

Evaluation der Assessment-Aufgabe „Turmbau“ zur Analyse berufsrelevanter Fähigkeiten

MASTERARBEIT

2015

Autor:
Wolfgang Meier

Betreuende Person:
Dr. Katja Pässler

Praxispartner:
impuls - Verein für bessere Chancen im Beruf

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Frage, ob eine neue Aufgabe eines Assessment-Centers (AC) berufskritische Fähigkeiten abbildet. Mit einer Stichprobe von N=42 Arbeitssuchenden wurde die Konstruktvalidität erhoben und die Fähigkeiten Lernen/Merken, Arbeitsplanung, kritische Kontrolle und Misserfolgstoleranz auf Zusammenhänge mit Konstrukten aus der Intelligenzforschung und Persönlichkeitsmerkmalen geprüft. Die Multitrait-Multimethod-Analyse konnte aufzeigen, dass es konvergente Validität zwischen der Arbeitsplanung und schlussfolgernder Denkfähigkeit von $r_s = .446$ ($p < .05$, $N=21$) gibt. Konvergente und diskriminante Validität wurden ansonsten nicht festgestellt. Aus den Ergebnissen lässt sich schliessen, dass es Zusammenhänge zwischen einzelnen Konstrukten gibt. Es wird vermutet, dass bereits während des Lernens und Merkens Arbeitsplanung stattfindet und dabei durch die Misserfolgstoleranz beeinflusst wird. Dem Praxispartner wird empfohlen, diesen Sachverhalt in einer Folgestudie empirisch zu prüfen. Um zu verifizieren, dass diese Zusammenhänge nicht durch eine schlechte Trennschärfe der Konstrukte bedingt sind, müssen die bestehenden Beobachtungseinheiten überprüft werden.

Schlüsselwörter:

Assessment-Center, Konstruktvalidität, Multitrait-Multimethod, Verhaltensbeobachtung, Arbeitsintegration

Zeichen im Bericht: 177'591 (mit Leerzeichen, ab Einleitung, exklusive Anhang)

Abstract

The purpose of this thesis is to examine the construct validity of a new Assessment-Center (AC) exercise based on the multitrait-multimethod. Participants were 42 clients of a non-profit-organisation in the sector of occupational integration. Therefore, professionally relevant capabilities (planning, memorize, frustration tolerance and critical examining) were compared with personality traits (Big-Five) and constructs from intelligence research. Significant correlation of $r_s = .446$ ($p < .05$, $N=21$) was found between the construct planning capability and reasoning. Apart from that, neither convergent nor divergent validity were found. Findings suggest that there is a positive relationship between the capability to memorize and planning capability, which is influenced by frustration tolerance. However, further studies will be needed to confirm these findings. Furthermore, discriminatory power among the constructs must be proved.

Key words:

Assessment-Center, construct validity, multitrait-multimethod, behaviour observation, occupational integration, behavioral observation

Inhaltsverzeichnis

Zusammenfassung	3
Abstract	4
Inhaltsverzeichnis	5
1. Einleitung	1
2. Theoretische Grundlagen	4
2.1 Das Assessment-Center (AC).....	4
2.1.1 Assessments nach MELBA bei impuls	5
2.1.2 Die Turmbauaufgabe von impuls.....	7
2.1.3 Ablauf Turmbauaufgabe von impuls.....	11
2.1.4 Die Qualität von Assessment-Centern	14
2.2 Verhalten und Dispositionen	15
2.2.1 Die Intelligenz	16
2.2.2 Persönlichkeitsmerkmale – Die Big Five	17
2.3 Systematische Verhaltensbeobachtung.....	19
2.3.1 Arten der Verhaltensbeobachtung.....	19
2.3.2 Verhaltensbeobachtung systematisieren.....	23
2.3.3 Beobachtungs- und Beurteilungsfehler	25
2.3.4 Das Beobachtungsraster	27
2.4 Gütekriterien	29
2.4.1 Objektivität.....	29
2.4.2 Reliabilität	31
2.4.3 Validität.....	33
2.4.4 Wechselbeziehung zwischen den Gütekriterien.....	35
2.5 Zwischenfazit	35
2.6 Hypothesen.....	37
3. Methodisches Vorgehen	43
3.1 Operationalisierung.....	43
3.1.1 Objektivität.....	43
3.1.2 Reliabilität	45
3.1.3 Validität.....	45
3.2 Vorbesprechung der Datenerhebung und Pretest	50
3.3 Stichprobe und Datenerhebung	51
3.4 Datenaufbereitung	55
3.5 Datenauswertung.....	55
4. Ergebnisse	59
4.1 Objektivität und Reliabilität.....	59
4.2 Konstruktvalidität.....	61
5. Diskussion.....	68
6. Schlussfolgerung und Ausblick	73

7. Literaturverzeichnis	77
Abbildungsverzeichnis	82
Tabellenverzeichnis	83
Selbstständigkeitserklärung	84
Anhang A	85
Anhang B	86
Anhang C	87
Anhang D	88
Anhang E	89
Anhang F	91
Anhang G	92

1. Einleitung

Impuls - Verein für bessere Chancen im Beruf - ist eine Non-Profit Organisation im Bereich der Arbeitsintegration mit Standorten in Ibach und Lachen im Kanton Schwyz. Der Name *impuls* steht für sinnvolle Beschäftigung bei geregelter Tagesstruktur, für individuelle Förder- und Weiterbildungsprogramme und für hochwertige Produkte, die *impuls* für den Gross- und Fachhandel herstellt. Bei *impuls* stehen ausgewählte Arbeitsplätze in der Holzproduktion, Logistik, Hauswirtschaft sowie im Bereich Empfang und Bürodienstleistung zur Verfügung. In diesen können Erwerbssuchende während einer Dauer von drei bis sechs Monaten Berufspraxis (wieder)erlangen. Neben dem Angebot der praxisorientierten Tagesstruktur am Arbeitsplatz berät *impuls* Personen bei der Stellensuche und bietet berufsvorbereitende und –qualifizierende Förder- und Weiterbildungsprogramme wie z.B. Deutsch- oder Staplerfahrerкурse an. Die Angebote von *impuls* richten sich an stellensuchende Personen (RAV), Sozialhilfeempfänger¹ (Sozialämter), IV-Versicherte (IV-Stellen) sowie anerkannte Flüchtlinge und vorläufig Aufgenommene (Amt für Migration) und haben alle die Stärkung der Leistungsfähigkeit von arbeitssuchenden Personen zum Ziel. Dadurch wird bei Erwerbssuchenden die Chance auf einen erfolgreichen Wiedereinstieg ins Erwerbsleben erhöht. Neben den genannten Angeboten führt *impuls* berufseignungsdiagnostische Abklärungen in Form von Assessments (standardisierte Abklärungen) durch. In diesen Assessments arbeiten Personen während drei Wochen in den Arbeitsplätzen von *impuls* im Bereich der Holzproduktion, Hauswirtschaft und im Büro. Dabei werden sie von qualifizierten Beobachtenden von *impuls* anhand eines Beobachtungsrasters nach MELBA auf 29 Schlüsselqualifikationen hin beobachtet und beurteilt. Dadurch können für zuweisende Stellen konkrete Aussagen zur Erwerbs- sowie Arbeitsfähigkeit der Teilnehmenden gemacht und Entwicklungspunkte aufgezeigt werden (Impuls, 2015).

Zurzeit lassen sich mit dem Assessment bei den Teilnehmenden nur 25 der gesamt- haft 29 berufsrelevanten Fähigkeiten bei der Verrichtung der in den verschiedenen Arbeitsplätzen anfallenden Tätigkeiten beobachten. Nicht beobachtbar und beurteilbar sind die Kriterien *kritische Kontrollfähigkeit*, *Lern-/Merkfähigkeit*, *Fähigkeit zur Misserfolgstoleranz* und *Arbeitsplanungsfähigkeit*. Um ein ganzheitliches Bild über die berufsrelevanten Fähigkeiten von Personen zu erlangen, ist es wichtig, dass

¹ In dieser Arbeit wird der Einfachheit halber oftmals die männliche Form verwendet. Die weibliche Form ist selbstverständlich immer mit eingeschlossen.

auch diese Fähigkeiten erhoben werden. Deshalb erweitert *impuls* das Assessment nun um eine selbst konzipierte Übung - der sogenannten Turmbauaufgabe. Bei dieser Übung bauen Teilnehmende mit Hilfe einer Anleitung nach bestimmten Kriterien einen Holzturm aus unterschiedlich grossen Holzleisten. Aufgrund der Aufgabenstellung sollen dabei die vier Fähigkeiten im Sinne von Verhaltensweisen sichtbar werden, wodurch auf die Ausprägung der jeweiligen Fähigkeit geschlossen werden kann. *Impuls* legt grossen Wert auf seriöse Assessments. Denn Assessment-Berichte zuhanden der zuweisenden Stellen, die Entwicklungsempfehlungen beinhalten, können weitreichende Konsequenzen für die Teilnehmenden haben. So kann eine zuweisende Stelle wie etwa die IV aus dem Bericht von *impuls* für den Teilnehmenden durchaus existenzielle Entscheide ableiten wie etwa die Kürzung von Geldleistungen oder gar deren Einstellung. Dass *impuls* deshalb bei der Durchführung der Assessments eine grosse soziale Verantwortung zukommt, liegt nahe. Das Assessment qualitativ hochstehend zu gestalten, ist unerlässlich für *impuls*. Aus diesem Grund hat sich *impuls* zum Ziel gesetzt, die neu konzipierte Turmbauaufgabe zu evaluieren, um herauszufinden, ob sich mit der Turmbauaufgabe die genannten Fähigkeiten beurteilen lassen. Daraus lässt sich für die vorliegende Arbeit die folgende Fragestellung ableiten:

Lassen sich mit der Turmbauaufgabe die vier Fähigkeiten Arbeitsplanung, kritische Kontrolle, Merken/Lernen und Misserfolgstoleranz abbilden?

Um die Frage zu beantworten, wird die Turmbauaufgabe auf die drei Hauptgütekriterien Validität, Reliabilität und Objektivität untersucht.

