praxis umzusetzen.

Schuler (1990) stellt fest, dass die Soziale Validität kein Konzept für einzelne Untersuchungen, sondern für ein sich entwickelndes Forschungsprogramm darstellt. Er führt aus, dass das Konzept (noch) kein überprüftes Modell sei, sondern lediglich als Heuristik verstanden werden soll. Diese Aussage besitzt trotz weiteren Untersuchungen seit den 90er Jahren noch ihre Gültigkeit.

In einer aktuellen Untersuchung (Schuler, Hell, Trapmann, Schaar & Boramir, 2007) zur Nutzung psychologischer Verfahren externer Personalauswahl wurde die Akzeptanz – und nicht nur die Validität – bereits in die Untersuchung integriert. Bereits diese Tatsache zeigt deutlich, dass die Wahrnehmung einiger Aspekte der Sozialen Validität immer mehr an Wichtigkeit gewinnt. Bei der besagten Untersuchung zeigte sich klar, dass bezüglich Akzeptanz Bewerbungsunterlagen und diverse Arten von Bewerbungsgesprächen deutlich besser als psychologische Tests sowie graphologische Gutachten bewertet werden. Die Validität wird dort besonders hoch eingeschätzt, wo Verhaltensbeobachtungen erfolgen können, aktuelle oder künftige Vorgesetzte involviert sind und wo es dem Beobachter möglich ist, ein gewisses Mass an Situationskontrolle auszuüben. Gemäss Schuler et al. (2007) sorgt dies schlussendlich auch für eine günstige Bewertung der Akzeptanz des Verfahrens durch die Bewerber.

### 3. Hypothesen

Die Fragestellungen beziehen sich auf drei Themen: 1) Es soll überprüft werden, inwiefern die Reliabilität der eingesetzten Verfahren den Werten aus anderen Ländern und in anderen Settings entspricht. 2) Es ist von Interesse, wie die konvergente und divergente Validität der zwei eingesetzten Verfahren (Fragebogen zur sozialen Validität von Klingner, 2003, und die Selection Procedural Justice Skale SPJS) ausfallen. 3) Es soll untersucht werden, wie die Teilnehmenden den Multicheck als Auswahlverfahren beurteilen. Dabei ist sowohl an Vergleiche innerhalb der verwendeten Verfahren (über die einzelnen Skalen) als auch Vergleiche mit Werten aus anderen Studien zu denken.

#### 3.1 Reliabilität

Es wird angenommen, dass die eingesetzten Fragebogen aus den Studien zur sozialen Validität (Klingner, 2003) und zur prozeduralen Gerechtigkeit (Müller et al., 2003) eine hohe Reliabilität aufweisen, weil in den entsprechenden Studien diese entweder bereits überprüft wurde (Müller et al., 2007) oder das Instrument auf dem Hintergrund einer Faktorenanalyse (Klingner, 2003) konzipiert worden ist. Da das Instrument der Sozialen Validität neu und mit einer geringen Itemanzahl bedacht ist, ist nicht sicher, dass eine hohe Reliabilität für alle Teilskalen resultiert. Entsprechend den Angaben von Bortz (2003, S. 199) soll für eine hohe Reliabilität ein Koeffizient von  $.80$  gültig sein, damit von einem guten Test gesprochen werden kann.

Hypothesen:

- 3.1.1 Die Reliabilität des Fragebogens von Klingner (2003) ist als gut ( $>.80$ ) zu bezeichnen.
- 3.1.2 Die Reliabilität der SPJS ist als gut ( $>.80$ ) zu bezeichnen.
- 3.1.3 Die Reliabilität aller Subskalen der SPJS und des Fragebogens von Klingner (2003) sind mindestens als genügend ( $>.70$ ) zu bezeichnen.

#### 3.2 Konvergente Validität

Zwischen Teilskalen und insbesondere den beiden übergeordneten Konstrukten „Soziale Validität“ und „Prozedurale Gerechtigkeit“ sollen Zusammenhänge bestehen. Die bisherigen Erfahrungen mit den Verfahren zur Sozialen Validität und der Prozeduralen

Gerechtigkeit zeigen aber, dass es sogar schwierig ist, einzelne Facetten auseinander zuhalten (z. B. Müller et al., 2007). So dürfen, um den Einsatz von Instrumenten mit Facetten zu gerechtfertigen, die Teilskalen (Facetten) nicht zu hoch miteinander korrelieren ( $< .70$ ), damit die Skalen eigenständige Konstrukte darstellen (siehe unten / Hypothese 5.3). Auf Grundlage der Items und der von den Autoren intendierten Namensgebung einzelner Facetten gibt es allerdings einige Skalen, die mit anderen Skalen des anderen Instrumentes relativ hoch (mindestens  $.50$ ) korrelieren sollten (z. B. Feedback der Sozialen Validität mit Feedback der Prozeduralen Gerechtigkeit), mindestens aber höher, als es un-ähnliche Skalen über die Verfahren tun. Die Regeln der Zuordnung, auf die sich die Hypothesen stützen, finden sich bei Kersting (1998), wobei wir uns bei der sozialen Validität auf die Facetten von Klingner (2003) stützen. In derselben Annahme wird davon ausgegangen, dass die Soziale Validität und die Prozedurale Gerechtigkeit jeweils als Gesamtes Ähnliches messen und deshalb hoch miteinander korrelieren (vgl. die Gegenüberstellung beider Konstrukte durch Kersting, 1998). Dies gilt auch für die Fairness, welche als subjektives Mass für die Akzeptanz ebenfalls hoch (mindestens  $.50$ ) mit beiden anderen Werten zusammenhängen sollte.

Hypothesen (alle postulierten Zusammenhänge = grosser positiver Effekt  $.50$ ):

- 3.2.1 Die Skalen Information und Feedback von Klingner (2003) korrelieren mit den Skalen „Information“, „Feedback“ sowie „Ehrlichkeit und Offenheit“ der SPJS.
- 3.2.2 Die Skalen Transparenz und Kontrolle von Klingner (2003) korrelieren mit den Skalen „Berufsbezogenheit (Inhalt und Vorhersage)“, „Möglichkeit der Leistungsdarstellung“ sowie „Konsistenz“ der SPJS.
- 3.2.3 Die Interkorrelationen der Gesamtskalen „Prozedurale Gerechtigkeit“ und „Soziale Validität“ sowie dem Wert der subjektiv beurteilten Validität (Fairness) entsprechen einem hohen Zusammenhang.

### 3.3 Diskriminante Validität

Zusammenhänge sollen nur dann bestehen, wenn ein Konstrukt theoretisch mit einem anderen Ähnlichkeit hat. Es gibt aber Variablen, für die auf theoretischer Basis keine Konvergenz vorgesehen ist. Für die Diskriminanz mit soziodemografischen Variablen konnten bisher nur geringe Zusammenhänge aufgezeigt werden (s. o.). Es wird nicht angenommen, dass die Anzahl absolvierter Eignungstests in Zusammenhang mit der

Akzeptanz steht. Auch zu psychologischen Variablen wie der Intelligenz konnten die Autoren in der gesichteten Literatur ebenfalls keine bedeutenden Zusammenhänge finden. Deshalb prognostizieren wir keine Korrelationen mit der Testleistung. Zusätzlich wird angenommen, dass die Konvergenz einzelner Skalen innerhalb der beiden verwendeten Instrumente nicht über .70 liegen sollte, um eine Redundanz der Skalen zu vermeiden. So hat Klingner (2003) bei der Faktorenanalyse festgestellt, dass Information und Transparenz keine separaten Faktoren sind. Auch Müller et al. (2007) bekunden Schwierigkeiten, einzelne Facetten der Prozeduralen Gerechtigkeit zu finden.

- 3.3.1 Sowohl die Skala „Prozedurale Gerechtigkeit“ als auch die Skala „Soziale Validität“ korrelieren nicht mit dem Geschlecht resp. dem Alter.
- 3.3.2 Sowohl die Skala „Prozedurale Gerechtigkeit“ als auch die Skala „Soziale Validität“ korrelieren nicht mit der Testerfahrung.
- 3.3.3 Sowohl die Skala „Prozedurale Gerechtigkeit“ als auch die Skala „Soziale Validität“ korrelieren nicht mit dem Resultat im Multicheck.
- 3.3.4 Die Interkorrelationen der Subskalen innerhalb eines Instrumentes sind nicht grösser als .70.

3.4 Vergleich der Mittelwerte der Teil-Skalenwerte der sozialen Validität und der prozeduralen Gerechtigkeit dieser Studie (zum Multicheck) mit den Werten aus anderen Studien

In dieser Hypothese wird die Gegenüberstellung zwischen dem Multicheck und anderen Verfahren vorgenommen, wobei nur Vergleiche möglich sind, wo die nötigen Daten aus den Studien zur Verfügung standen oder auf Anfrage der Autoren eingetroffen sind (für die Teilskalen der Prozeduralen Gerechtigkeit in der Studie von Müller et al., 2007). So werden die Ergebnisse zur Sozialen Validität des Multicheck mit den Resultaten von Klingner (2003) verglichen und es wird eine geringere Akzeptanz erwartet, weil die AZUBI-BK Ähnlichkeit mit Arbeitsproben hat (welche besser abschneiden). Die Werte zur Prozeduralen Gerechtigkeit vergleichen wir mit den Ergebnissen von Müller et al. (2007), die den Fragebogen Absolvierenden von Assessment Centers vorlegen. Hier werden aufgrund des hoch akzeptierten Verfahrens weitgehend niedrigere Werte für den Multicheck erwartet. Zur Kontrolle werden die Daten von Klingner (2003) mit dem Multicheck und einem „klassischen“ Intelligenztest, dem BIS-4 verglichen. Dabei werden keine signifikanten Unterschiede erwartet.

Hypothesen:

- 3.4.1 Die Facetten und Items der Sozialen Validität des Multicheck werden signifikant geringer beurteilt als die der AZUBI-BK (gemäss Klingner, 2003).
- 3.4.2 Die Prozedurale Gerechtigkeit und deren Facetten werden im Multicheck signifikant geringer beurteilt als das Assessment Center (Müller et al., 2007).
- 3.4.3 Es besteht kein Unterschied zwischen dem Multicheck und dem BIS-4 bezüglich der Facetten und Items zur Sozialen Validität.

## 4. Methode

### 4.1 Fragebogen

Als Basis für die in dieser Studie verwendeten Tests zur Akzeptanz wurden Verfahren eingesetzt, die entweder bereits mehrfach in der anglo-amerikanischen oder deutschsprachigen Wissenschaft verwendet worden sind oder im Rahmen einzelner Untersuchungen neu konstruiert worden sind. Ersteres gilt für die *selection procedural justice scale* (Bauer, 2001), deren Anpassung an die Studie im nächsten Abschnitt aufgezeigt wird, letzteres gilt für den Fragebogen zur Sozialen Validität, welcher in Anlehnung an Klingner (2003) leicht verändert wurde. Das Instrument stellt eine Operationalisierung der Theorie zur Sozialen Validität von Schuler & Stehle (1983) resp. Schuler (1990) dar. Klingner (2003) macht mit jeweils zwei Fragen zu den vier Facetten der Sozialen Validität, nämlich Information, Transparenz, Feedback und Kontrolle, ein relativ „schlankes“ Instrument, das sehr wohl für den Einsatz in der Praxis geeignet ist. Die acht Items, welche Klingner (2003) benutzt, wurden übernommen aber teilweise marginal sprachlich modifiziert. Beispielsweise die Frage „Die einzelnen Arbeitsproben sind praxisnah“ wird zu „Die einzelnen Testaufgaben sind praxisnah“ angepasst. In derselben Weise wurden zudem die Fragen zur Fairness, welche nach Klingner (2003) die subjektive Wahrnehmung der Akzeptanz des Tests durch die Teilnehmenden messen sollen übernommen. Die Aussagen können jeweils von 1 = „trifft nicht zu“ bis 5 = „trifft zu“ bewertet werden.