Aufbau der Arbeit

Zu Beginn dieser Arbeit wird im Kapitel 2 erklärt, was unter einem Assessment Center zu verstehen ist und welches Prinzip dahinter steckt. In diesem Zusammenhang wird denn auch das Assessment Center von *impuls* und die neu konzipierte Assessment-Aufgabe „Turmbau“ vorgestellt. In einem Exkurs wird das Problem der zunehmenden mangelnden Qualität bei Assessment-Centern beleuchtet. Da sich Assessment Center unter anderem der systematischen Verhaltensbeobachtung bedienen, wird erklärt, was unter Verhalten zu verstehen ist und was das Verhalten bestimmt. Daher wird auf den eigenschaftsorientierten Ansatz der psychologischen Eignungsdiagnostik Bezug genommen und zwei Konstrukte vorgestellt, die Einfluss auf das Verhalten haben - die Intelligenz und die Persönlichkeitsmerkmale. Es wird aufge-

zeigt, welche Arten von Verhaltensbeobachtungen es gibt und wie die Verhaltensbeobachtung systematisiert werden kann, um Beobachtungs- und Beurteilungsfehler zu vermeiden. Der Abschnitt 2.4 beschäftigt sich mit den Gütekriterien. In diesem Abschnitt der Arbeit wird beschrieben, was unter den einzelnen Gütekriterien zu verstehen ist und welche Arten es gibt. Darauf folgt ein erstes Fazit und es werden Hypothesen abgeleitet, mit denen die (Konstrukt)Validität der Turmbauaufgabe überprüft wird. Im methodischen Teil der Arbeit im Kapitel 3 wird aufgezeigt, wie die Hypothesen operationalisiert und wie bei der Datenerhebung vorgegangen wurde. Danach werden im Kapitel 4 die Ergebnisse der Erhebung dargelegt, die anschliessend im Kapitel 5 diskutiert werden. Zum Schluss wird ein Fazit gezogen, die Fragestellung beantwortet und es werden praktische Implikationen für *impuls* gegeben.

2. Theoretische Grundlagen

Nach Amelang und Schmidt-Atzert (2006, zitiert nach Schmid-Atzert & Amelang, 2012) liegt die Aufgabe der psychologischen Diagnostik darin, Merkmale und Veränderungen so zu erfassen, „dass hinlänglich präzise Vorhersagen künftigen Verhaltens und Erlebens sowie deren eventuellen Veränderung in definierten Situation möglich werden“ (S. 2). Ein eignungsdiagnostisches Verfahren, das in der Praxis immer mehr eingesetzt wird (dem sich auch *impuls* bedient) und das im Zentrum dieser Arbeit steht, ist das sogenannte Assessment-Center.

2.1 Das Assessment-Center (AC)

Ein Assessment-Center ist ein Beurteilungsverfahren, das vor allem in der Personalpsychologie eingesetzt wird, um die Eignung von Bewerbenden festzustellen (Schmidt-Atzert & Amelang, 2012). Es besteht aus mehreren eigens konzipierten Assessment-Übungen resp. Aufgaben wie zum Beispiel Gruppendiskussionen, Rollenspielen, Präsentationen, Postkorbübungen oder etwa Übungen wie die Turmbaufgabe bei *impuls*. Die Aufgaben werden von den Teilnehmenden einzeln oder in Gruppen bearbeitet, weshalb nach Gruppen- oder Einzelassessments unterschieden wird (Moosbrugger & Rauch, 2005). Dabei sind die Übungen so konzipiert, dass sie den Anforderungen der späteren Tätigkeiten möglichst ähnlich sind. Folglich sollte dem AC – aber auch bei eignungsdiagnostischen Verfahren generell – immer eine Anforderungsanalyse vorausgehen (Schmidt-Atzert & Amelang, 2012). Das Prinzip des AC's liegt darin, dass Bewerbende während der Bearbeitung der Aufgaben von mindestens zwei Assessoren bzw. Beobachtenden auf Verhaltensweisen hin beobachtet und beurteilt werden, woraus sich auf die Ausprägung eines dem Verhalten übergeordneten Merkmals schliessen lässt (Moosbrugger & Rauch, 2005).

Meist werden im Rahmen von AC's neben den eigens konzipierten Aufgaben weitere Verfahren wie etwa das Interview, Analysen der Bewerbungsunterlagen, Leistungs- und Persönlichkeitstest etc. eingesetzt. Dies führt zu einem ganzheitlichen Bild über die Fähigkeiten des Teilnehmenden, was ein grosser Mehrwert des Verfahrens ist (Lissmann, 2013; Schmidt-Atzert & Amelang, 2012). Die Ergebnisse aus den einzelnen Aufgaben und eingesetzten Verfahren werden zum Schluss zu einer Gesamtbeurteilung zusammengefasst (Moosbrugger & Rauch, 2005). Auf dieser Grundlage wird beurteilt, ob sich eine Person etwa für eine bestimmte Stelle eignet, woraus sich wiederum ein Entscheid für die Personalselektion ableiten lässt.

AC's werden heute längst nicht mehr einzig in der Personalauswahl eingesetzt, sondern wie etwa bei *impuls* zur Potenzialabklärung im Rahmen der beruflichen Rehabilitation (Schuler, 2007). Bei diesen sogenannten Potenzial-Assessments werden bei Erwerbssuchenden Stärken und Schwächen analysiert. Dadurch ergeben sich Hinweise auf den Entwicklungsbedarf sowie die Entwicklungs-/Einsatzmöglichkeiten (Schmidt-Atzert & Amelang, 2012).

2.1.1 Assessments nach MELBA bei impuls

Wie in der Einleitung erwähnt bietet *impuls* Dienstleistungen im Bereich der Berufspraxis und beruflichen Qualifizierung sowie Beratung an. Daneben führt *impuls* berufseignungsdiagnostische Abklärungen² in Form von 15-tägigen Potenzialassessments resp. Entwicklungs-/Standort-Assessments durch (*impuls*, 2015; swiss assessment, 2015). Bei diesen standardisierten Abklärungen werden arbeitssuchende Personen bei der Verrichtung von Arbeiten durch Mitarbeitende von *impuls*, die eine fünftägige Beobachterschulung durchlaufen haben, in ausgewählten Arbeitsplätzen auf 29 Schlüsselfähigkeiten hin beobachtet und beurteilt. Zuweisende Stellen erhalten einen ausführlichen und nachvollziehbaren Bericht, der Aussagen zur Arbeitsfähigkeit und Empfehlungen für die Entwicklung der Teilnehmenden enthält (*Impuls*, 2015).

Zur Beobachtung und Beurteilung der Fähigkeiten bedient sich *impuls* dem Profilvergleichsverfahren MELBA. MELBA - Merkmalprofile zur Eingliederung Leistungsgewandelter und Behinderter in Arbeit - ist ein von Kleffmann, Weinmann, Föhres und Müller (1997) inhaltlich validiertes und praktisch erprobtes Verfahren, das ursprünglich dazu entwickelt wurde, Personen mit einer Behinderung und psychischer Beeinträchtigung ihren Fähigkeiten und den Anforderungen von Arbeitsplätzen entsprechend zu platzieren. Das Verfahren besteht aus zwei Profilbögen: einem Dokumentationsbogen für Tätigkeitsanforderungen und einem damit kompatiblen Dokumentationsbogen für Arbeitsfähigkeiten. Ein Dokumentationsbogen umfasst 29 Merkmale (vgl. Anhang A). Bei den Merkmalen handelt es sich um Schlüsselqualifikationen, „die eine hinreichende und umfassende Beschreibung der tätigkeitsrelevanten psychischen Fähigkeiten einer Person sowie der behinderungsrelevanten psychischen

² Nach Schuler und Höft (2006, zitiert nach Thunsdorff & Schmitt, 2013) wird mit Berufseignungsdiagnostik „die Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zum Zwecke eignungsbezogener Erfolgsprognosen und Entscheidungshilfen im beruflichen Kontext“ gemeint (S. 128).

Anforderungen einer Tätigkeit erlauben“ (Kleffmann et al., 1997, S. 9). Die 29 Merkmale sind das Ergebnis einer aufwendigen Literaturrecherche und Analyse bestehender Verfahren, die sich mit kritischen Fähigkeiten bei psychisch Beeinträchtigten befassen. Aus der Analyse resultierte eine Merkmalskatalog mit folgenden Fähigkeiten:

Tabelle 1: Schlüsselfähigkeiten nach MELBA (Kleffmann et al., 1997, S. 10)

Kognitive Merkmale	Soziale Merkmale	Merkmale zur Art der Arbeitsausführung	Psychomotorische Merkmale	Kulturtechniken/Kommunikation
Arbeitsplanung	Durchsetzung	Ausdauer	Antrieb	Lesen
Auffassung	Führungsfähigkeit	Mißerfolgstoleranz	Feinmotorik	Rechnen
Aufmerksamkeit	Kontaktfähigkeit	Kritische Kontrolle	Reaktionsgeschwindigkeit	Schreiben
Konzentration	Kritikfähigkeit	Ordnungsbereitschaft		Sprechen
Lernen/Merken	Kritisierbarkeit	Pünktlichkeit		
Problemlösen	Teamarbeit	Selbständigkeit		
Umstellung		Sorgfalt		
Vorstellung		Verantwortung		

Jedes Merkmal ist genau definiert. Die Definitionen haben Kleffmann et al. (1997) in Expertenrunden und in Anlehnung an die einschlägige Literatur zu den einzelnen Merkmalen zusammengetragen. Pro Merkmal gibt es sogenannte Leitfragen sowie Hilfestellungen. Mit diesen können die Beobachtenden während der Assessment-Aufgabe Verhaltensweisen identifizieren, diese einer Fähigkeit zuordnen und beurteilen. Die Hilfestellungen stellen die Definition für die einzelnen Skalenwerte dar. Denn die Profilbögen von MELBA verwenden zur Bewertung von Fähigkeiten Ratingskalen von 1=*stark eingeschränkte Fähigkeit* bis 5=*weit überdurchschnittliche Fähigkeit*. Auch die Definition der Leitfragen und Hilfestellungen sind Resultat dieser Expertenrunde und Literaturrecherche.

Die Definitionen zu allen 29 Fähigkeiten samt Leitfragen und Hilfestellungen finden sich in Handbuch III von Föhres, Kleffmann, Strutz und Weinmann (2011). Jene vier Fähigkeiten, die es im Rahmen dieser Arbeit zu untersuchen gilt, werden im nächsten Abschnitt 2.1.2 genauer erklärt.

Ob die beiden Dokumentationsbögen als Profilvergleichsverfahren eingesetzt und somit das Anforderungsprofil einer bestimmten Stelle erhoben und mit den erhobenen Fähigkeiten einer Person verglichen werden, ist abhängig vom Anwendungskontext. Die Profil- resp. Dokumentationsbögen können auch einzeln eingesetzt werden. Da es bei sogenannten Potenzialassessments darum geht, die Fähigkeiten von Per-

sonen zu erheben, verwendet *impuls* „nur“ die Fähigkeitsprofile nach MELBA. Deshalb wird in dieser Arbeit nicht weiter auf die Möglichkeit der Profilvergleiche eingegangen.

MELBA ist kein eigenständiges Diagnoseinstrument, sondern ein Dokumentationsbogen, auf dem Ausprägungen tätigkeitsrelevanter Fähigkeitsmerkmale vermerkt werden können. Die Anwendungsmöglichkeiten der Dokumentationsbögen für Arbeitsfähigkeiten sind vielfältig und werden in der Regel in Kombination mit diagnostischen Erhebungsverfahren eingesetzt, zum Beispiel im Rahmen eines Gesprächs mit dem Klienten oder dessen Angehörigen (Fremdanamnese), Testaufgaben oder psychometrischen Verfahren usw. Die Dokumentationsbögen dienen dazu, die Ergebnisse aus den Verfahren, mit welchen die Fähigkeiten erfasst werden, festzuhalten. Folglich ist ein Dokumentationsbogen nicht als Beobachtungsraster oder dergleichen zu verstehen. Dazu dient als Vorlage das Handbuch III, in dem die Definitionen, Leitfragen und Hilfestellungen ausführlich beschrieben sind. Ein praktikables Beobachtungsraster in Anlehnung an das Handbuch III von Föhres et al. (2011) zu kreieren, liegt somit nach Kleffmann et al. (1997) in der Verantwortung des jeweiligen Anwenders.

2.1.2 Die Turmbauaufgabe von *impuls*

25 der total 29 Fähigkeiten nach MELBA lassen sich gemäss Martin Birchler, Assessmentverantwortlicher von *impuls*, in den verschiedenen Arbeitsplätzen momentan abbilden. Vier Fähigkeiten können jedoch nicht beobachtet werden, die Fähigkeit zur *kritischen Kontrolle*, *Arbeitsplanung*, *Misserfolgstoleranz* und *Lernen/Merken*. Die Definition der vier Merkmale sowie die dazugehörigen Leitfragen und Hilfestellungen, die den Beobachtenden bei der Identifizierung und Beurteilung der einzelnen Schlüsselqualifikationen unterstützen, finden sich nachfolgend in den Tabellen 1-4 detailliert aufgeführt.

Tabelle 2: Schlüsselqualifikation Kritische Kontrolle (Föhres et al., 2011, S. 13)

KRITISCHE KONTROLLE	
Fähigkeitsdefinition: Kritische Kontrolle ist die Fähigkeit, mit dem Arbeitsprozess in Zusammenhang stehende eigene Verhaltensweisen und das selbst erbrachte Arbeitsergebnis auf sachbezogene Richtigkeit hin prüfen und bewerten zu können.	
Fragen zur Fähigkeitsanalyse:	
<ul style="list-style-type: none"> • Kann die Person ihr Arbeitsergebnis richtig einschätzen? Auch bei schwierigen Arbeiten • Kontrolliert sie sich selbst? • Kann sie im Hinblick auf das Arbeitsziel (-ergebnis) realistische Massstäbe erkennen und / oder entwickeln? • Kann sie sich selbst eigene Fehler eingestehen? 	
Hilfestellung zur Beurteilung der Merkmalsausprägung:	
1	<ul style="list-style-type: none"> • benötigt ständig Kontrolle von aussen • schätzt das eigene Arbeitsergebnis nur nach Aufforderung und dann in der Regel falsch ein • erkennt eigene Fehler nicht an oder übernimmt unreflektiert die Einschätzung anderer
2	<ul style="list-style-type: none"> • benötigt von Zeit zu Zeit Kontrolle von aussen • neigt dazu, sich tendenziell zu über- oder unterschätzen • kontrolliert zwanghaft
3	<ul style="list-style-type: none"> • kontrolliert sich bei der Arbeit immer wieder selbst, ohne zusätzliche Aufforderung • liegt in der Einschätzung der eigenen Leistung tendenziell richtig • unüberschaubare Arbeitsergebnisse werden manchmal unrealistisch eingeschätzt
4	<ul style="list-style-type: none"> • kann auch wenig anschauliche Arbeitsergebnisse richtig einschätzen und an realistischen Massstäben beurteilen • kontrolliert die eigene Arbeit selbstständig und zuverlässig
5	<ul style="list-style-type: none"> • kann auch sehr schwer kalkulierbare Arbeitsergebnisse richtig einschätzen und an realistischen Massstäben beurteilen • kontrolliert die eigene Arbeit selbstständig und fortlaufend zuverlässig

Tabelle 3: Schlüsselqualifikation Arbeitsplanung (Föhres et al., 2011, S. 2)

ARBEITSPLANUNG	
Fähigkeitsdefinition: Die Fähigkeit zur Arbeitsplanung besteht darin, eine gestellte Arbeitsaufgabe unter Berücksichtigung der technischen, administrativen und personellen Bedingungen im Hinblick auf ein optimales Zusammenwirken der einzelnen Elemente zu einem effektiven Ergebnis gliedern und strukturieren zu können.	
Fragen zur Fähigkeitsanalyse:	
<ul style="list-style-type: none"> • Ist die Person in der Lage, verschiedene Arbeitsaufgaben zu koordinieren? • Wie verhält sie sich bei einer Tätigkeit, bei der sie viele verschiedene Arbeitsschritte beachten muss, die nicht alle vorgegeben sind? • Muss die Arbeitsaufgabe extrem vorstrukturiert sein? • Ist sie in der Lage, Wichtiges von Unwichtigem zu trennen? • Kann sie einzelne Arbeitsschritte als solche erkennen? • Ist sie in der Lage, sich technische, administrative und / oder personelle Ressourcen zunutze zu machen? 	
Hilfestellung zur Beurteilung der Merkmalsausprägung:	
1	<ul style="list-style-type: none"> • Kann vorstrukturierte Tätigkeiten ausführen, bei denen nichts selbst geplant werden muss • Kann Abweichungen im Arbeitsablauf erkennen und darauf reagieren, wenn die Handlungsalternative vorher festgelegt worden ist
2	<ul style="list-style-type: none"> • Kann bei einfachen Tätigkeiten, die im Wesentlichen vorstrukturiert sind, verschiedene Arbeitsschritte koordinieren • Kann bei leichten Abweichungen von der Routine die unmittelbaren Konsequenzen erkennen und zwischen vorgegebenen Handlungsalternativen die sinnvollere erkennen
3	<ul style="list-style-type: none"> • Kann verschiedene Arbeitsschritte so koordinieren, dass sie zu einem effektiven Arbeitser-

	gebnis führen <ul style="list-style-type: none"> • Kann bei auftretenden Abweichungen auch längerfristige Konsequenzen erkennen und aus verschiedenen Handlungsalternativen eine effektive Strategie entwickeln
4	<ul style="list-style-type: none"> • Kann eine Vielzahl von Elementen, die zur Erledigung einer komplexen Arbeitsaufgabe gehören, effektiv koordinieren • Kann unter Zeitdruck auch langfristig Konsequenzen erkennen und aus verschiedenen Handlungsalternativen eine effektive Strategie entwickeln
5	<ul style="list-style-type: none"> • Kann eine Vielzahl von Elementen, deren gegenseitiger Einfluss schwer vorhersagbar ist, effektiv koordinieren • Kann auch die unterschiedlichsten und in ihren Interaktionen ausserordentlichen differenzierten Elemente so koordinieren, dass auch langfristig ein optimales Ergebnis erzielt wird • Findet auch unter Zeitdruck die effektivste Strategie zur Lösung der Arbeitsaufgabe

Tabelle 4: Schlüsselqualifikation Misserfolgstoleranz (Föhres et al., 2011, S. 16)

MISSERFOLGSTOLERANZ

Fähigkeitsdefinition:

Misserfolgstoleranz ist die Fähigkeit, sich einer Arbeitsaufgabe auch dann stetig zuwenden zu können, wenn ihre Erfüllung nicht gesichert ist.

Fragen zur Fähigkeitsanalyse:

- Kann die Person mit einer Arbeitsaufgabe beginnen, auch wenn sie weiss, dass ein erfolgreicher Abschluss nicht in jedem Fall sichergestellt ist?
- Ist ihr Engagement in einem solchen Fall von Beginn an eingeschränkt?
- Gerät sie in Rage oder wird sie deprimiert, wenn sie eine Aufgabe nicht erfolgreich beenden kann?
- Kann sie mit der Erledigung einer Arbeitsaufgabe erneut beginnen, wenn diese Aufgabe zuvor nicht erfolgreich abgeschlossen werden konnte?