Die „selection procedural justice scale“ (SPJS) von Bauer (2001) stellt die operationalisierte Form der Prozeduralen Gerechtigkeit von Gilliland (1993) dar. Dieses Instrument besteht in der englischen Originalversion aus 11 Skalen und 39 Items, die diese Form der Akzeptanz in eignungsdiagnostischen Situationen erheben sollen. Müller et al. (2007) übersetzten den Fragebogen von Bauer (2001) ins Deutsche und verwendeten ihn erstmals in einer Untersuchung. Das Verfahren besteht aus neun Skalen: Berufsbezogenheit Vorhersage, Information, Möglichkeit der Leistungsdarstellung, Ehrlichkeit und Offenheit, Konsistenz, Behandlung, Kommunikation, Angemessenheit der Fragen und Berufsbezogenheit Inhalt. Diese Skalen wurden übernommen und einige Anpassungen an die jetzige Studie wurden vorgenommen. Allgemein beziehen sich diese Änderungen nur auf die Rahmenbedingungen dieser Studie oder sind dem Verständnis halber vorgenommen worden. Zusätzlich wurde eine Skala von Bauer (2001)

übernommen, weil der Bezug zu den Skalen der Sozialen Validität interessierte: die Skala „Feedback“. Diese wurden frei übersetzt. Alle übersetzten Items und Fragen finden sich im Anhang. Dort finden sich ebenfalls alle zusätzlichen Fragen (demografische Daten zum selbstberichteten Testergebnis oder der angegebenen Testerfahrung).

Insgesamt beinhaltet der eingesetzte Fragebogen 44 (34+10) Items zu 15 (10+4+1) Skalen. Bewertungsformat ist jeweils eine siebenstufige Likert-Skala (1 = stimme gar nicht zu, 7 = stimme voll zu). Wichtig ist folgende Überlegung zu unserem Bewertungsformat: Da die Mittelwerte der Verfahren aus unserer Studie jeweils mit den Werten aus den ursprünglichen Untersuchungen zu vergleichen sind, wurde das Antwortskalen-Format wie im Original beibehalten. Die durchschnittliche Verweildauer der beantwortenden Personen (also auch solche, die den Fragebogen frühzeitig beendet haben) liegt bei ca. 13 Minuten. Die Beantwortung der Items zur Prozeduralen Gerechtigkeit dürfte dabei ungleich länger gedauert haben als bei allen anderen Fragen zusammen.

## 4.2 Datenerhebung

Die Datenerhebung erfolgte, nachdem die Teilnehmenden die Tests bei der Firma Multicheck absolviert haben. Weil die Daten aufgrund von zeitlichen Bedingungen nicht direkt nach der Testung erhoben werden konnten, musste dies (anschliessend) in einem eigenständigen Schritt gemacht werden. Aufgrund der Tatsache, dass die Testungen zu unterschiedlichen Zeitpunkten stattfanden, wurde im Abstand von circa einem Monat jeweils eine neue Online-Umfrage gestartet und den Teilnehmenden, die den Test zu diesem Zeitpunkt absolviert haben, via e-Mail eine Aufforderung zur Teilnahme gesendet. Damit alle Teilnehmenden die Testergebnisse erhalten haben, wurden die 10 Tage abgewartet, die die Firma Multicheck im Rahmen des Qualitätsmanagement als Zeitrahmen für das Eintreffen der Testergebnisse bei den Teilnehmenden angibt. Die Fragebogen selber konnten die Teilnehmenden jeweils zwei Wochen lang beantworten, so dass die Umfrage frühestens nach 10 Tagen und spätestens nach ca. zwei Monaten von den Testabsolvierenden beantwortet wurde. Die Umfrage fand zwischen November 2008 und Februar 2009 statt, wobei es aus Organisationsgründen nicht möglich war, die SPJS bereits im Jahr 2008 einzusetzen.

Die Universität Basel stellt ein e-Research-System (eRes) zu Verfügung, mit dem man formular-basierte Online-Befragungen und Experimente erstellen und durchführen kann.

Das System stellt eine einfache Möglichkeit dar, eine Umfrage für wissenschaftliche Zwecke online selbst zu erstellen und auszuwerten. Die Autoren haben sich für dieses Online-System entschieden, weil von den Teilnehmenden die e-Mail-Adressen zur Verfügung standen und damit eine einfache Möglichkeit gegeben war, möglichst viele Testabsolventen der Firma Multicheck zu erreichen. Zudem hat das System weitere Vorteile, wie z. B. Hinweise zur Rücklaufquote und automatische Archivierung der Daten.

Die Rücklaufquote der „beginner“ (Verhältnis der Teilnehmenden, die mit der Umfrage angefangen haben, zu denen, die angefragt wurden) beträgt 23 Prozent. Die Rücklaufquote der „completer“ (Verhältnis der Teilnehmenden, die die Umfrage beendet haben, zu denen, die angefragt wurden) ist demgegenüber mit 11 Prozent deutlich geringer. Es wird davon ausgegangen, dass dies ein Artefakt der Online-Erhebung ist und auch mit der Redundanz der Fragen zu tun hat (Phänomen der „Ausmergelung“). Aufgrund der Tatsache, dass der Fragebogen zur Prozeduralen Gerechtigkeit erst später eingesetzt wurde, wurde für die Soziale Validität eine deutlich höhere Stichprobe erreicht. Grund dafür ist die Anpassung des Fragebogens in Anlehnung an Müller et al. (2007; s. o.). Insgesamt konnte in etwa eine äquivalente Stichprobengrösse erreicht werden, wie es die Autoren für die Mittelwertsvergleiche in ihren Akzeptanzüberprüfungen der herbeigezogenen Instrumente erreicht haben.

#### 4.3 Stichprobe

Die Teilnehmenden der Untersuchung stellen alles Absolvierende der Eignungstests der Firma Multicheck dar, die sich im Hinblick auf das Erhalten einer Lehrstelle bei einer Firma bewerben müssen. In Bezug auf die Repräsentativität wurden alle für den Abschnitt Ende Jahr 2008 und Anfang Jahr 2009 angemeldeten Testkandidaten um eine Teilnahme gebeten. Die Eignungstests der Firma Multicheck sind berufsabhängig, weshalb die Teilnehmenden je nach Ausbildungs- resp. Berufsziel einen unterschiedlichen Test für die Branche Kaufmännische, Technische oder Gewerbliche Branche, Detailhandel, Gesundheit oder Schönheit absolvieren müssen. Implizit gilt also die Voraussetzung, dass die Berufswahl (oder zumindest die Branchen-Wahl) erfolgt ist und sich die Personen für einen entsprechenden Beruf bewerben wollen (mit Ausnahme derer, die den Test der Erfahrung halber machen).

Die Stichprobe kann wie folgt beschrieben werden (für die „completer“): In der Stichprobe haben 783 (47 Prozent) der Personen einen Kaufmännischen, 314 (19 Prozent) einen Technischen, 120 (7 Prozent) einen Gewerblichen Multicheck, 296 (18 Prozent) einen Test für den Detailhandel, 150 (9 Prozent) einen Test für den Bereich Gesundheit und 7 (ca. 1 Prozent) für den Bereich Schönheit absolviert. Ein Grossteil der Teilnehmenden (69 Prozent) hat nur diesen einen Multicheck absolviert, 31 Prozent haben bereits einen (25 Prozent) oder mehrere andere Eignungstests (6 Prozent) gemacht. 17 Prozent sind 14 Jahre oder jünger, 46 Prozent sind genau 15 Jahre, 21 Prozent der Jugendlichen sind 16 Jahre alt, und nur 6 Prozent sind 18 Jahre oder älter. Der Mittelwert und Median des Alters betragen beide 15,7 Jahre (Standardabweichung des Mittelwertes 2.7 Jahre). 58 Prozent der Teilnehmenden sind weiblichen, 42 Prozent sind männlichen Geschlechts. Insgesamt haben 2'170 Personen den Fragebogen begonnen, wovon leider nur ein Teil beide Fragebogen beantworten hat.

#### 4.4 Datenauswertung

Für die Soziale Validität wurden fünf Skalen zu je zwei Items und für die Prozedurale Gerechtigkeit zehn Skalen mit insgesamt 34 Items in die Berechnungen miteinbezogen. Diese Variablen sind alle intervall-skaliert, hinsichtlich grafischer Auswertung (Sichtung) der Daten normalverteilt und können dementsprechend parametrisch ausgewertet werden. Die Zusammenfassung der Skalen erfolgt anhand der Zuordnungen, wie sie in den zur Konstruktion beigezogenen Studien vorgenommen wurde. Eine Ausnahme stellt die Transformation der Items (ohne die Skala „Fairness“) zur Gesamtskala der Sozialen Validität dar, was Klingner (2003) in ihrer Studie unterlässt. Es ist das Ziel die Bedeutung einzelner Facetten (Aspekte) der Akzeptanz-Konstrukte im Vergleich mit einem Gesamtwert zu überprüfen. Die Reliabilität der Skalen wurde mit Cronbachs Alpha überprüft.

Zur Auswertung der Hypothese 3 wurden folgende Variablen hinzugenommen: Geschlecht, Alter, Testleistung (ein normiertes Aggregat einzelner Testergebnisse) und Testerfahrung (Anzahl bereits absolvierter Eignungstests). Diese Variablen weisen generell kein verteilungsfreies Format auf und sind nur zum Teil intervallskaliert; so ist zum Beispiel die selbstberichtete Test-Leistung prozentrang-normiert, weil diese von der Firma Multicheck so kommuniziert wird. Für Auswertungen, welche diese Variablen mit einbeziehen (Diskriminante Validität, siehe Kap. 5.3), wurden nicht-parametrische

Verfahren (Spearman statt Pearson-Korrelationen) benutzt. Ansonsten wurden parametrische Auswertungsmethoden wie Produkt-Moment-Korrelationen (Pearson) und t-Tests für unabhängige Stichproben verwendet, da die Voraussetzungen dafür erfüllt sind (vgl. Bortz, 1999, S. 138 resp. 204).

Für die Mittelwertsvergleiche ist es nötig, die aus den Stichproben erwarteten Varianzen zu vergleichen, um bei unterschiedlichen Varianzschätzungen eine Korrektur vorzunehmen (gemäss Clauss & Ebner, 1971; vgl. Bortz, 1999, S. 138). Wir haben diese überprüft, in den Ergebnissen aber nicht aufgelistet, da sich im Signifikanzniveau der Resultate keine Differenzen ergeben haben. Es wurde unterlassen, die t-Tests für die nach dem Multicheck-Typ (also nach Berufsbranchen) getrennten Stichproben vorzunehmen (wegen der in diesem Fall zu geringen Stichprobengrössen). Die Auswertungen wurden generell mit SPSS (Version 15.0) vorgenommen, mit Ausnahme der t-Tests, weil nur die geschätzten Populationsvarianzen und keine Datensätze aus den relevanten Untersuchungen (Klingner, 2003, sowie Müller et al., 2007) zur Verfügung standen. Diese Berechnungen wurden mit Microsoft Excel (Version 2003) durchgeführt, wobei die vereinfachte Formel für den Standardfehler gemäss Bortz (1999, S. 138, Formel 5.13) genutzt und dann die t-Werte berechnet wurden, indem die Differenzen der Mittelwerte an den berechneten Standardwerten relativieren wurden.