Hilfestellung zur Beurteilung der Merkmalsausprägung:

1	<ul style="list-style-type: none"> • Kann eindeutige Misserfolge als solche erkennen, erwartet aber auch Misserfolge, die nicht eintreten • Reagiert sehr empfindlich auf eingetretenen Misserfolg (z.B. grosser Ärger, Wut) • Kann erfolglos ausgeführte Arbeiten erneut beginnen, wenn dies mit geringem Aufwand möglich ist und ein erneuter Misserfolg nicht wahrscheinlich ist
2	<ul style="list-style-type: none"> • Kann deutliche Misserfolge erkennen und manchmal vorwegnehmen, tut dies oft auch unberechtigt • Reagiert mit Frustration, Ärger oder Nervosität, wenn sich während der Arbeit ein Misserfolg abzeichnet • Kann erfolglos ausgeführte Arbeiten dann erneut ausführen, wenn dies mit geringem Aufwand möglich ist
3	<ul style="list-style-type: none"> • Kann eine Aufgabe auch dann unbeeinträchtigt weiterführen, wenn der Erfolg nicht gesichert ist • Versucht mehrmals, einen erfolgreichen Abschluss zu erreichen • Zeigt Engagement, um ein Arbeitsergebnis bei sich abzeichnendem Misserfolg zu Ende zu führen
4	<ul style="list-style-type: none"> • Kann eine Aufgabe auch dann unbeeinträchtigt erneut beginnen, wenn die Aufgabe bereits mehrfach erfolglos bearbeitet worden ist • Zeigt Engagement auch nach mehreren erfolglosen Versuchen, um ein Arbeitsergebnis zu Ende zu führen
5	<ul style="list-style-type: none"> • Kann eine mehrfach erfolglos bearbeitete Aufgabe auch dann unbeeinträchtigt erneut beginnen und zu Ende führen, wenn die Bearbeitung sehr viel Aufwand verlangt • Zeigt ein hohes Mass an Engagement auch bei Aufgaben, bei denen ein Misserfolg wahrscheinlicher als ein Gelingen ist

Tabelle 5: Schlüsselqualifikation Lernen/Merken (Föhres et al., 2011, S. 14)

LERNEN/MERKEN	
Fähigkeitsdefinition:	
Lernen/Merken ist die Fähigkeit, arbeitsrelevante Informationen auffassen, im Gedächtnis speichern und zu einem gegebenen Zeitpunkt verfügbar machen zu können.	
Anmerkung: Lernen/Merken ist bestimmt durch die jeweilige Ausprägung der Komponenten Lerngeschwindigkeit, Umfang und Komplexität des Materials.	
Fragen zur Fähigkeitsanalyse:	
<ul style="list-style-type: none"> • Ist die Person in der Lage, sich Arbeitsinhalte über einen langen Zeitraum zu merken? • Ist sie in der Lage, sich schnell an etwas zu erinnern, was sie vor langer Zeit gelernt hat? • Wenn der Person gerade etwas erklärt worden ist, vergisst sie das schnell wieder oder erst nach langer Zeit? • Kann sie sich viele Dinge gleichzeitig merken? 	
Hilfestellung zur Beurteilung der Merkmalsausprägung:	
1	<ul style="list-style-type: none"> • kann sich einfache, kurze Anweisungen merken, die sich auf sich wiederholende Tätigkeiten beziehen • kann sich einfache, sich wiederholende Tätigkeiten nach einiger Zeit merken • vergisst neu Gelerntes wieder, wenn es nicht ständig aktiviert wird
2	<ul style="list-style-type: none"> • kann sich einfache Anweisungen merken • erlernt neue Tätigkeiten nur langsam • kann sich nur schwer mehrere Dinge gleichzeitig merken • vergisst schnell, wenn neue Gelerntes längere Zeit nicht angewendet oder abgerufen wird
3	<ul style="list-style-type: none"> • kann sich auch längere Anweisungen merken • erlernt neue Tätigkeiten nach einigen Durchgängen in angemessener Zeit • kann sich mehrere Dinge gleichzeitig merken • vergisst einiges, wenn neu Gelerntes längere Zeit nicht angewendet oder abgerufen wird
4	<ul style="list-style-type: none"> • kann sich auch umfangreiche Anweisungen, bei denen mehrere Aspekte berücksichtigt werden müssen, gut merken • erlernt neue Tätigkeiten meist sehr schnell • kann sich auch über längere Zeit mehrere Dinge gleichzeitig merken • vergisst wenig, wenn neu Gelerntes längere Zeit nicht angewendet oder abgerufen wird
5	<ul style="list-style-type: none"> • kann sich sehr komplexe Anweisungen, bei denen viele Aspekte berücksichtigt werden müssen, immer gut merken • erlernt neue Tätigkeiten auf Anhieb • kann sich auch über lange Zeit sehr viele Dinge gleichzeitig merken • vergisst wenig, wenn neu Gelerntes über sehr lange Zeit nicht angewendet oder abgerufen wird.

Um die vier oben aufgeführten Merkmale bei den Assessment-Teilnehmenden beurteilen zu können, möchte *impuls* das AC um eine Assessment-Aufgabe erweitern – der sogenannten Turmbaufgabe. Die Turmbaufgabe wurde von Mitarbeitenden von *impuls* entwickelt.

2.1.3 Ablauf Turmbauaufgabe von impuls

Wie es die Bezeichnung vermuten lässt, geht es bei der Aufgabe darum, dass eine Person aus unterschiedlich grossen Holzleisten einen Holzturm errichtet. Dazu erhält sie zu Beginn der Aufgabe eine Anleitung, die ihr vorgibt, wie der Turm im Endzustand aussehen und nach welchen Kriterien der Turm gebaut werden soll (siehe Abbildung 1).

1. Was wird beobachtet:

Mit dem Turmbau beobachten wir losgelöst vom Assessment-Arbeitsplatz folgende Schlüsselqualifikationen:

- Arbeitsplanung
- Kritische Kontrolle
- Lernen / Merken
- Misserfolgstoleranz

2. Auftrag:

1. Vorbereitung und lesen der Anleitung Turmbau (inkl. Rückfragen)
Zeit: 10 Minuten
2. Pause
Zeit: 15 Minuten
3. Abfragen des Auftrages
Zeit: ca. 5-10 Minuten
4. Bauen Sie einen Turm mit Hilfe der Stückzahlliste, den Holzteilen und der Skizze-Turm
Zeit: 30 Minuten

Stückliste Holzteile:

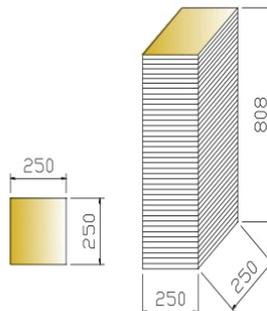
Holzteile:	Dicke / Länge	20 / 200 mm	140 Stück
Holzteile:	Dicke / Länge	20 / 150 mm	35 Stück
Holzteile:	Dicke / Länge	20 / 100 mm	35 Stück
Holzteile:	Dicke / Länge	20 / 50 mm	140 Stück
Holzteile:	Dicke / Länge	18 / 200 mm	24 Stück
Holzteile:	Dicke / Länge	18 / 150 mm	6 Stück
Holzteile:	Dicke / Länge	18 / 100 mm	6 Stück
Holzteile:	Dicke / Länge	18 / 50 mm	24 Stück

Bitte beachten Sie dabei folgendes:

- Beim Turm dürfen sich innen und aussen keine Hohlräume ergeben und sichtbar sein
- Es dürfen keine vorstehende Holzteile sichtbar sein
- Der fertig erbaute Turm soll in sich stabil sein
- Der Turm wird alleine und ohne Mithilfe erbaut
- Während des Baus sind keine Rückfragen erlaubt

Skizze-Turm:

Grundriss: 250 mm x 250 mm
Höhe: 808 mm
Lagen: 35 à 20 mm
06 à 18 mm



Material:

- Holzteile gemäss Stückliste
- Grundrissplatte
- Massstab (Meterstab)

Abbildung 1: Anleitung Turmbau (Impuls, 2015)

In Anlehnung an die Anleitung erklärt ein Moderator den Teilnehmenden während fünf Minuten die Aufgabe. Danach hat dieser zehn Minuten Zeit, die Anleitung zu studieren und Verständnisfragen zu stellen. Nach Ablauf dieser zehn Minuten gibt es eine 15-minütige Pause, über die der Teilnehmende frei verfügen kann. Nach der Pause wird der Teilnehmende durch den Moderator aufgefordert, die in der Anleitung enthaltenen Informationen und insbesondere den Arbeitsauftrag so gut wie möglich verbal wiederzugeben. Dabei wird der Teilnehmende durch zwei Assessoren auf die *Lern-/Merkfähigkeit* hin beobachtet und beurteilt. Dazu verwenden sie ein von *impuls* in Anlehnung an das Handbuch III und in diesem Zusammenhang an die obig vorgestellten Definitionen, Leitfragen und Hilfestellungen eigens entwickeltes Beobachtungs-/Bewertungsraster. Die beobachteten Ergebnisse tragen die Assessoren im Beobachtungsraster auf einer sogenannten fünf-Punkte-Ratingskala ein. Danach hat der Teilnehmende 30 Minuten Zeit, um den Turm mit Hilfe der Anleitung zu bauen. Um den Turm zu bauen, liegen dem Teilnehmenden verschiedene unsortierte Holzteile mit verschiedenen Massen und ein Doppelmeter vor. Um den Turm innerhalb von 30 Minuten nach den Vorgaben der Anleitung zu erstellen, bedarf es einer *guten Arbeitsplanung* und *kritischen Kontrolle* der jeweiligen Arbeitsschritte. So müssen die Teilnehmenden mental einen Plan darüber erstellen, welche Holzstücke aneinander gereiht werden müssen, um den Grundriss von 25x25 cm zu erreichen. Ferner müssen die Holzstücke so geschichtet werden, dass keine Hohlräume entstehen und der Turm stabil ist d.h. Holzstücke müssen sortiert, abgemessen werden etc. Um den Turm vorgabengerecht zu erstellen, bedarf es einer fortlaufenden Überprüfung der eigenen Arbeitsschritte hin auf ihre sachbezogene Richtigkeit. Für die Turmerstellung stehen 30 Minuten zur Verfügung. Die Aufgabe ist so konzipiert, dass es für den Teilnehmenden kaum möglich ist, den Turm innerhalb der vorgegebenen Zeit nach allen Qualitätsansprüchen fertigzustellen. Dadurch soll geprüft werden, wie es um die *Misserfolgstoleranz* des Teilnehmenden steht (wird die Person nervös oder aggressiv, wenn sie merkt, dass sie mit dem Turm nicht fertig wird innerhalb der vorgegebenen Zeit).