### **Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben (t-Test)**

Der t-Test für unabhängige Stichproben untersucht die Nullhypothese, dass die Mittelwerte zweier Stichproben aus Populationen stammen, deren Parameter  $\mu_1$  und  $\mu_2$  identisch sind. Formalisiert bedeutet dies für die  $H_0$  und die  $H_1$ :

$$H_0: \mu_1 - \mu_2 = 0$$

und

$$H_1: \mu_1 - \mu_2 \neq 0$$

Bortz (2001) gibt neben der Gleichung, die sich auf die Berechnung vorhandener Daten bezieht, auch eine Gleichung an, welche verwendet werden kann, wenn nur die Mittelwerte, Varianzen und Stichprobengrössen der zu vergleichenden Stichprobendaten vorhanden sind (S. 137). Diese Gleichung für den Standardfehler der Differenz lautet:

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{(n_1 - 1) \cdot \hat{\sigma}_1^2 + (n_2 - 1) \cdot \hat{\sigma}_2^2}{(n_1 - 1) + (n_2 - 1)}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Dabei bezeichnen die  $\sigma$ -Werte in der Formel die geschätzten Populationsvarianzen und die  $n$ -Werte die Stichprobengröße zu den beiden zu vergleichenden Mittelwerten.

Der empirische t-Test lässt sich schlussendlich folgendermassen darstellen (Bortz, 2001, S. 138):

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} \quad df = n_1 + n_2 - 2$$

Dabei sind die beiden  $x$ -Werte die aus den empirischen Daten berechneten Mittelwerte aus den beiden Stichproben.

Die Voraussetzungen für diesen Test tangieren die Anwendung in der Studie nur insofern, als dass die Varianzen in den zu vergleichenden Populationen gleich sein müssen. Sollte dies nicht so sein, sollen die Freiheitsgrade korrigiert werden (vgl. Clauss & Ebner, 1971, S. 192 f.). Die korrigierten Freiheitsgrade werden jeweils angegeben und diskutiert, auch wenn aufgrund der (hohen) Stichprobengröße in dieser Studie diese Korrektur kaum verwendet wurde.

### **Vergleich zweier Stichprobenmittelwerte aus abhängigen Stichproben (t-Test)**

Wir beschäftigen uns zudem mit Differenzen (der sozialen Validität und prozeduralen Gerechtigkeit) innerhalb einer Stichprobe (der untersuchten Stichprobe). Dazu wurde der Mittelwert der Verteilung von Differenzen benötigt (geschätzter Standardfehler), welcher folgendermassen lautet (Bortz, 2001, S. 141):

$$\hat{\sigma}_{\bar{x}_d} = \sqrt{\frac{\hat{\sigma}_d^2}{n}}$$

Wobei das geschätzte  $\sigma$  in der Formel die Verteilung der Mittelwerte in der Stichprobe darstellt.

Die (empirische) Bedeutsamkeit der Differenz wurde dann folgendermassen überprüft (wobei angenommen wird, dass  $\mu_D = 0$  ist):

$$t_{emp} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}$$

Der t-Wert wurde mit den Irrtumswahrscheinlichkeiten der Normalverteilung verglichen, weil aufgrund unserer (hohen) Stichprobengröße die Verteilung des t-Wertes in eine Normalverteilung übergeht.

## 5. Ergebnisse

Die Ergebnisse widerspiegeln die gemäss den Hypothesen im Kapitel 5 postulierten Erwartungen. Um die Anzahl der Tabellen möglichst gering zu halten, wurden teilweise Resultate zu unterschiedlichen Fragestellungen in einer Tabelle dargestellt; in diesem Fall wird an der entsprechenden Stelle darauf hingewiesen. Zusätzliche Ergebnisse, die für die Beantwortung dieser Fragestellungen nur indirekt von Belang sind, werden im Anhang aufgeführt.

### 5.1 Reliabilität der eingesetzten Instrumente

In den Hypothesen wird postuliert, dass die Reliabilität der eingesetzten Instrumente auf übergeordneter Ebene als gut (mindestens .80) oder auf Subskalen-Ebene mindestens als genügend (.70) zu bezeichnen ist. In Tabelle 1 und 2 sind die Resultate zu den beiden Verfahren, die entweder Soziale Validität oder Prozedurale Gerechtigkeit messen sollen, in den Diagonalen dargestellt. Für die Soziale Validität zeichnet sich ein zwiespältiges Ergebnis ab: Während für die Gesamtskala ein Cronbachs Alpha von .86 und für die Subskalen Transparenz und Feedback jeweils ein Wert von über .70 resultiert, sind die Koeffizienten für die Subskalen Information sowie Kontrolle als unbefriedigend zu beurteilen (.55 resp. .45). Im Hinblick auf die Messgenauigkeit des Instrumentes bedürfen diese Resultate der Diskussion. Ein möglicher Grund dafür könnte sein, dass alle Teilskalen jeweils nur durch zwei Items gebildet werden und dass die Schweizerische Stichprobe differenzierter auf beide Items reagiert als die Stichprobe der Deutschen aus der Studie von Klingner (2003).

**Tabelle 1: Mittelwerte, Standardabweichungen, Korrelationen und Reliabilitäten (Cronbach's Alpha) der Skalen der Sozialen Validität**

	M	SD	1	2	3	4	5
1 Information	3.19	1.08	.55	.65	.49	.50	.80
2 Transparenz	3.37	.94		.74	.59	.50	.85
3 Feedback	3.65	1.18			.88	.49	.82
4 Kontrolle	3.33	1.01				.45	.77
5 Soziale Validität	3.39	.85					.86

Anmerkungen: N = 1290 bis 1676; alle Korrelationen sign. \*\*  $p < .01$  (1-seitig); Cronbach's Alpha in der Diagonale

Ganz im Gegensatz dazu bewegen sich die Reliabilitäten zum Verfahren zur Prozeduralen Gerechtigkeit praktisch immer in der erwarteten Grössenordnung (siehe Tabelle 2). So übertrifft die Reliabilität des Gesamtwertes zur Prozeduralen Gerechtigkeit deutlich die Erwartungen, welche in der Hypothese 5.2 formuliert wurden. Sie egalisiert denselben Wert, den Müller et al. (2007) in ihrer Studie gefunden haben. Zusätzlich erreichen alle Teilskalen mit einer Ausnahme die geforderten .70, welche als Kriterium für eine angemessene Reliabilität von Teilskalen halten. Die Ausnahme ist die Skala „Information“, welche mit .68 mit vernachlässigbarer Differenz den geforderten Wert nicht erreicht. Interessant ist allerdings, dass die Skala mit demselben Namen im Instrument zur Sozialen Validität ebenfalls nicht den Erwartungen entspricht. Hier besteht möglicherweise ein Problem, das über die Operationalisierung des Konstruktes hinweg besteht. Insgesamt konnten aber sowohl beim Instrument zur Messung der Sozialen Validität als auch beim Verfahren zur Messung der Prozeduralen Gerechtigkeit die Hypothesen 5.1 bis 5.3 mehrheitlich bestätigt werden, insbesondere was die Reliabilität der übergeordneten Skalen betrifft. Dies spricht für die Messgenauigkeit der eingesetzten Tests.

**Tabelle 2: Mittelwerte, Standardabweichungen, Korrelationen und Reliabilitäten (Cronbach's Alpha) der Skalen der Prozeduralen Gerechtigkeit**

	M	SD	1	2	3	4	5	6	7	8	9	10	11
1 Berufsbezogenheit Vorhersage	4.19	1.60	.83	.59	.64	.30	.23	.34	.38	.36	.65	.31	.54
2 Information	4.55	1.39		.68	.69	.57	.53	.62	.64	.61	.73	.63	.79
3 Möglichkeit der Leistungsdarstellun g	4.83	1.53			.89	.52	.47	.54	.55	.61	.77	.49	.78
4 Ehrlichkeit und Offenheit	5.38	1.28				.71	.78	.79	.74	.71	.45	.62	.85
5 Konsistenz	5.88	1.31					.82	.78	.68	.70	.42	.63	.79
6 Behandlung	5.65	1.37						.91	.82	.71	.53	.64	.88
7 Kommunikation	5.07	1.52							.87	.71	.58	.63	.87
8 Angemessenheit der Fragen	5.34	1.25								.70	.59	.58	.80
9 Berufsbezogenheit Inhalt	4.68	1.51									.79	.46	.71
10 Feedback	5.28	1.38										.70	.73
11 Prozedurale Gerechtigkeit	5.16	1.13											.96

Anmerkungen: N = 229 bis 292; alle Korrelationen sign. \*\* p < .01 (1-seitig); Cronbach's Alpha in der Diagonale

## 5.2 Konvergente Validität

Wie in den theoretischen Erwartungen formuliert, kommt Kersting (1998) beim Vergleich der Sozialen Validität von Schuler & Stehle (1983) und der Prozeduralen Gerechtigkeit von Gilliland (1993) zum Schluss, dass Ähnlichkeiten zwischen den Konstrukten bestehen und dass bestimmte Teilskalen der jeweiligen Instrumente einander entsprechen. Diese These wurde überprüft und die Zuordnung im Sinne von Kersting mit Hilfe der Analyse der Konstruktvalidität der Verfahren von Klingner (2003) und Bauer (2001) überprüft. In Tabelle 3 und 4 finden sich die Interkorrelationen der Teilskalen zwischen den Konstrukten, wobei die interessierenden Zusammenhänge unterstrichen sind. Es zeigt sich, dass man die Hypothese 5.2.1 und 5.2.2 zur Konkordanz der Skalen zur Sozialen Validität mit der zur Prozeduralen Gerechtigkeit nicht bestätigen kann (Tabelle 3 und 4). Nur gerade 4 von 14 Koeffizienten entsprechend mindestens einer hohen Effektstärke im Zusammenhang. Wenn man aus den restlichen 26 (vier mal zehn minus vierzehn) Konstruktvaliditäten der Teilskalen beider Verfahren die Koeffizienten grösser .50 zählt, so erhält man die Anzahl 6. Das Verhältnis von 4 zu 14 (für die erwarteten Zusammenhänge) ist nur wenig grösser als die 6 zu 26 Korrelationen, für die keine Korrelationen um .50 erwartet werden. Nach der Multitrait-Multimethod-Methode von Campbell und Fiske (1959) wäre aber zu erwarten, dass ein ungleich höheres Verhältnis resultieren müsste, wenn sich die gemäss Kersting (1998) zugeordneten Konstrukte entsprechen würden. In Tabelle 3 und 4 ist zusätzlich jeweils die Fairness als Kriterium der subjektiven Wahrnehmung der Akzeptanz eingefügt, wobei es sich zeigt, dass diese Facette in dieser Stichprobe vor allem mit der Transparenz resp. der Möglichkeit zur Leistungsdarstellung korreliert.

**Tabelle 3: Interkorrelationen der Teilskalen und Gesamtwerte der Sozialen Validität und Prozeduralen Gerechtigkeit (Teil 1)**

	1	2	3	4	5	6
Information	.48	<u>.50</u>	.58	<u>.36</u>	.30	.37
Transparenz	<u>.57</u>	.59	<u>.70</u>	.44	<u>.36</u>	.42
Feedback	.37	<u>.43</u>	.58	<u>.43</u>	.42	.40
Kontrolle	<u>.36</u>	.37	<u>.46</u>	.41	<u>.33</u>	.36
Soziale Validität	.57	.62	.75	.54	.46	.50
Fairness	.50	.59	.65	.45	.42	.50

Anmerkungen: N = 228 bis 348; alle Korrelationen sign. \*\*  $p < .01$  (1-seitig); 1 Berufsbezogenheit Vorhersage, 2 Information, 3 Möglichkeit der Leistungsdarstellung, 4 Ehrlichkeit und Offenheit, 5 Konsistenz, 6 Behandlung; Reliabilität der Fairness: .74.

Anders sieht das Ergebnis aus, wenn man Hypothese 5.2.3 überprüft: Hier stimmen die Korrelationen zwischen den drei zu vergleichenden Werten der Verfahren zur Akzeptanz wie erwartet überein. So korreliert der Gesamtwert der Soziales Validität mit der Prozeduralen Gerechtigkeit (.67) als auch mit der augenscheinlich wahrgenommenen Akzeptanz, der Fairness (.72). Auch die Prozedurale Gerechtigkeit korreliert mit der Fairness-Skala relativ hoch, wenn auch etwas tiefer (zu .61). Gemäss der Hypothese kann man also von einer Konvergenz der Verfahren ausgehen, nichtsdestotrotz dürfte die Eigenständigkeit der Konstrukte noch gegeben sein, wenn man die Höhe der Korrelationen für ähnliche Skalen auf .70 begrenzt. Inhaltlich spricht dies dafür, dass sowohl die Wahrnehmung der Fairness als auch die das Verfahren zur Sozialen Validität und das zur Prozeduralen Gerechtigkeit ähnliches (aber nicht dasselbe) erfassen. Zu beachten gilt, dass die Verfahren jeweils eine unterschiedliche Anzahl Items aufweisen (mit 2, 8 und 34), obwohl die Reliabilitäten nicht sehr stark divergieren (.74, .86 und .96).

**Tabelle 4: Interkorrelationen der Teilskalen und Gesamtwerte der Sozialen Validität und Prozeduralen Gerechtigkeit (Teil 2)**

	7	8	9	10	11	Fairness
Information	.41	.48	.60	<u>.39</u>	.50	.51
Transparenz	.45	.52	<u>.63</u>	.42	.59	.68
Feedback	.40	.46	.50	<u>.36</u>	.53	.54
Kontrolle	.39	.43	<u>.39</u>	.37	.48	.45
Soziale Validität	.53	.60	.68	.49	.67	.72
Fairness	.50	.56	.58	.42	.61	-

Anmerkungen: N = 228 bis 348; alle Korrelationen sign. \*\* p < .01 (1-seitig); 7 Kommunikation, 8 Angemessenheit der Fragen, 9 Berufsbezogenheit Inhalt, 10 Feedback, 11 Prozedurale Gerechtigkeit; Reliabilität der Fairness: .74.

Interessant im Ergebnis scheint die Fragestellung, welche Facetten (Teilskalen) des einen Verfahrens jeweils die der (anderen) Verfahren bestimmt (siehe auch Regressionen im Anhang 7.e/f). So scheint die Soziale Validität vor allem durch den Aspekt der Möglichkeit zur Leistungsdarstellung oder der Berufsbezogenheit (Inhalt) der Prozeduralen Gerechtigkeit bestimmt zu sein, während diese ihrerseits mehr durch die Transparenz oder das Feedback der Sozialen Validität bestimmt wird. Praktisch dieselben Facetten klären auch am meisten Varianz in der subjektiven Wahrnehmung der Fairness der Absolvierenden auf. So klärt insbesondere auch die Transparenz in der Fairness Varianz auf, was bereits Klingner (2003) in Regressionsanalysen festgestellt hat. Insgesamt sind die Zusammenhänge vielfach fast so hoch, wie die aggregierten Werte der Fragebögen

zueinander korrelieren. Ein ähnliches Bild ergibt sich, wenn die Interkorrelationen der Facetten innerhalb der Fragebögen angeschaut werden (siehe Tabelle 1 resp. 2 und Resultate zur diskriminanten Validität in Kapitel 5.3). Das spricht, auf das gesamte Bild der Resultate zur konvergenten Validität Bezug nehmend, von einer gewissen Redundanz (innerhalb) der Instrumente. Dies hat für die Wissenschaft natürlich andere Konsequenzen zur Folge als für die Praxis, worauf in der Diskussion (Kapitel 6) näher eingegangen wird.

### 5.3 Diskriminante Validität

Theoretisch wurde erwartet, dass sich die verschiedenen Konstrukte zur Akzeptanz von demographischen Variablen oder solchen, die die Testung betreffen, abgrenzen (Hypothesen 5.3.1 bis 5.3.4). In Tabelle 5 sind die Ergebnisse zusammengefasst. Wie es sich zeigt, korrelieren die Variablen „Alter“ und „Geschlecht“ trotz grossen Stichprobenzahlen nicht mit den jeweiligen Werten zur Sozialen Validität, zur Prozeduralen Gerechtigkeit oder zur Fairness. Dies entspricht damit Resultaten, wie es zum Beispiel für das Geschlecht auch Klingner (2003) gefunden hat. Für bestimmte Variablen, die mit dem Absolvieren des Berufseignungstests (Multicheck) zu tun haben, sieht es ähnlich aus. So spielt es in Bezug auf die Bewertung der Akzeptanz keine Rolle, wie viele Male eine Person bereits ein Eignungstest absolviert hat. Allerdings sind Zusammenhänge mit den Test-Leistungen tendenziell positiv und werden signifikant (entgegen der Hypothese 5.3.3); das heisst, bessere Absolvierende des Multicheck beurteilen die Soziale Validität, die Prozedurale Gerechtigkeit und vor allem die Fairness höher. Entgegen den Annahmen ist zum Teil auch der Zusammenhang von Facetten der jeweiligen Konstrukte innerhalb eines Instrumentes (Hypothese 5.3.4) sehr hoch (siehe Tabelle 2); dies gilt nur für die Prozedurale Gerechtigkeit. So weisen 11 von 45 Korrelationen eine höhere Effektstärke aus als .70, was die Anforderungen an Redundanz von Verfahren tangiert. Dies entspricht damit den Resultaten zur konvergenten Validität über die Instrumente (s.o.) und Bedarf der Diskussion.

**Tabelle 5: Korrelationen der sozialen Validität und prozed. Gerechtigkeit mit persönlichen Variablen**

	Alter	Geschlecht	Test-Leistung	Test-Erfahrung
Soziale Validität	-.029	-.069	.176**	.002
	[1268]	[1250]	[1169]	[1239]
Prozedurale Gerechtigkeit	-.089	.076	.102	.010
	[225]	[223]	[217]	[227]
Fairness	-.060	-.011	.227**	.001
	[1552]	[1529]	[1416]	[1513]

Anmerkungen: Nicht-parametrische Korrelationen (Spearman-Rho). \*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant; in Klammern [ ] ist N gesetzt.

## 5.4 Mittelwertsvergleiche

Im Rahmen der letzten Hypothesen wird Bezug auf den praktischen Teil der Arbeit genommen und es wird in einer Art „Benchmarking“ die Höhe der Akzeptanz relativ zu den anderen Verfahren überprüft (Vergleichbarkeit der Stichproben siehe Diskussion). Gemäss Hypothese 5.4.1 und 5.4.2 wird erwartet, dass die Akzeptanzwerte für den Multicheck tendenziell schlechter beurteilt wird, als für die in den Vergleich einbezogenen Verfahren, den AZUBI-BK und ein Assessment Center.

**Tabelle 6: Vergleich der 4 Facetten zur sozialen Validität für die Einschätzung des Multicheck und des AZUBI-BK**

	Stichprobe Multicheck			Stichprobe AZUBI			Differenz					
	N	M	SD	N	M	SD	M	SF	t	df	sig.	ES
Transparenz	1444	3.19	0.94	1173	3.07	0.97	0.12	0.98	3.13	2615	.001	0.12
Information	1517	3.37	1.08	1173	3.16	0.99	0.21	1.02	5.29	2688	.000	0.21
Feedback	1676	3.65	1.18	1173	3.70	1.08	-0.05	1.07	-1.23	2847	n.s.	0.05
Kontrolle	1527	3.33	1.01	1172	3.34	0.85	-0.01	0.97	-0.27	2697	n.s.	0.01
Fairness	1576	3.46	1.21	1174	3.40	1.24	0.06	1.11	1.41	2748	n.s.	0.05

Anmerkungen: t-Test für unabhängige Stichproben (2-seitig). SF = Standardabweichung, ES = Effektstärke.

In Tabelle 6 wird der Multicheck mit dem AZUBI-BK verglichen: Entgegen den Annahmen gibt es zwei signifikante Differenzen zu Gunsten des Multicheck, genauer in den Skalen Transparenz und Information. Hypothese 5.4.1 kann also nicht bestätigt werden, auch

wenn die entsprechenden Effektstärken klein (.12 resp. .21; kleiner Effekt = 0.1) ausfallen. Werden die entsprechenden Vergleiche auf Basis der verwendeten Items gemacht, fallen die Resultate ähnlich aus: Auch hier kann die Hypothese, dass der Multicheck schlechter beurteilt wird, nicht aufrecht erhalten werden, auch wenn im besten Fall nur eine niedrige Effektstärke resultiert. Da die signifikanten Ergebnisse nicht den Erwartungen entsprechen, müssen sie gründlich diskutiert werden (Kapitel 6).

Im Vergleich mit dem Assessment Center, das für die Rekrutierung von Fachpersonal in der Luftfahrt verwendet wurde, und für das Müller et al. (2007) in der entsprechenden Stichprobe die Prozedurale Gerechtigkeit untersuchen, wird die Erwartungen in den meisten Skalen bestätigt (vgl. Tabelle 7). So fällt die Akzeptanz ausser in den Skalen zur Berufsbezogenen Vorhersage, zur Konsistenz, zur Kommunikation und der Möglichkeit zur Leistungsdarstellung immer zu Gunsten des Assessment Centers aus (gemäss Hypothese 5.4.2). Mit Ausnahme der letzten Skala resultiert jeweils kein signifikanter Unterschied, was an der geringen Teststärke liegen könnte. Mit der Möglichkeit zur Leistungsdarstellung wird von der Stichprobe, die den Multicheck absolviert hat, eine Skala besser bewertet. Allerdings weisen diese Akzeptanzwerte vom Multicheck eingeschränkte Varianz auf (im Gegensatz zu den meisten anderen Werten), was eine Selektivität der antwortenden Personen bedeutet. Diese Selektivität könnte ein Grund dafür sein, dass die Akzeptanzwerte (fälschlicherweise) zu Gunsten des Multicheck ausfallen.