Wie schon erwähnt wird die Person, welche die Turmbaufaufgabe bearbeitet, während der ganzen Zeit von zwei Assessoren beobachtet. Zum Schluss der Beobachtung nehmen die Beobachtenden aufgrund ihren Beobachtungen eine Beurteilung zu den einzelnen Fähigkeiten vor und tragen diese im nachfolgend abgebildeten Beobachtungsraster ein (siehe Abbildung 2).

Name, Vorname:

02 Arbeitsplanung Bei der Wertung halten wir uns an die Melba-Fragestellungen. Beobachtungen auf separaten Blatt aufschreiben

Fragestellungen	MELBA-Wert
<input type="radio"/> Ist die Person in der Lage, verschiedene Arbeitsaufgaben zu koordinieren?	1
<input type="radio"/> Wie verhält sich die Person bei einer Tätigkeit, bei der sie viele verschiedene Arbeitsschritte beachten muss?	2
<input type="radio"/> Ist der Teilnehmende in der Lage, Wichtiges von Unwichtigem zu trennen?	3
<input type="radio"/> Kann die Person einzelne Arbeitsschritte als solche erkennen?	4
<input type="radio"/> Muss die Arbeitsaufgabe extrem vorstrukturiert sein?	5

12 Kritische Kontrolle Bei der Wertung halten wir uns an die Melba-Fragestellungen. Beobachtungen auf separaten Blatt aufschreiben

Fragestellungen	MELBA-Wert
<input type="radio"/> Kann die Person ihr Arbeitsergebnis richtig einschätzen? Auch bei schwierigen Arbeiten?	1
<input type="radio"/> Kontrolliert sie sich selbst?	2
<input type="radio"/> Kann die Person im Hinblick auf das Arbeitsziel(-ergebnis) realistische Massstäbe erkennen und / oder entwickeln?	3
<input type="radio"/> Kann die Person sich selbst eigene Fehler eingestehen?	4
	5

16 Misserfolgstoleranz Bei der Wertung halten wir uns an die Melba-Fragestellungen. Beobachtungen auf separaten Blatt aufschreiben

Fragestellungen	MELBA-Wert
<input type="radio"/> Kann die Person mit einer Arbeitsaufgabe beginnen, auch wenn sie weiss, dass ein erfolgreicher Abschluss nicht in jedem Fall sichergestellt ist?	1
<input type="radio"/> Ist das Engagement in einem solchen Fall von Beginn an eingeschränkt?	2
<input type="radio"/> Gerät die Person in Rage oder wird sie deprimiert, wenn sie eine Aufgabe nicht erfolgreich beenden kann?	3
<input type="radio"/> Kann die Person mit der Erledigung einer Arbeitsaufgabe erneut beginnen, wenn diese Aufgabe zuvor nicht erfolgreich abgeschlossen werden konnte?	4
	5

14 Lernen/ Merken Bei der Wertung halten wir uns an die Melba-Fragestellungen. Beobachtungen auf separaten Blatt aufschreiben

Fragestellungen	MELBA-Wert
<input type="radio"/> Ist die Person in der Lage, sich Arbeitsinhalte über einen langen Zeitraum zu merken (min. 15 Min)	1
<input type="radio"/> Ist die Person in der Lage, sich schnell an etwas zu erinnern, was er vor 15 Min. gelernt hat?	2
<input type="radio"/> Kann die Person sich viele Dinge gleichzeitig merken?	3
	4
	5

Erinnerungspunkte zu Punkt 14 (nach 15. minütiger Pause)

		erfüllt	Nicht erfüllt
Auftrag:	- Bauen sie einen Turm mit Hilfe der Stückzahlliste		
	- Bauen sie einen Turm mit den vorhanden Holzteilen		
	- Die Skizze kann als Vorlage benutzt werden		
	- Zeitdauer 30 Min.		
Zu beachten sind:	- Beim Turm dürfen sich innen und aussen keine Hohlräume ergeben und sichtbar sein		
	- Es dürfen keine vorstehende Holzteile sichtbar sein		
	- Der fertig erbaute Turm soll in sich stabil sein		
Erklärung der Skizze	- Das Aussehen des Turms in eigenen Worten erläutern		
Materialliste:	- Holzteile gemäss Stückliste		
	- Grundrissplatte		
	- Doppelmeter (Massstab)		

Assessor inkl. Melba-Lizenznummer

Ort, Datum, Unterschrift

Abbildung 2: Beobachtungsraster Turmbau (Impuls, 2015)

Als Hinweis für die Beurteilung der Fähigkeit Lernen/Merken sind auf dem Beobachtungsraster die wichtigsten Inhalte der Anleitung als Erinnerungspunkte aufgeführt. Damit kann geprüft und „abgehakt“ werden, an welche Punkte der Anleitung sich der Teilnehmende noch zu erinnern vermag. Daraus lässt sich eine Beurteilung auf der fünf-Punkte-Skala vornehmen.

Pro Fähigkeitsmerkmal sind auf dem von *impuls* entwickelten Beobachtungsraster die Leitfragen aus dem Handbuch III von Föhres et al. (2011) aufgeführt. Die Definitionen und Hilfestellungen sind nicht aufgeführt.

2.1.4 Die Qualität von Assessment-Centern

Das Assessment Center erfährt eine zunehmende Popularität. Eine Befragung von Höft und Obermann (2010, zitiert nach Schmidt-Atzert & Amelang, 2012) hat ergeben, dass von 233 befragten Unternehmen im deutschsprachigen Raum 171 Unternehmen AC's verwenden. Die Verbreitungsquote von rund 70% in Grossunternehmen und grossen Behörden bestätigen auch Kanning und Schuler in einer Studie (2014). Weiter dürfte die Popularität der doch eher kostenintensiveren AC's nicht zuletzt mit der dem Verfahren hohen zugesprochenen prognostischen Validität zusammenhängen (Gutknecht, Semmer & Annen, 2004). Die zunehmende Validität dürfte zur Popularität beitragen. Dies ist auf die methodisch erzielten Fortschritte zurückzuführen. Dies gilt jedoch leider nicht für das Assessment Center (Obermann, 2013; Sarges, 2009; Schuler, 2007). Während ältere Studien grosse Validitätswerte beim AC verzeichnen konnten, werden in neueren Studien zunehmend tiefe Werte festgestellt. Dies wirft gemäss Schuler (2007) die Frage auf, wie sich erklären lässt, dass ein eignungsdiagnostisches Verfahren wie das AC heute trotz Anwendungserfahrung und Forschung schlechter dasteht als noch vor gut einem halben Jahrhundert. Schuler (2007) ist der Ansicht, dass das Assessment aufgrund der methodischen Unzulänglichkeit, die sich in der Konzeption und Anwendung von AC's niederschlägt, weit hinter seinen Möglichkeiten zurückbleibt. Er erklärt die Abnahme der prognostischen Validität unter anderem damit, dass AC's „in weiten Teilen inzwischen zu einer Spielwiese der Laiendiagnostik verkommen sein, also in mangelnder psychologischer Expertise der Designer und Assessoren von ACn“ (2009, zitiert nach Sarges, 2015, S. 2). Dies zeigt sich darin, dass oftmals auf eine fundierte Anforderungsanalyse verzichtet wird, die Assessoren oftmals nur unzureichende diagnostische Qualifi-

kationen vorweisen und keine Verfahrensevaluationen durchgeführt werden³. Genau diesem Problem schwindender Qualität versuchen etwa die Norm DIN 33430, die Qualitätsstandards für die berufsbezogene Eignungsbeurteilung festlegt, oder Vereine wie Swiss Assessment, die das Ziel der Qualitätssicherung und Weiterentwicklung verfolgen, entgegen zu wirken (Reimann, 2005; Schuler, 2007; Swiss-Assessment, 2015). Swiss Assessment gibt neun Qualitätsstandards vor, durch die man zu einem qualitativ hochstehenden Assessment Center gelangt. Auch *impuls* orientiert sich an diesen Standards. Im Zentrum dieser Arbeit steht das 9. Qualitätsstandard von Swiss Assessment (2015). Dieses gibt vor, dass Assessment Center regelmässig einer Güteprüfung unterzogen werden sollten, insbesondere bei der Neueinführung. Dadurch soll die Qualität sichergestellt werden, sodass die mit dem AC angestrebten Ziele auch nachhaltig erreicht werden.

Wie vorgängig erwähnt handelt es sich bei einem Assessment Center um ein multimodales Beurteilungsverfahren, das sich aus verschiedenen Aufgaben zusammensetzt (z.B. Intelligenztest, Persönlichkeitsfragebogen, Simulationsübung etc.) und somit auch verschiedene Messmethoden verwendet. Im Fokus dieser Arbeit steht nicht die Evaluation des gesamten Assessment Centers von *impuls*, sondern die Evaluation der neuen Assessment-Aufgabe „Turmbau“. Deshalb werden in dieser Arbeit nur die Gütekriterien dieser Aufgabe erhoben. Um diese zu eruieren, stellt sich die Frage, welcher Messmethode sich diese Aufgabe bedient. Die Antwort auf die Frage lautet: durch Verhaltensbeobachtung und -beurteilung (Schmidt-Atzert & Amelang, 2012). Um beurteilen zu können, ob die Messmethode Verhaltensbeobachtung gut oder schlecht ist, soll nachfolgend erklärt werden, was überhaupt unter Verhaltensbeobachtung zu verstehen ist. Dies wiederum setzt voraus, dass zuerst definiert werden muss, was mit Verhalten gemeint ist.