**Tabelle 7: Vergleich der 9 Facetten und der Gesamtskala zur Prozeduralen Gerechtigkeit für die Einschätzung des Multicheck und eines Assessment Centers**

	Stichprobe Multicheck			Stichprobe AC			Differenz					
	N	M	SD	N	M	SD	M	SF	t	df	sig.	ES
1	292	4.19	1.60	300	4.09	1.73	.10	1.29	0.94	590	n.s.	0.08
2	281	4.56	1.38	300	4.92	1.49	-.36	1.20	-3.61	579	.000	0.30
3	274	4.86	1.50	300	4.46	1.70	.37	1.27	3.77	572	.000	0.32
4	276	5.40	1.25	300	5.65	1.28	-.25	1.13	-2.66	574	.004	0.22
5	270	5.87	1.33	300	5.82	1.33	.05	1.16	0.52	568	n.s.	0.04
6	266	5.61	1.39	300	6.17	1.11	-.56	1.12	-5.96	564	.000	0.50
7	259	5.07	1.50	300	5.20	1.43	-.13	1.21	-1.27	557	n.s.	0.11
8	267	5.31	1.24	300	5.65	1.29	-.34	1.13	-3.58	565	.000	0.30
9	285	4.68	1.51	300	5.45	1.59	-.77	1.25	-7.46	583	.000	0.62
10	229	5.16	1.09	300	5.34	1.12	-.18	1.05	-1.95	527	.026	0.17

Anmerkungen: *t*-Test für unabhängige Stichproben (2-seitig). SF = Standardfehler, ES = Effektstärke. 1 Berufsbezogenheit Vorhersage, 2 Information, 3 Möglichkeit der Leistungsdarstellung, 4 Ehrlichkeit und Offenheit, 5 Konsistenz, 6 Behandlung, 7 Kommunikation, 8 Angemessenheit der Fragen, 9 Berufsbezogenheit Inhalt, 10 Prozedurale Gerechtigkeit.

In einer Art Kontroll-Untersuchung wurde der Unterschied zwischen sehr ähnlichen Verfahren, nämlich dem Multicheck und dem BIS-4, einem „klassischen“ Intelligenztest (gemäss Hypothese 3.4.3). Die Differenzen für den Vergleich wird berechnet, so kann die Hypothese verworfen werden, dass kein Unterschied besteht (vgl. Tabelle 8). Die Effektgrößen sind dabei klein bis gross (0.1 = kleiner, 0.5 = mittlerer und 0.8 = grosser Effekt). Am Besten beurteilt werden dabei die Subskalen Transparenz und Information, gefolgt von der Gesamtskala zur Sozialen Validität. Demnach wird der Multicheck deutlich transparenter, informativer und akzeptabler erlebt als ein klassischer Intelligenztest. Es soll auch hier der Vergleich auf Itemebene gemacht werden (siehe Anhang 7.F): Das Resultat für die Subskalen wiederholt sich. Deutlich besser wird insbesondere ein Item zur Information bewertet („Der Test informiert mich, welche Anforderungen in einem Büroberuf gestellt werden.“). Der geringste Unterschied zum BIS-4 ergibt sich für Item 5 zur Skala „Kontrolle“ („Durch eigene Anstrengung konnte ich das Testergebnis beeinflussen.“).

**Tabelle 8: Vergleich der 4 Facetten zur sozialen Validität für die Einschätzung des Multicheck und des BIS-4**

	Stichprobe Multicheck			Stichprobe BIS-4			Differenz					
	N	M	SD	N	M	SD	M	SF	t	df	sig.	ES
Transparenz	1444	3.19	0.94	1173	2.49	1.02	0.70	0.99	18.02	2615	.000	0.71
Information	1517	3.37	1.08	1173	2.69	1.03	0.68	1.03	17.00	2688	.000	0.66
Feedback	1676	3.65	1.18	1173	3.29	1.10	0.36	1.07	8.83	2847	.000	0.34
Kontrolle	1527	3.33	1.01	1172	3.11	0.88	0.22	0.98	5.80	2697	.000	0.23
Fairness	1576	3.46	1.21	1174	3.06	1.19	0.40	1.10	9.46	2748	.000	0.36

Anmerkungen: *t*-Test für unabhängige Stichproben (2-seitig). *SF* = Standardfehler, *ES* = Effektstärke.

## 6. Diskussion

### 6.1 Reliabilität / Konstruktvalidität

Um die Nützlichkeit der Instrumente zur sozialen Validität und der Prozeduralen Gerechtigkeit zu überprüfen, wurden die Gütekriterien der Fragebogen, also insbesondere die Reliabilität und die Konstruktvalidität untersucht. Die Reliabilität gibt nur in Bezug auf das von Klingner (2003) konstruierte Instrument zur Sozialen Validität Anlass zu Kritik: Während die Skalen Transparenz und Feedback gute Kennwerte liefern, sind die Skalen Information und Kontrolle als nicht oder bedingt akzeptabel zu bezeichnen. Klingner (2003) hat leider die Reliabilität der Skalen in ihrer Untersuchung nicht angegeben; allerdings hat sie das Verfahren auf Grundlage einer Faktorenanalyse konzipiert. Den Grund die diskussionswürdige Messgenauigkeit dieser Skalen sehen die Autoren vor allem in der Eigenart der Stichprobe. Für die Skala Information kann die Stichprobe die Items zur Skala nicht beantworten, weil die meisten Schüler noch nicht genau wissen, wie es im Beruf zu und hergeht. Das führt automatisch zu einem Messfehler und damit zur geringeren Kohärenz der Items untereinander. Im Gegensatz dazu beziehen sich die Fragen zur Berufsbezogenen Vorhersage der SPJS auf die Ausbildung, also die Lehre, und da die meisten Schüler eine Schnupperlehre absolviert haben, ist hier mehr Klarheit gegeben.

Bei der Skala Kontrolle halten die Autoren das Studien-Design als Grund für das schlechte Abschneiden bezüglich Messgenauigkeit. Da Daten zur Sozialen Validität erhoben wurden, nachdem die Testergebnisse bekannt waren, hat Item 8 („Ich kann ungefähr einschätzen, wie ich bei diesem Test abgeschnitten habe.“) keinen Sinn mehr gemacht. Die meisten Teilnehmenden werden hier also einen höheren Wert angeben, als sie es zum Vorneherein gemacht hätten. Das führt zu einer (ungewollten) Varianzeinschränkung,

was wiederum das schlechte Resultat bedingt. Insgesamt sehen die Autoren hier keinen Handlungsbedarf bezüglich des Instrumentes von Klingner (2003); das gilt auch für die Skala Information. Man sollte sich in anderen Untersuchungen aber bewusst sein, dass die Bedingungen (Studiendesign und Stichprobe) die Messgenauigkeit des Instrumentes beeinflussen kann. Damit ist aber ein Einsatz des Instrumentes immer noch möglich, ähnlich, wie die Autoren es für den Fragebogen zur Prozeduralen Gerechtigkeit propagieren würden; denn die Resultate zu diesem Verfahren zeigen ganz klar, dass die Reliabilität entsprechend unseren (positiven) Erwartungen ausfällt. Dies zeigen auch die Reliabilitäten zu den Gesamtwerten, die bei beiden Verfahren nahezu hervorragend ist.

Mehr zu diskutieren gibt es bei der Konstruktvalidität der Verfahren, insbesondere zur konvergenten Konstruktvalidität bei der SPJS. Hier zeigt sich, dass etliche der Skalen innerhalb des Verfahrens als redundant taxiert werden können, weil gewisse Korrelationen so hoch sind. So korreliert z. B. die Skalen Ehrlichkeit und Offenheit, Konsistenz, Behandlung und Kommunikation nahezu mit einer grossen Effektstärke miteinander. Ähnliches gilt für die Skalen Information, Möglichkeit der Leistungsdarstellung und Berufsbezogenheit Inhalt, wobei sogar inhaltlich erstaunlich ist, weshalb hier so hohe Korrelationen zu Tage treten. Allgemein sind diese Effekte gemäss den Autoren nach zu hoch, um die Konstruktfacetten trennen zu können; zudem können wir diese Resultate nicht mit den Rahmenbedingungen der Studie erklären. Diese hohe Konvergenz einiger Skalen haben auch Müller et al. (2007) gefunden, weshalb sie nur den Gesamtwert in ihrer Studie verwenden. Wir tendieren deshalb darauf, dies dem Verfahren zuzuschreiben. Weniger erstaunen die Verfahrens-internen Korrelationen zur Sozialen Validität; hier finden sich weitgehend mittlere Effekte. Ausnahme stellt die hohe Konvergenz zwischen Transparenz und Information dar, welche bereits Klingner (2003) festgestellt hat.

Diskussionen geben die Ergebnisse zu den Konstruktvaliditäten über die beiden verwendeten Instrumente. Wir haben erwartet, dass die Zuordnungen der Konstrukte, wie es Kersting (1998) aufgrund der Beziehungen zwischen Sozialer Validität und Prozeduraler Gerechtigkeit propagiert, auch in den Korrelationen sichtbar werden. Dem ist aber nicht so, denn die Korrelationen sind kaum höher als die anderen Zusammenhänge. Eine Ausnahme stellt die Korrelation zwischen der Transparenz und der Möglichkeit zur Leistungsdarstellung dar; diese zwei Konstrukte scheinen auch anderweitig wichtig zu sein (s. u.). Auch sind die Effekte der Zusammenhänge zwischen unterschiedlichen Konstrukten (gemäss Kersting, 1998) meist nicht deutlich geringer als .50. Betrachten wir

die Korrelationen zwischen den Gesamttestwerten der Sozialen Validität, Prozeduralen Gerechtigkeit und der in diesem Fall subjektiven Fairnessbeurteilung so zeigt sich aber, dass die Verfahren immerhin in etwa dasselbe erfassen (Korrelationen um .70). Die Frage stellt sich deshalb, weshalb man in der Praxis ein solch umfangreiches Instrument wie das der SPJS einsetzen soll, wenn man nur einen undifferenzierten Wert zur Akzeptanz erheben möchte. Wir nehmen schon hier voraus, dass wir einen Einsatz der hier untersuchten Instrumente nur dann propagieren, wenn eine differenzierte Beurteilung gewünscht wird. Interessant ist deshalb auch, dass die Fairness, also die subjektiv wahrgenommene Akzeptanz, insbesondere und in hohem Masse durch die Transparenz der Sozialen Validität und der Möglichkeit zur Leistungsdarstellung aus der SPJS vorhergesagt werden kann (korrelations- und regressionsanalytisch).

Zum Gesamtbild der Konstruktvalidität trägt auch die diskriminante Validität bei. Diese ist in unserem Falle mit den eingesetzten Demographie- und Testparametern gegeben, mit Ausnahme der Korrelation zwischen den Akzeptanzwerten und der Testleistung. Diese geringen Effekte zwischen Akzeptanz und Leistung schreiben die Autoren aber dem Design zu, weil die Datenerhebung nach dem Erhalt der Testergebnisse gemacht wurden. Insofern ist anzunehmen, dass hier ohne diesen Einfluss ein Null-Zusammenhang wahrscheinlich wäre. Den Erwartungen zur Diskriminanz entgegengerichtet sind auch die bereits genannten Zusammenhänge zwischen Konstrukten innerhalb oder über die Instrumente der Studie, die nach Kersting (1998) keinen grossen Effekt erwarten lassen.

Wir führen die genannten Probleme der konvergenten resp. diskriminanten Validität auf die unterschiedliche Operationalisierung von Skalen mit derselben oder einer ähnlichen Bezeichnung der beiden untersuchten Instrumente zurück. So zeigt sich z. B., dass die Skala Information aus der SPJS inhaltlich eher auf den Vorgang der Information durch die eignungs-abklärende Firma zielt, während die Skala mit demselben Namen eher auf die Tätigkeitsinformation der zu bewerbenden Ausbildung abzielt. Dementsprechend findet sich eine hohe Korrelation zwischen der letztgenannten Skala mit der Skala Berufsbezogenheit Inhalt zur Prozeduralen Gerechtigkeit. Ähnliches gilt auch für die Skalen Feedback aus der SPJS und dem Fragebogen zur Sozialen Validität. Damit ist das Problem der fehlenden Konvergenz resp. der zu geringen Diskriminanz z. T. erklärbar. Insbesondere die geringe Diskriminanz kann aber auch durch den Umstand erklärt werden, dass die SPJS eine zu hohe Anzahl Skalen beinhaltet, die inhaltlich sehr ähnlich sind (s. o.).

Das Fazit lautet, dass nur die Reliabilität, nicht aber die Konstruktvalidität als günstig eingestuft werden kann. Dies macht einen Einsatz der Instrumente in der Praxis möglich, allerdings vorerst nur dann, wenn ein zusammengefasstes Urteil über die Akzeptanz eines Eignungs- oder Abklärungsverfahrens zum Einsatz kommen soll. In diesem Fall ist es dann allerdings bereits ausreichend, mit nur zwei Fragen (zur Fairness-Beurteilung) bei nahezu akzeptabler Reliabilität ein sehr guter Hinweis auf die Akzeptanz zu erfahren, ohne dass umfangreichere Verfahren eingesetzt werden müssen. Sollen differenziertere Angaben gewonnen werden, so müssen unserer Meinung nach die bestehenden Verfahren weiter untersucht beziehungsweise entwickelt werden, weil die Facetten der Konstrukte innerhalb des Instrumentes zu hoch und über die Instrumente teilweise zu wenig zusammenhängen. Verbesserungspotenzial liegt möglicherweise bei der Umsetzung der theoretischen Vorgaben in den Fragebogen, wenn die Abgrenzungen einzelner Facetten innerhalb des Fragebogens besser gemacht würden (oder die Anzahl Facetten reduziert wird).

## 6.2 Mittelwertsvergleiche

Im Rahmen der Untersuchung war es ein Ziel, die Akzeptanz, und damit sowohl die Soziale Validität als auch die Prozedurale Gerechtigkeit, des Multicheck zu untersuchen und mit der anderer Verfahren zu vergleichen. Vergleiche wurden mit drei Verfahren, nämlich mit einer Integration von Intelligenztests mit Arbeitsproben, dem AZUBI-BK, einem Assessment Center und einem „klassischen“ Intelligenztest, dem BIS-4 durchgeführt. Für den Vergleich der Sozialen Validität von Multicheck versus AZUBI-BK erhalten wir zwei signifikante Ergebnisse zu den Skalen Transparenz und Information, die zu Gunsten des Multicheck ausfallen. Der Multicheck wird also transparenter und informativer erlebt, auch wenn der Effekt nur gering ausfällt. Dies ist doch erstaunlich, wenn man bedenkt, dass die AZUBI-BK eine Integration aus Intelligenztest und Arbeitsprobe darstellen soll und die Arbeitsprobe im Gegensatz zu Tests generell als transparent(er) erlebt wird. Wir können diesen Effekt wiederum nicht damit erklären, dass die Stichprobe selektiv ist, weil wir in keinem Fall eine (deutlich) geringere Standardabweichung der Skalenwerte erhalten. Wir können folgende Vermutung anstellen: Der Multicheck nimmt ebenfalls Bezug zur Arbeitswelt auf, weil das Verfahren berufsabhängig konstruiert worden ist (vgl. Widmer, 2006). So enthält z. B. der Test für die technische Branche ein Testgebiet zum Technischen Verständnis mit Elementen, die eindeutig Bezug zum Beruf nehmen. Dies

dürfte einen für die Jugendlichen hohen Anforderungsbezug aufrufen, auch wenn diese noch gar keine grosse Erfahrung mit der Arbeitswelt haben.

Der Vergleich mit dem Assessment Center zeigt, was erwartet worden ist: In fast allen Belangen ergeben sich Akzeptanzwerte (der Prozeduralen Gerechtigkeit) zu Gunsten des Assessment Centers. Dass dies normal ist, dürfte der Umstand bezeugen, dass ein Assessment ein sehr aufwändiges Verfahren mit einem mehrfach höheren Preis als ein einfacher Eignungstest ist. Erstaunlicherweise ergeben sich aber drei Werte, die weder für das eine noch das andere Verfahren oder sogar für den Multicheck sprechen. Aufgrund der etwas geringen Power für die Vergleiche ist insbesondere das Resultat zu Gunsten des Multicheck erstaunlich. So wird die Möglichkeit zur Leistungsdarstellung für den Multicheck um mindestens einen kleinen Effekt höher eingestuft als für das Assessment. Das verwundert umso mehr, als dass gerade Arbeitsproben in dieser Hinsicht als optimal erlebt werden. Wir erklären das Ergebnis mindestens teilweise damit, dass wir aufgrund der Selektivität der Stichprobe (erkennbar an der Varianzeinschränkung) zu hohe Werte erhalten.

Nicht aufgrund der Selektivität erklärbar erscheinen die Resultate zum Vergleich der Sozialen Validität von Multicheck versus BIS-4. Hier erhalten wir immer signifikante und mit teilweise fast hoher Effektstärke ausgeprägte Ergebnisse zu Gunsten des Multicheck. Demnach wird der Multicheck transparenter, informativer, mit besserem Feedback und grösserer Kontrolle erlebt als der BIS-4. Allerdings ist zu sagen, dass der BIS-4 ein relativ antiquarischer Intelligenztest darstellt und dementsprechend von den Absolvierenden in der Studie von Klingner (2003) als wahrscheinlich sehr altbacken aufgefasst worden ist. Zudem könnte es sein, dass wir in unserer Studie aufgrund von Erinnerungseffekten „geschönte“ Ergebnisse bekommen, weil wir den Test erst deutlich später als bei dessen Durchführung haben evaluieren lassen. Dementsprechend sollte der Effekt nicht so deutlich ausgelegt werden.

Eine andere Möglichkeit, die zum Teil sehr optimistischen Resultate zu erklären, wäre nicht die Selektivität, sondern die Unterschiedlichkeit der Stichproben im Vergleich zwischen dem Multicheck und den anderen Verfahren. Die Unterschiedlichkeit bezieht sich dabei vor allem auf demografische Variablen. Während beim Multicheck vielfach jüngere Jugendliche den Test machen (Altersspektrum ca. 14 bis 16 Jahre), sind es bei den anderen Verfahren vielfach ältere Personen. Daraus - und in Anbetracht der Ergebnisse - könnte abgeleitet werden, dass ältere Personen Verfahren aufgrund umfassenderen Erfahrungen (in welcher Hinsicht auch immer) die Verfahren kritischer beurteilen. Dazu passt auch, dass die Personen der Stichproben sowohl in der Studie von Klingner (2003)

und Müller et al. (2007) meistens gebildeter waren und dass Klingner (2003) zum Teil Effekte zwischen Berufsschülern und Schülern findet. Andere Faktoren, wie die Länderzugehörigkeit, können wir hier nicht mehr diskutieren. Diese demografischen resp. persönlichen Variablen müssten in weiteren Untersuchungen in die Auswertung einbezogen und kontrolliert werden.

### 6.3 Empfehlungen für die Firma Multicheck

Die Ergebnisse zeigen, dass der Multicheck hinsichtlich der Sozialen Validität besser abschneidet, als es zu erwarten wäre. Diesbezüglich ist dem Verfahren und damit der Firma Multicheck ein gutes Zeugnis auszustellen. Dennoch sind sicher Punkte zu erwähnen, die verbesserungswürdig sind; deshalb wurde diese Studie (auch im Sinne der Firma Multicheck) durchgeführt. Auf Basis der Ergebnisse ist es schwierig, entsprechende Verbesserungsmöglichkeiten aufzuzeigen, obwohl es relativ einfach ist, Kritikpunkte anzugeben (diese finden sich in der Diskussion im Kap. 6). Wir fassen deshalb hier die Möglichkeiten zusammen, die wir aufgrund der Kenntnisse zum Verfahren oder zur Organisation der Firma Multicheck haben. Die gemachten Empfehlungen stützen sich insbesondere auf den Vergleich des Multicheck mit den Intelligenztests AZUBI-BK und BIS-4 und damit den Daten von Klingner (2003), weil wir nicht „Äpfel mit Birnen“, sprich Test mit Assessment Center, vergleichen wollen.

Die Ergebnisse zeigen, dass es in den Skalen Feedback und Kontrolle einen zwar nicht signifikanten, aber betragsmässigen Unterschied zu Ungunsten des Multichecks im Vergleich mit dem AZUBI-BK gibt. Dieses Resultat deutet darauf hin, dass die Teilnehmenden das Feedback, welches sie im Anschluss an die Testdurchführung erhalten, als nicht sehr zuvorkommend interpretieren. Da wir wissen, dass die Firma Multicheck die Ergebnisse nur brieflich und mit einem Serienbrief kommentiert versendet, würden wir die Ergebnisse so interpretieren, dass das Feedback als „zu spärlich“ angesehen wird. Wir würden deshalb empfehlen, den Telefondienst, den die Firma anbietet, in ihrem Schreiben als gute Möglichkeit zu empfehlen, ein persönliches Feedback einzuholen. Andere Möglichkeiten sehen wir nicht, da es schwierig werden könnte, mehr als 30'000 getestete Personen pro Jahr ein persönliches Schreiben oder anderweitiges Feedback zu geben.

Zur Kritik in Beziehung auf das Akzeptanz-Kriterium „Kontrolle“ ist zu erwähnen, dass die meisten Jugendlichen der Stichprobe erstmals einen solchen Test absolviert haben und deshalb der Stressfaktor relativ hoch sein dürfte. Andererseits gilt es genau deswegen, genauer auf die Bedingungen vor oder während der Testdurchführung Acht zu geben. Hier sollte also die Firma Multicheck ansetzen und Verbesserungsmöglichkeiten wahrnehmen. Die exemplarischen Empfehlungen der Autoren lauten:

- Die Testlänge von ca. 4 Stunden für die Tests im kaufmännischen und technischen Bereich sollte heruntergesetzt werden. Wie Widmer (2006) gezeigt hat, sind die beiden eingesetzten Verfahren (Papier-Bleistift- und Computer-Test) zu einem grossen Teil redundant.
- Einige Testgebiete dürften unserer Meinung nach durchaus stressvoller sein als andere; womöglich wären diese durch andere zu ersetzen oder womöglich ganz weg zu lassen. Es ist auch sinnvoll, bei der Revidierung des Verfahrens vermehrt Kriterien wie „Akzeptanz“ oder „Spas“ zu beachten.
- Es sollte dafür gesorgt werden, dass die Rahmenbedingungen auch die physiologischen Voraussetzungen beachten. Es ist anzunehmen, dass ein Grossteil der Jugendlichen vor der Testdurchführung zu wenig trinkt und es deshalb zu einer Leistungseinbusse kommen kann. Wir empfehlen, zumindest vor der Testung kostenlos Wasser oder Tee anzubieten.
- Die Autoren empfehlen ein Internet-Forum (oder Blog), in welchem die Teilnehmer/innen ihre Meinung zum Testverfahren und zu den Rahmenbedingungen äussern können.