2.2 Verhalten und Dispositionen

Fröhlich (2008) versteht unter Verhalten „die Gesamtheit aller beobachtbaren, feststellbaren oder messbaren Aktivitäten des lebenden Organismus, meist aufgefasst als Reaktion auf bestimmte Reize oder Reizkonstellationen, mit denen der Organis-

³ Nach Schuler (2007) und Sarges (2009) liegt das wohl grösste Defizit bei AC's jedoch in der methodischen Einseitigkeit der eingesetzten Aufgaben. Demnach habe sich in AC-Guidelines und in der Praxis eingebürgert, nur noch mit verhaltensbezogene, arbeitsprobenartigen Aufgaben zu bestücken, wodurch die Multimodalität und die damit verbundenen Vorteile verloren gehen, komplexe Phänomene nicht mehr erfasst werden können, was in eine Abnahme der prognostischen Validität führt. Da sich die vorliegende Arbeit jedoch nicht mit dem Gesamt-Assessment befasst, sondern nur mit der einzelnen Aufgabe „Turmbau“, wird diese Thematik nicht weiter thematisiert in dieser Arbeit. Folglich wird auf die genannte Literatur verwiesen.

mus in experimentellen oder lebensweltlichen Situationen konfrontiert wird bzw. konfrontiert ist“ (S. 500). In der Wissenschaft gibt es verschiedene Auffassungen darüber, weshalb wir uns so verhalten, wie wir uns verhalten. Verhaltenstheoretiker sehen das Verhalten als Funktion einer Situation und die erfahrene Konsequenz aus dem Verhalten in einer früheren Situation. Oder einfacher ausgedrückt: die Gegebenheit einer Situation bestimmt, wie wir uns verhalten (Schmidt-Atzert & Amelang, 2012). Demgegenüber stehen die Eigenschaftstheoretiker. Diese würden das Verhalten wie folgt beschreiben: „Fragt man Laien, warum Menschen sich so verhalten, wie sie es tun, lautet eine gängige Antwort: Das liegt an ihrem Charakter“, so Schmidt-Atzert und Amelang (2012, S. 11). Demzufolge sind es die Eigenschaften, die bestimmen, wie wir uns verhalten. Stemmler, Hagmann, Amelang und Bartussek (2010, zitiert nach Schmid-Atzert & Amelang, 2012) definieren Eigenschaften („traits“) als zeitlich stabile und relativ breite Dispositionen zu bestimmten Verhaltensweisen, „die konsistent in verschiedenen Situationen auftreten“ (S. 11). Es handelt sich um Konstrukte d.h. hypothetisch, gedanklich konstruierte Gebilde, die nur erschlossen werden können aus direkt beobachtbaren Verhaltensweisen. Zerstört beispielsweise ein Kind ein Spielzeug, schlägt oder beschimpft es einen Kameraden, so schliessen wir aufgrund dieser Verhaltensweisen wahrscheinlich auf die Eigenschaft „Aggressivität“. Schmidt-Atzert und Amelang (2012) betonen, dass sich der verhaltens- und eigenschaftstheoretische Ansatz nicht gegenseitig ausschliesst, sondern ergänzt. Demnach ist Verhalten als Folge von Eigenschaften und Situationen zu betrachten, ganz im Sinne des Interaktionismus nach Ender und Magnusson (1976, zitiert nach Schmidt-Atzert & Amelang, 2012). Auch in dieser Arbeit sollen die Eigenschaften und Situation als komplementär und nicht als konkurrierend verstanden werden. Dennoch wird nachfolgend die eigenschaftstheoretische Brille aufgesetzt und zwei Eigenschaften näher gebracht, die gemäss Schmidt-Atzert und Amelang (2012) empirisch gut erforscht sind und somit eine gute Grundlage für die vorliegende Arbeit bieten. Diese werden nun in den Abschnitten 2.2.1 und 2.2.2 behandelt.

2.2.1 Die Intelligenz

Es gibt eine Vielzahl von Definitionen von Intelligenz. Eine verbale Umschreibung von Intelligenz, welcher die meisten Intelligenzforscher zustimmen würden, lautet „Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various

forms of reasoning, to overcome obstacles by taking thought“ (Neisser, Boodoo, Bouchard, Boykin, Brody & Ceci, 1996, zitiert nach Stemmler, Hagemann, Amelang & Bartussek, 2011, S. 139).

In der obigen Definition geht es im Kern darum, dass Intelligenz die Fähigkeit ist, durch Nachdenken Probleme mehr oder weniger gut lösen zu können (Stemmler et al., 2011). Damit ist jedoch noch nicht geklärt, welches die Bestandteile der Intelligenz sind. Heute gibt es in der Intelligenzforschung eine Vielzahl von Vorstellungen und Modellen darüber. Einige Modelle gehen davon aus, dass Intelligenz im Sinne eines Gesamtkonstrukts als allgemeine Intelligenz zu verstehen ist. Nach anderen Modellen verfügt man nicht über eine Intelligenz, sondern über multiple Intelligenzfaktoren (Gardner, 1983, 1999, zitiert nach Myers, 2008). Demzufolge ist Intelligenz als komplexes Konstrukt zu verstehen, das aus einer Vielzahl von Teilfähigkeiten besteht. Trotz der vielen unterschiedlichen Intelligenztheorien kann heute eine inhaltliche Konvergenz in Bezug auf die einzelnen Fähigkeiten von Intelligenz festgestellt werden. Typische Teilfähigkeiten sind: Schlussfolgerndes Denken, numerische, räumlich-figurale und verbale Fähigkeiten, Kreativität, Merkfähigkeit und wahrnehmungsbezogene Fähigkeiten (Liepmann, Beauducel, Brocke & Amthauer, 2007). Doch was hat dies nun alles mit Verhalten zu tun? Nach Rammstedt, Kemper, Klein, Beierlein und Kovaleva (2012) belegen verschiedene Studien, dass der beste Prädiktor für ein erfolgreiches Leben (z.B. Erfolg in der Schule, im Beruf usw.) die kognitiven Fähigkeiten einer Person sind, insbesondere die Intelligenz. Zu diesem Schluss kommt auch Strenze (2007, zitiert nach Schmidt-Atzert & Amelang, 2012) in einer Metaanalyse, wonach der sozioökonomische Erfolg eines Menschen sich mit Intelligenz vorhersagen lässt. Daraus lässt sich schliessen, dass sich eine Person mit hoher Intelligenz in gewissen Situationen so verhalten kann, dass sie in einer bestimmten Kultur als erfolgreich gilt (Stemmler et al., 2011). Die Intelligenz hat folglich beträchtlichen Einfluss darauf, wie wir uns verhalten.

2.2.2 Persönlichkeitsmerkmale – Die Big Five

Neben der Intelligenz und dem (Fach-)Wissen spielen gemäss Hossiep und Schulz (2013) im gesellschaftlichen und speziell im Berufsleben die Persönlichkeitseigenschaften eine wichtige Rolle. Persönlichkeitsmerkmale haben einen wesentlichen Einfluss auf die persönliche Einstellung und somit auch auf das individuelle Erleben und Verhalten. Das am weitesten verbreitete Modell zur Beschreibung der Gesamt-

persönlichkeit ist das Fünf-Faktoren-Modell der Persönlichkeit nach Costa und McCrae (Rammstedt et al., 2012, S. 7). Das Modell (auch Big-Five-Modell genannt) geht davon aus, dass es fünf breite Persönlichkeitsdomänen gibt (Hossiep & Schulz, 2013; Kemper, Ziegler, Krumm, Heene & Bühner, 2015). Diese fünf Dimensionen sind in Tabelle 6 anhand von beispielhaften Bedeutungsinhalten erklärt.

Tabelle 6: Big-Five-Persönlichkeitsfaktoren und ihnen zugeordnete Bedeutungsinhalte (Hossiep & Mühlhaus, 2005, zitiert nach Hossiep & Schulz, 2013, S. 202)

Persönlichkeitsmerkmal	Beispielhafte Bedeutungsinhalte	
Neurotizismus	- Nervosität - Ängstlichkeit - Reizbarkeit	- Verlegenheit - Unsicherheit - Besorgtheit
Extraversion	- Geselligkeit - Frohsinn - Impulsivität	- Gesprächigkeit - Aktivität - Originalität
Offenheit	- Gebildetheit - Kreativität - Kultiviertheit	- Vielseitigkeit - Aufgeschlossenheit - Originalität
Verträglichkeit	- Wärme - Hilfsbereitschaft - Toleranz	- Freundlichkeit - Bescheidenheit - Kooperation
Gewissenhaftigkeit	- Sorgfalt - Beharrlichkeit - Zuverlässigkeit	- Selbstdisziplin - Verantwortungsbewusstsein

Gemäss Rammstedt et al. (2012) konnte in diversen Studien belegt werden, dass die Big-Five ein guter Prädiktor für verschiedene Aspekte alltäglichen Lebens sind. So konnte etwa herausgefunden werden, dass weniger neurotische oder sehr gewissenhafte Personen über eine bessere Gesundheit verfügen und eine längere Lebensdauer haben. Ferner hat die individuelle Persönlichkeit einen Einfluss auf die Berufswahl. So wählen verträgliche Personen beispielsweise eher soziale Berufe, während gewissenhafte Personen eher einer konventionellen Tätigkeit nachgehen. Vedel (2014) konnte zudem im Rahmen einer Metastudie herausfinden, dass Gewissenhaftigkeit den Studienerfolg voraussagen kann. Ferner wurde in einer Metastudien festgestellt, dass insbesondere die Gewissenhaftigkeit und das Merkmal Neurotizismus die berufliche Leistungsfähigkeit vorhersagen können (Barrick, Mount & Judge, 2001, zitiert nach Hülshager & Maier, 2008).

2.3 Systematische Verhaltensbeobachtung

Unter Verhaltensbeobachtung ist das durch die Sinnesorgane ausschnittshafte Wahrnehmen der Gesamtheit aller beobachtbaren, feststellbaren oder messbaren Aktivitäten des lebenden Organismus zu verstehen. Aktivitäten sind, wie soeben dargelegt, meist eine Reaktion auf Reize oder Reizkonstellationen, mit denen der Organismus in bestimmten Situationen konfrontiert wird bzw. konfrontiert ist (Fröhlich, 2009; Stemmler & Margraf-Stiksrud, 2015). Wie sich eine Person auf einen bestimmten Reiz oder Reizkonstellation hin verhält, ist, wie aus dem Abschnitt 2.2 herausgeht, beispielsweise abhängig von den Persönlichkeitsmerkmalen und der Intelligenz. Folglich soll nun näher darauf eingegangen werden, wie Verhaltensweisen mittels der Verhaltensbeobachtung erfasst werden können.

2.3.1 Arten der Verhaltensbeobachtung

Es kann nach drei Arten von Beobachtung unterschieden werden: der alltäglichen, freien und systematischen Beobachtung (Stemmler & Margraf-Stiksrud, 2015). Die alltägliche Beobachtung dient einer Person zur Orientierung in der Welt. Bei ihr geht es darum, dass eine Person aus einem Meer von Informationen die wichtigsten „herausfischt“ und deutet, um sich in einer Situation zurechtzufinden und adäquat zu verhalten. Bei der alltäglichen Beobachtung steht deshalb die beobachtende Person selbst im Zentrum (Atteslander, 2010; Stemmler & Margraf-Stiksrud, 2015). Von der alltäglichen Beobachtung abzugrenzen ist die freie Beobachtung. Bei ihr wird bewusst und zu einem bestimmten Zweck das Verhalten anderer oder ein bestimmtes Geschehen beobachtet. Meist liegt dabei die Aufmerksamkeit auf einem gewissen Bereich wie beispielsweise dem Sozial-, Zwangs- oder Spielverhalten. Das beobachtete Verhalten wird bei der freien Beobachtung möglichst objektiv beschrieben, ohne bereits das Wahrgenommene zu interpretieren. Dennoch entscheidet der Beobachter meist selbst, welche Verhaltensweisen er genau beobachtet (Schmidt-Atzert & Amelang, 2012; Stemmler & Margraf-Stiksrud, 2015). Meist wird die freie Beobachtung in neu zu erschliessenden Forschungsgebieten angewendet. Dabei hat man keine Hypothese, aufgrund welcher man nach bestimmten Informationen sucht. Das Ziel der freien Beobachtung liegt darin, einen Gegenstand umfassend zu beschreiben, um daraus mögliche Hypothesen abzuleiten (Rentzsch & Schütz, 2009). Bei der dritten

Art der Beobachtung - der systematischen Beobachtung⁴ - liegt hingegen schon im Vorfeld der Beobachtung eine Hypothese vor, was in der Beobachtungssituation wichtig ist. Dem Beobachter wird genau vorgegeben, worauf er seine Aufmerksamkeit zu richten und was er zu protokollieren hat (Schmidt-Atzert & Amelang, 2012). Sowohl die Beobachtung als auch die Registrierung ihrer Inhalte erfolgt nach klaren Regeln. Die systematische Beobachtung stellt deshalb ein Messverfahren dar, mit dem Verhaltensmerkmale von Personen oder Personengruppen quantitativ erfasst werden (Stemmler & Margraf-Stiksrud, 2015). Dabei müssen die zu sammelnden Daten stets den Hauptgütekriterien Objektivität, Reliabilität und Validität genügen. Dadurch sollen die der systematischen Beobachtung vorangestellten Hypothesen und die zugrundeliegende Fragestellung so genau wie möglich beantwortet werden. Neben der Unterscheidung, ob eine Verhaltensbeobachtung alltäglich, frei oder systematisch ist, lässt sich eine Beobachtung weiter darüber charakterisieren, ob sie direkt oder indirekt erfolgt. Bei der indirekten Beobachtung wird die Beobachtungssituation mithilfe eines Videos aufgezeichnet. Bei der direkten Beobachtung hingegen wird in Echtzeit beobachtet. Beide Varianten haben ihre Vor- und Nachteile. Ein grosser Vorteil der direkten Beobachtung liegt darin, dass der Beobachtende seinen Blick flexibel auf jenes Geschehen richten kann, das ihm gerade interessant und wichtig erscheint. Er ist nicht auf den starren Blickwinkel der Kamera eingeeengt. Ein grosser Nachteil liegt jedoch darin, dass nicht gleichzeitig beobachtet und registriert werden kann. Registriert der Beobachtende gerade eine Verhaltensweise, indem er diese beispielsweise auf einem Stück Papier notiert, ist seine Aufmerksamkeit für eine kurze Zeit gezwungenermassen nicht beim Beobachtungsgegenstand. Er läuft dadurch Gefahr, eine wichtige Beobachtungssequenz zu verpassen. Belässt der Beobachtende seine Aufmerksamkeit permanent auf dem Geschehen und vernachlässigt das Registrieren des Verhaltens, entsteht die Gefahr des Vergessens oder des Wiedergebens falscher Fakten bei der nachträglichen Protokollierung. Der grösste Vorteil der indirekten Beobachtung liegt deshalb darin, dass dank der Videoaufzeichnung das Beobachten und Registrieren separat vorgenommen werden kann. Ausserdem können Szenen nach Belieben noch einmal angeschaut werden oder mittels einer speziellen Software die Dauer eines Verhaltens exakt bestimmt werden (Schmidt-Atzert & Amelang, 2012). Eine Beobachtung kann im Feld oder im Labor

⁴ Oftmals ist synonym auch von wissenschaftlicher Beobachtung die Rede (Granach & Frenz, 1969, zitiert nach Schnell, Hill und Esser, 2011).

stattfinden. Beobachten im Feld bedeutet, dass eine Person in ihrem gewöhnlichen Umfeld beobachtet wird. Dadurch können Kontextbedingungen mitbeobachtet werden, welche für ein Verhalten mitverantwortlich sind. Bei einer Laborbeobachtung wird eine Situation künstlich hergestellt. Der Untersucher kann die Situation passend auf seine Fragestellung gestalten und etwa bestimmte Reize beim Probanden initiieren, die ein bestimmtes Verhalten zur Folge haben sollen. Ein weiterer Vorteil liegt darin, dass die Situation standardisiert werden kann und dadurch für alle Probanden die gleichen Rahmenbedingungen gelten. Dies lässt Vergleiche unter den Probanden zu (Schmidt-Atzert & Amelang, 2012). Hingegen stellt sich bei der Laboruntersuchung die Frage, inwieweit sich die Beobachtungsergebnisse auf die Alltagssituation übertragen lassen. Ob sich eine Person im Labor gleich verhält wie im gewohnten Kontext ist fraglich (Rentzsch & Schütz, 2009; Schmidt-Atzert & Amelang, 2012).

Eine Beobachtung kann des Weiteren verdeckt und offen erfolgen. Bei der verdeckten Beobachtung ist der Beobachtende nicht sichtbar. Der Beobachtete weiss nichts davon, dass er beobachtet wird. Dadurch kann sichergestellt werden, dass sich die beobachtete Person weitgehend natürlich und aufgrund der Beobachtungssituation nicht anders verhält (reaktives Verhalten). Aus ethischen Gründen braucht es bei der verdeckten Beobachtung immer die Zustimmung der Beteiligten (Schmidt-Atzert & Amelang, 2012).

Weiter kann nach teilnehmender und nicht-teilnehmender Beobachtung unterschieden werden. Von aktiv-teilnehmender Beobachtung ist die Rede, wenn der Beobachtende an der interessierten Situation teilnimmt und mit den übrigen Teilnehmenden interagiert, z.B. indem er mitdiskutiert oder etwas erklärt. Bei der passiv-teilnehmenden Beobachtung ist der Beobachter zwar für die Teilnehmenden sichtbar, jedoch ohne mit ihnen zu interagieren. Bei der nicht-teilnehmenden Beobachtung ist der Beobachtende physisch nicht anwesend (Rentzsch & Schütz, 2009). Ein Vorteil der aktiven-teilnehmenden Beobachtung liegt darin, dass beispielsweise in einer Beobachtungssituation ein Reiz impliziert werden kann, um ein gewünschtes Verhalten zu provozieren – ganz im Sinne der Laborbeobachtung. Ein grosser Nachteil der aktiv-teilnehmenden Beobachtung ist jener, dass die Beobachtungskapazität in hohem Masse eingeschränkt ist. Dies hat Einfluss auf die Qualität der Beobachtung. Doch einige Situationen setzen es voraus, dass teilnehmend beobachtet wird. So würde es einen etwas komischen Eindruck erwecken, wenn eine Person an einer geselligen Runde teilnehmen würde, ohne ein Wort zu sagen, zu trinken oder zu essen. Dies

könnte zu Reaktivität führen (Schmidt-Azert & Amelang, 2012). Nach Stemmler (2005) kann weiter nach strukturierter und unstrukturierter Beobachtung unterteilt werden. Ob eine Beobachtung strukturiert oder unstrukturiert ist, hängt von den Regeln ab, die für die Durchführung einer Beobachtung aufgestellt werden. Die unstrukturierte Beobachtung definiert sich darüber, dass es keine oder nur sehr allgemeine Regeln dazu gibt, was es in einer Beobachtungssituation zu beobachten gilt. Somit entscheidet der Beobachtende in der Beobachtungssituation intuitiv, welche Verhaltensweisen ihm wichtig erscheinen und welche nicht (Stemmler, 2005). Attestlander (2010) bezeichnet die unstrukturierte Beobachtung auch als qualitativ orientierte Beobachtung. Beispiele von unstrukturierter Beobachtung sind die freie oder die aktivteilnehmende Beobachtung. Eher unstrukturierte Beobachtungen machen dann Sinn, wenn es darum geht, einen noch nicht erschlossenen Forschungsgegenstand explorativ zu untersuchen (Attestlander, 2010; Flick, 2007). Demgegenüber steht laut Stemmler (2005) die strukturierte Beobachtung oder quantitativ orientierte Beobachtung, wie es Attestlander (2010) nennt. Während die unstrukturierte Beobachtung eher explorativen Zweck hat, ist die strukturierte Beobachtung dann sinnvoll, wenn es darum geht, aus der Theorie abgeleitete Hypothesen zu überprüfen (Attestlander, 2010). Die strukturierte Beobachtung hat den Anspruch, das zu beobachtende Verhalten einer Person möglichst objektiv zu erheben d.h. zu beschreiben und zu beurteilen. Die Beobachtung soll wertefrei und unabhängig von Interpretationen des Beobachtenden sein, damit die Wirklichkeit so genau wie möglich abgebildet wird. Deshalb werden Verhaltensbeobachtung und -beurteilung strikte getrennt. Die strukturierte Beobachtung erfolgt wie schon erwähnt theoriegeleitet und zeichnet sich durch eine kontrollierte Wahrnehmung, Aufzeichnung und Auswertung aus (Attestlander, 2010). Dazu wird der beobachtenden Person mit einem Beobachtungsplan genau vorgegeben, zu welchem Zeitpunkt welches Verhalten in welcher Form registriert werden soll. Meist werden den beobachtenden Personen deshalb sogenannte Protokollbögen resp. Beobachtungsraster ausgeteilt, in denen festgelegte Regeln in sogenannten Kodiersystemen übertragen sind (Stemmler, 2005).