## 7. Anhänge

### 7. A) Literatur

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* 2 (pp. 267-299). New York: Academic Press.
- Arnold, B. & Helbling, R. (2007). *Stand und Bedürfnisse bezüglich Testanwendung in der allgemeinen Berufs-, Studien- und Laufbahnberatung sowie in der IV-Berufsberatung der deutschen Schweiz - eine Befragung unter Berufsberaterinnen und Berufsberatern*. Unveröffentlichte Masterarbeit. Bern: Institut für Psychologie der Universität Bern.
- Bauer, T.N., Truxillo, D.M., Sanchez, R.J., Ferrara, Ph., Craig, J.M. & Campion, M.A. (2001). Applicant Reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, 54, 387-420.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Berlin: Springer Verlag.
- Bortz, J. (2003). *Statistik: Für Human- und Sozialwissenschaftler*. Berlin: Springer Verlag.
- Campbell, D.T. & Fiske, D.T. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56 (3), 83-105.
- Claus, G. & Ebner, H. (1971). *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*. Frankfurt a. M.: Verlag Harri Deutsch.
- Cohen-Charash, Y. & Spector, P. E. (2001). The role of justice in organizations: a meta-analysis. *Organizational Behavior and Human Decision Processes*, 86, 278-321.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H. & Ng, K.Y. (2001). Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86, 425-445.
- Festinger, L. (1954) A theory of social comparison processes, *Human Relations*, 7, 117-140.
- Fruhner, R., Schuler, H., Funke, U., & Moser, K. (1991). Einige Determinanten der Bewertung von Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 170-178.
- Gilliland, S.W. (1993). The perceived fairness of selection systems: an organizational justice perspective. *Academy of Management Review*, 18 (4), 694-734.
- Grandjean, Ch. (2007). *Un essai de compréhension de l'usage des tests d'aptitudes professionnelles multichек et basic--check dans la sélection des apprenants*.

(unveröffentlichte Lizentiatsarbeit am Psychologischen Institut der Universität Lausanne).

- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
- Jungo, D. & Zihlmann, R. (2002). Lehrlingsselektion. Empirische Untersuchung des basic-check. *Panorama*, 2, 30-32.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61-75.
- Klingner, Y. (2003). *Integration von Tests und Arbeitsproben: Entwicklung eines neuen Verfahrenskonzepts am Beispiel der Arbeitsprobe zur berufsbezogenen Intelligenz für Büro- und kaufmännische Tätigkeiten (AZUBI-BK)*. Berlin: dissertationen.de – Verlag im Internet.
- Köchling, A. C. & Körner, S. (1996). Personalauswahl aus der Sicht der Betroffenen: Zur bewerberorientierten Gestaltung von Beurteilungssituationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 40, 22-37.
- Leventhal, G. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social Exchange: Advances in theory and research* (pp. 27-55). New York: Plenum Press.
- Leventhal, G.G. (1976). The distribution of rewards and resources in groups and organizations. *Advances in experimental social psychology*, 9, 92 – 131.
- Lind, E., Lissak, R. & Conlon, D. (1983). Decision control and process control effects on procedural fairness judgments. *Journal of Applied Social Psychology*, 13, 338-350.
- Lind, E.A., Kanfer, R., Early, P.C. (1990). Voice, Control and Procedural Justice. *Journal of Personality and Social Psychology*, 59, 952-959.
- Lupfer, M., Weeks, K., Doan, K., Houston, D. (2000), Folk conceptions of fairness and unfairness. *European Journal of Social Psychology*, 30, 405-428.
- Müller, A., Janssen, J., Jarzina, D. (2007). *Das Bewerbungsverfahren aus Sicht der Bewerbenden - Vorhersageunsicherheit als Moderator des Einflusses prozeduraler Fairness auf die Attraktivität von Organisationen*. Forschungsprojekt 07/17. Universität Mannheim (in press).
- Rassinski, A., Tyler, T.R. (1987). Social Psychology and Political Behavior. In S. Long (Ed.), *Political behavior annual*. 1, 103 – 128. Boulder CO: Westview press.
- Schachter, S. (1951). Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 46, 190-207.

- Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- Schuler, H. (1990). Personenauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 4, 184-191.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H., Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. *Zeitschrift für Personalpsychologie*, 6, 2, 60-70.
- Skinner, H.A & Pakula A. (1986). Challenge of Computers in Psychological Assessment. Professional Psychology: Research and Practice Copyright. *American Psychological Association*, 17, 1, 44-50.
- Thibaut, J. and L. Walker. (1975). *Procedural Justice: A Psychological Analysis*. Hillsdale, NJ: Erlbaum.
- Tyler, T. R. (1987). Conditions Leading to Value-Expressive Effects in Judgments of Procedural Justice: A Test of Four Models. *Journal of Personality and Social Psychology*, 52, 2, 333-344.
- Tyler, T. (1989). The psychology of procedural justice: A test of the group-value model. *Journal of Personality and Social Psychology*, 57, 830-838.
- Tyler, T.R. & Lind, E.A. (1992). A relational model of authority in groups. *Advances in Experimental Social Psychology*, 25, 115-119.
- Van den Bos, K., Vermunt, R., & Wilke, H. A. M. (1997). Procedural and distributive justice: What is fair depends more on what comes first than on what comes next. *Journal of Personality and Social Psychology*, 72, 95-104.
- Vidmar, Neil (1990). The Origins and Consequences of Procedural Fairness. *Law and Social Inquiry*, 15, 877-892.
- Walster, E., E. Berscheid und G. W. Walster. (1973). New Directions in Equity Research, *Journal of Personality and Social Psychology*, 6, 435-441.
- Widmer, M. (2006). *Reliabilität und Validität eines Berufseignungstests bei kaufmännischen Lehrlingen*. (unveröffentlichte Lizentiatsarbeit am Psychologischen Institut der Universität Bern).

## 7. B) Fragebogen zur Sozialen Validität nach Klingner (2003)

Frage 1	Ich kann mir vorstellen, dass durch dieses Verfahren berufliche Eignung festgestellt wird.
Frage 2	Der Test informiert mich, welche Anforderungen in einem Büroberuf gestellt werden.
Frage 3	durch den Test werden meine Stärken sichtbar.
Frage 4	durch den Test erkenne ich meine Schwächen.
Frage 5	durch eigene Anstrengung konnte ich das Testergebnis beeinflussen.
Frage 6	Die einzelnen Testaufgaben sind praxisnah.
Frage 7	Mir scheint, durch diesen Test können Personen, die für einen Büroberuf geeignet sind, auch ermittelt werden.
Frage 8	Ich kann ungefähr einschätzen, wie ich bei diesem Test abgeschnitten habe.
Frage 9	Der Test ist fair.
Frage 10	Diesen Test würde ich als Auswahlverfahren für einen Büroberuf akzeptieren.

→ Beantwortung auf 5-stufiger Likert-Skala (1 = trifft nicht zu; 5 = trifft zu).

Faktorstruktur der sozialen Validität resp. Zuordnung der Items zu Skalen im Fragebogen

Item	Faktor
Frage 1	Transparenz
Frage 2	Information
Frage 3	Feedback
Frage 4	Feedback
Frage 5	Kontrolle
Frage 6	Information
Frage 7	Transparenz
Frage 8	Kontrolle
Frage 9	Subjektive Akzeptanzbeurteilung
Frage 10	Subjektive Akzeptanzbeurteilung

## 7. C) Fragebogen zur Prozeduralen Gerechtigkeit nach Müller et al. (2007) resp. Bauer (2001)

Item 1	Bei diesem Auswahlverfahren gut abzuschneiden, bedeutet, dass man auch in der Lehre erfolgreich sein kann.
Item 2	Ich bekam von den Mitarbeiterinnen und Mitarbeitern der Firma Multicheck ausreichend Informationen darüber, wie das Auswahlverfahren ablaufen würde.
Item 3	Ich hatte während des Auswahlverfahrens die Möglichkeit, mein Können und meine Fähigkeiten darzustellen.
Item 4	Während des Auswahltages wurde ich offen und ehrlich behandelt.
Item 5	Das Auswahlverfahren wurde für alle Bewerber/Bewerberinnen auf dieselbe Art und Weise durchgeführt.
Item 6	Ich wurde während des gesamten Auswahlverfahrens höflich und zuvorkommend behandelt.
Item 7	Während des Auswahlverfahrens gab es für mich genügend Möglichkeiten zur Kommunikation mit den Mitarbeitern der Firma Multicheck.
Item 8	Die Inhalte des Tests erschienen mir in keiner Weise voreingenommen.
Item 9	Dieser Test misst Fähigkeiten für das erfolgreiche Absolvieren einer Lehre.
Item 10	Mir wurde klar, wann ich die Testresultate erhalten würde.
Item 11	Der Test gab mir die Möglichkeit, meine berufsrelevanten Fähigkeiten deutlich zu machen.
Item 12	Die Mitarbeiter und Mitarbeiterinnen der Firma Multicheck nahmen sich Zeit und beantworteten aufrichtig alle Fragen während des Tests.
Item 13	Während des Auswahlverfahrens waren die Mitarbeiter und Mitarbeiterinnen der Firma Multicheck mir gegenüber immer rücksichtsvoll und entgegenkommend.
Item 14	Die Fragen im Test verletzen meine Privatsphäre nicht.
Item 15	Jemand, der bei diesem Auswahlverfahren gut abschneidet, wird auch in der Lehre gut sein.
Item 16	Schon von Beginn an war für mich klar, wie das Auswahlverfahren ablaufen würde.
Item 17	Die Mitarbeiter und Mitarbeiterinnen der Firma Multicheck machten keine Unterschiede in der Art und Weise wie sie die einzelnen Bewerber/Bewerberinnen behandelten.
Item 18	Ich bin zufrieden mit der Art und Weise, wie ich während des Auswahlverfahrens behandelt wurde.
Item 19	Ich hatte das Gefühl, dass ich gegebenenfalls meine Bedenken und Anliegen hätte äußern können.
Item 20	Ich wusste, wann ich Feedback zu meinen Testresultaten erhalten würde.
Item 21	Während des Auswahlverfahrens versuchten die Mitarbeiter der Firma Multicheck nicht, mir bestimmte Informationen vorzuenthalten.
Item 22	Ich wurde während des Auswahltages mit Respekt behandelt.
Item 23	Ich konnte während des Auswahlverfahrens immer Fragen stellen.
Item 24	Ich konnte in dem Auswahlverfahren zeigen, was ich kann.
Item 25	Die beantworteten Fragen über den weiteren Ablauf des Tests waren ehrlich und konkret.
Item 26	Der Inhalt des Tests steht in Bezug zur Ausbildung.
Item 27	Ich war von Beginn an ausreichend darüber informiert, was mich bei dem Auswahlverfahren erwarten würde.
Item 28	Das Auswahlverfahren gibt den Bewerbern/Bewerberinnen die Möglichkeit zu zeigen, was sie wirklich können.
Item 29	Es gab keine Unterschiede in der Behandlung einzelner Bewerber/Bewerberinnen.
Item 30	Ich war zufrieden mit der Art, wie das Testresultat dargestellt wurde.
Item 31	Die Mitarbeiter und Mitarbeiterinnen der Firma Multicheck schufen während

	des Auswahltages eine angenehme Atmosphäre.
Item 32	Ich hatte jederzeit das Gefühl, Fragen (über das Auswahlverfahren) stellen zu können.
Item 33	Die Inhalte des Auswahlverfahrens schienen mir für die Situation angemessen.
Item 34	Ich bin zufrieden mit der Kommunikation während des Bewerbungsprozesses.

Faktorstruktur der Prozeduralen Gerechtigkeit resp. Zuordnung der Items zu Skalen

Item 1	Berufsbezogenheit Vorhersage
Item 2	Information
Item 3	Leistungsdarstellung
Item 4	Offenheit
Item 5	Konsistenz
Item 6	Behandlung
Item 7	Kommunikation
Item 8	Angemessenheit
Item 9	Berufsbezogenheit Inhalt
Item 10	Feedback
Item 11	Leistungsdarstellung
Item 12	Offenheit
Item 13	Behandlung
Item 14	Angemessenheit
Item 15	Berufsbezogenheit Vorhersage
Item 16	Information
Item 17	Konsistenz
Item 18	Behandlung
Item 19	Kommunikation
Item 20	Feedback
Item 21	Offenheit
Item 22	Behandlung
Item 23	Kommunikation
Item 24	Leistungsdarstellung
Item 25	Offenheit
Item 26	Berufsbezogenheit Inhalt
Item 27	Information
Item 28	Leistungsdarstellung
Item 29	Konsistenz
Item 30	Feedback
Item 31	Behandlung
Item 32	Kommunikation
Item 33	Angemessenheit
Item 34	Kommunikation

## 7. D) Exkurs: Was ist subjektive Fairness?

**Tabelle 9: Schrittweise Regression für das Kriterium Fairness des Multicheck durch Facetten der Sozialen Validität und Prozeduralen Gerechtigkeit**

Modell	R	R <sup>2</sup>	R <sup>2</sup> <sub>korr</sub>	Inkrement in R <sup>2</sup>	Prädiktoren (stand. Beta-Gewichte)
Modell 1	.66	.43	.43	.43	Möglichkeit der Leistungsdarstellung (.66)
Modell 2	.72	.51	.50	.08	Möglichkeit der Leistungsdarstellung (.39), Transparenz (.39)
Modell 3	.73	.53	.53	.02	Möglichkeit der Leistungsdarstellung (.30), Transparenz (.37), Angemessenheit der Fragen (.18)
Modell 4	.74	.55	.54	.01	Möglichkeit der Leistungsdarstellung (.23), Transparenz (.35), Angemessenheit der Fragen (.16), Feedback (.15)

Anmerkungen: für alle R p<.01, zweiseitige Testung, N=221.

**Tabelle 10: Schrittweise Regression für das Kriterium Fairness des Multicheck durch Facetten der Sozialen Validität**

Modell	R	R <sup>2</sup>	R <sup>2</sup> <sub>korr</sub>	Inkrement in R <sup>2</sup>	Prädiktoren (stand. Beta-Gewichte)
Modell 1	.70	.49	.49	.49	Transparenz (.70)
Modell 2	.74	.54	.54	.06	Transparenz (.52), Feedback (.30)
Modell 3	.75	.56	.56	.02	Transparenz (.42), Feedback (.27), Information (.18)
Modell 4	.75	.57	.57	.01	Transparenz (.40), Feedback (.24), Information (.16), Kontrolle (.09)

Anmerkungen: für alle R p<.01, zweiseitige Testung, N=1269.

**Tabelle 11: Schrittweise Regression für das Kriterium Fairness des Multicheck durch Facetten der Prozeduralen Gerechtigkeit**

Modell	R	R <sup>2</sup>	R <sup>2</sup> <sub>korr</sub>	Inkrement in R <sup>2</sup>	Prädiktoren (stand. Beta-Gewichte)
Modell 1	.65	.43	.43	.43	Möglichkeit der Leistungsdarstellung (.65)
Modell 2	.68	.46	.45	.03	Möglichkeit der Leistungsdarstellung (.53), Angemessenheit der Fragen (.21)
Modell 3	.68	.47	.46	.01	Möglichkeit der Leistungsdarstellung (.45), Angemessenheit der Fragen (.18), Information (.15)

Anmerkungen: für alle R p<.01, zweiseitige Testung, N=226.

**7. E) Exkurs: Welche Attribute des einen Instrumentes zur Akzeptanz klärt die des anderen auf?**

**Tabelle 122: Schrittweise Regression für das Kriterium „Soziale Validität“ des Multicheck durch Facetten der Prozeduralen Gerechtigkeit**

Modell	R	R <sup>2</sup>	R <sup>2</sup> <sub>korrr</sub>	Inkrement in R <sup>2</sup>	Prädiktoren (stand. Beta-Gewichte)
Modell 1	.76	.58	.58	.58	Möglichkeit der Leistungsdarstellung (.76)
Modell 2	.78	.60	.60	.02	Möglichkeit der Leistungsdarstellung (.67), Ehrlichkeit und Offenheit (.17)
Modell 3	.79	.62	.61	.02	Möglichkeit der Leistungsdarstellung (.51), Ehrlichkeit und Offenheit (.17), Berufsbezogenheit Inhalt (.21)
Modell 4	.79	.63	.62	.01	Möglichkeit der Leistungsdarstellung (.52), Ehrlichkeit und Offenheit (.29), Berufsbezogenheit Inhalt (.23), Behandlung (-.17)

Anmerkungen: für alle R p<.01, zweiseitige Testung, N=224.

**Tabelle 13: Schrittweise Regression für das Kriterium „Prozedurale Gerechtigkeit“ des Multicheck durch Facetten der Sozialen Validität**

Modell	R	R <sup>2</sup>	R <sup>2</sup> <sub>korrr</sub>	Inkrement in R <sup>2</sup>	Prädiktoren (stand. Beta-Gewichte)
Modell 1	.59	.35	.35	.35	Transparenz (.59)
Modell 2	.65	.43	.42	.08	Transparenz (.43), Feedback (.32)
Modell 3	.67	.45	.44	.02	Transparenz (.37), Feedback (.29), Kontrolle (.17)

Anmerkungen: für alle R p<.01, zweiseitige Testung, N=223.

## 7.F) Vergleich der 4 Facetten der sozialen Validität des Multicheck mit anderen Verfahren

**Tabelle 13: Vergleich der Items zur sozialen Validität für die Einschätzung des Multicheck und des AZUBI-BK**

Item Nr.	Stichprobe Multicheck			Stichprobe AZUBI			Differenz					
	N	M	SD	N	M	SD	M	SF	t	df	sig.	ES
1	1656	3.41	1.25	1175	3.19	1.20	0.22	1.11	5.20	2829	.000	0.20
2	1661	3.29	1.20	1176	3.21	1.25	0.08	1.11	1.90	2835	.030	0.07
3	1688	3.70	1.27	1176	3.59	1.21	0.11	1.12	2.59	2862	.000	0.10
4	1687	3.60	1.23	1174	3.81	1.20	-0.21	1.10	-5.01	2859	.000	0.19
5	1596	3.40	1.30	1174	3.51	1.17	-0.11	1.12	-2.56	2768	.010	0.10
6	1475	3.09	1.05	1174	2.93	1.15	0.16	1.05	3.91	2647	.000	0.15
7	1552	3.32	1.16	1175	3.13	1.16	0.19	1.08	4.56	2725	.000	0.18
8	1626	3.26	1.21	1176	3.16	1.20	0.10	1.10	2.38	2800	.010	0.09
9	1620	3.61	1.31	1174	3.40	1.24	0.21	1.13	4.84	2792	.000	0.19
10	1597	3.29	1.40	1172	2.98	1.28	0.31	1.16	6.94	2767	.000	0.27

Anmerkungen: t-Test für unabhängige Stichproben (2-seitig). SF = Standardfehler, ES = Effektstärke. Items: 1) Ich kann mir vorstellen, dass durch dieses Verfahren berufliche Eignung festgestellt wird. 2) Der Test informiert mich, welche Anforderungen in einem Büroberuf gestellt werden. 3) durch den Test werden meine Stärken sichtbar. 4) durch den Test erkenne ich meine Schwächen. 5) durch eigene Anstrengung konnte ich das Testergebnis beeinflussen. 6) Die einzelnen Testaufgaben sind praxisnah. 7) Mir scheint, durch diesen Test können Personen, die für einen Büroberuf geeignet sind, auch ermittelt werden. 8) Ich kann ungefähr einschätzen, wie ich bei diesem Test abgeschnitten habe. 9) Der Test ist fair. 10) Diesen Test würde ich als Auswahlverfahren für einen Büroberuf akzeptieren.

**Tabelle 14: Vergleich der Items zur sozialen Validität für die Einschätzung des Multicheck und des BIS-4**

Item Nr.	Stichprobe Multicheck			Stichprobe BIS-4			Differenz					
	N	M	SD	N	M	SD	M	SF	t	df	sig.	ES
1	1656	3.41	1.25	1175	2.77	1.19	0.64	1.11	15.2	2829	.000	0.58
2	1661	3.29	1.20	1176	2.49	1.21	0.80	1.10	19.1	2835	.000	0.73
3	1688	3.70	1.27	1176	3.23	1.21	0.47	1.12	11.1	2862	.000	0.42
4	1687	3.60	1.23	1174	3.35	1.21	0.25	1.11	5.9	2859	.000	0.23
5	1596	3.40	1.30	1174	3.23	1.11	0.17	1.10	4.0	2768	.000	0.15
6	1475	3.09	1.05	1174	2.49	1.11	0.60	1.04	14.8	2647	.000	0.58
7	1552	3.32	1.16	1175	2.62	1.18	0.70	1.08	16.7	2725	.000	0.65
8	1626	3.26	1.21	1176	2.98	1.14	0.28	1.09	6.7	2800	.000	0.26
9	1620	3.61	1.31	1174	3.06	1.19	0.55	1.12	12.8	2792	.000	0.49
10	1597	3.29	1.40	1172	2.59	1.20	0.70	1.15	15.9	2767	.000	0.61

Anmerkungen: t-Test für unabhängige Stichproben (2-seitig). SF = Standardfehler, ES = Effektstärke. Items: 1) Ich kann mir vorstellen, dass durch dieses Verfahren berufliche Eignung festgestellt wird. 2) Der Test informiert mich, welche Anforderungen in einem Büroberuf gestellt werden. 3) durch den Test werden meine Stärken sichtbar. 4) durch den Test erkenne ich meine Schwächen. 5) durch eigene Anstrengung konnte ich das Testergebnis beeinflussen. 6) Die einzelnen Testaufgaben sind praxisnah. 7) Mir scheint, durch diesen Test können Personen, die für einen Büroberuf geeignet sind, auch ermittelt werden. 8) Ich kann ungefähr einschätzen, wie ich bei diesem Test abgeschnitten habe. 9) Der Test ist fair. 10) Diesen Test würde ich als Auswahlverfahren für einen Büroberuf akzeptieren.