Das Ausführen der verschiedenen Arten von Verhaltensbeobachtung - insbesondere in Bezug auf den Strukturierungsgrad - ermöglicht es, die angewandte Verhaltensbeobachtung im Rahmen der Turmbauaufgabe einzuordnen. Im Zentrum dieser Arbeit steht die AC-Aufgabe „Turmbau“, die bestimmte Verhaltensweisen bei Teilnehmenden auslösen soll, wodurch auf Fähigkeiten geschlossen werden kann. Diesem

Sachverhalt unterliegt somit eine Hypothese: Um die Aufgabe lösen zu können, braucht es Planungsfähigkeit. Dies äussert sich beispielsweise in der Verhaltensweise, dass der Teilnehmende eine Auslegeordnung macht, bevor er mit dem Bau des Turms beginnt usw. Folglich handelt es sich in dieser Arbeit um einen Sachverhalt, den es eher mit einem strukturierten Beobachtungsverfahren zu prüfen gilt. Im nachfolgenden Abschnitt 2.3.2 wird deshalb näher ausgeführt, wie die Beobachtung strukturiert und dadurch systematisiert werden kann.

2.3.2 Verhaltensbeobachtung systematisieren

Nach Schmid-Atzert und Amelang (2012) liegt das Ziel der systematischen Verhaltensbeobachtung nicht darin, das vollständige Verhalten zu beschreiben. Es soll nur jener Teilaspekt des Verhaltens erfasst werden, der im Zentrum des Forschungsinteresses steht, z.B. Misserfolgstoleranz (Schmid-Atzert & Amelang, 2012). Dazu werden sogenannte Beobachtungseinheiten bestimmt, die Atteslander (2010) als „denjenigen Teilbereich sozialen Geschehens, der konkreter Gegenstand der Beobachtung sein soll“ bezeichnet (S. 82). Beobachtungseinheiten müssen inhaltlich definiert werden. Damit ist gemeint, dass im Vorfeld der effektiven Beobachtung festgelegt wird, welche Verhaltensweisen beispielsweise einen Hinweis auf die Planungsfähigkeit geben, z.B. macht eine Auslegeordnung, macht Notizen usw. Die Definitionen der Verhaltensweisen sollen dabei möglichst genau und verständlich definiert sein, sodass diese von den Beobachtenden ohne Missverständnisse erkannt werden. Das Merkmal Aggressivität könnte beispielsweise mit der Verhaltensweise „schlägt mit der Hand auf den Tisch“ definiert werden. Jedoch könnte dieses Verhalten auch ein Zeichen für Ungeduld und nicht Aggressivität sein – es gilt hier folglich abzuwägen. Laut Margraf-Stiksrud (2005) kann anhand von vier Kriterien bestimmt werden, ob eine Verhaltensbeschreibung als Beobachtungseinheit geeignet ist:

1. „Kann die Einheit zuverlässig von allen Beobachtenden identifiziert werden (versteht jeder unter „reagiert wütend“ dasselbe)?“
2. Ist mit den gewählten Beschreibungen das Merkmal in allen relevanten Facetten beschrieben?
3. Kann die Liste der Verhaltensweisen insgesamt von einem einzelnen Beobachter während des Beobachtungszeitraums überblickt werden?
4. Kommt das beschriebene Verhalten in der Beobachtungssituation überhaupt vor?“ (S. 57)

Sind Beobachtungseinheiten erst einmal definiert, können sie quantifiziert werden. Das heisst, es kann beurteilt werden, wie häufig, wie stark, wie andauernd und in wie vielen Verhaltensvarianten eine bestimmte Verhaltensweise auftritt (Schmidt-Atzert, 2005). Es gibt drei wesentliche Methoden und Techniken der Quantifizierung: Die Ereignisstichprobe (event sampling), die Zeitstichprobe (time sampling) und das Ratingverfahren (Fassnacht, 1995). Von Ereignisstichproben ist die Rede, wenn gewählte Ereignisse resp. Verhaltensweisen immer dann registriert werden, wenn sie auftreten. Folglich werden die festgelegten Ereignisse während der gesamten Beobachtungssituation erfasst und zwar, wie oft oder wie lange sie auftreten (meist ist die Häufigkeit und nicht die Dauer von Interesse). Dadurch können quantitative Aussagen und überdies Interpretationen gemacht werden, z.B. Person A hat sich zehn Mal aggressiv verhalten in der Situation X, Person B nur zweimal. Person A hat folglich eine aggressivere Veranlagung. Beim zweiten Verfahren, dem Zeitstichprobeverfahren, liegt das Prinzip darin, die Beobachtungssituation in exakt gleich lange Zeitabschnitte zu unterteilen (beispielsweise 30 Sekunden pro Zeitintervall). Dabei wird registriert, ob eine bestimmte Verhaltensweise in einem Zeitintervall vorkommt oder nicht (Schmidt-Atzert, 2005). Die Ereignis- und Zeitstichprobe werden meist zusammen mit sogenannten Kategoriensystemen angewendet. Das wesentliche Prinzip des Kategoriensystems liegt darin, das registrierte Verhalten, beispielsweise das erste beobachtete Ereignis oder ein festgelegter Zeitintervall, in klar definierte und voneinander abgrenzbare Kategorien einzuordnen. Dies geschieht in der Regel mit einer Kennzeichnung im Sinne eines Bleistiftstriches auf einem vorgefertigten Kategorienschema. Daraus lässt sich zum Schluss der Beobachtungssituation die Auftretenshäufigkeit erkennen. Das dritte Verfahren der Quantifizierung, dem sich *impuls* bedient und das überhaupt mehrheitlich bei AC verwendet wird, nennt sich Ratingverfahren. Das Ratingverfahren ist weniger aufwendig. Im Gegensatz zu den genannten Verfahren werden beim Ratingverfahren nicht einzelne Verhaltensmerkmale registriert und Auftretenszeiten sowie -häufigkeiten ermittelt. Die beobachtende Person bildet sich beim Ratingverfahren eine Art Pauschalurteil zum Verhalten einer Person (z.B. wie strukturiert ist die Person, welche die Turmbauaufgabe bearbeitet hat, während der gesamten Beobachtungssituation vorgegangen?). Beim Ratingverfahren muss der Beobachtende in einem ersten Schritt „nur“ beobachten resp. Informationen aufnehmen. Nach einer längeren Beobachtungsphase nimmt er eine Verhaltensbeurteilung vor resp. fällt ein eher globales Urteil über das Verhalten des Pro-

banden, das er ebenfalls in einem Kategorienschema einträgt. Dies geschieht in der Regel mit einer sogenannten Intervallskala (z.B. von 1= *sehr schwach ausgeprägt* bis 5=*sehr stark ausgeprägt*) entweder am Ende der gesamten Beobachtungssituation oder, falls das Verhalten sinnvoll segmentiert werden kann, zu mehreren Zeitpunkten (z.B. erstes Segment von Minute 0-10: Planungsfähigkeit, zweites Segment von Minute 10-20: kritische Kontrolle). Im Gegensatz zur Ereignis- und Zeitstichprobe ist das Ratingverfahren in Bezug auf die Datenerhebung - obschon Werte für ein bestimmtes Verhalten vergeben werden - eher qualitativer Natur. Die Subjektivität der Beobachtenden wird bei der Beobachtung und Beurteilung im Sinne des der qualitativen Sozialforschung zugrundeliegenden interpretativen Paradigmas bewusst miteinbezogen in die Beobachtung und Beurteilung (Flick, 2007). Die beobachtende Person nimmt bei Ratingverfahren viele einzelne Verhaltensmerkmale wahr, die nach ihrer Auffassung bedeutsam und Hinweisreiz für das zu beurteilende Merkmal sind. Dabei werden die einzelnen Informationen intuitiv gewichtet (Schmidt-Atzert, 2005). Bei allen Vorteilen, die eher qualitativ orientierte Datenerhebungsmethoden mit sich bringen, bergen sie auch einige Schwierigkeiten in sich⁵. Denn die in die Beobachtung einflussende Subjektivität des Beobachtenden kann zu Ergebnisverzerrungen führen. Wie wir beobachten und eine Situation beurteilen, ist abhängig von diversen Faktoren wie etwa der individuellen Erfahrung und Kognition (Schmidt-Atzert, 2005; Aronson, Wilson & Akert, 2008). Nachfolgend sollen daher einige mögliche Beobachtungsfehler und Beobachtungsverzerrungen ausgeführt werden, bevor im Anschluss daran erläutert wird, wie solche Fehler vermieden werden können.

2.3.3 Beobachtungs- und Beurteilungsfehler

Laut Greve und Wentura (1997) kann nach drei Arten von Fehlern unterschieden werden: Fehler zu Lasten des Beobachters, Fehler zu Lasten der Beobachtung und Fehler zu Lasten äusserer Bedingungen. Dieser Abschnitt befasst sich mit der ersten Fehlerart. Die Fehler zu Lasten der Beobachtung werden im Abschnitt 2.3.4 und jene zu Lasten äusserer Bedingungen im Abschnitt 2.4.1 angesprochen. Nachfolgend werden nun einige wichtige Fehler erster Art erläutert.

- **Beobachterdrift**

Die Beobachtenden entdecken im Verlauf der Beobachtung zunehmend mehr

⁵ Weitere Informationen zum Verhältnis zwischen qualitativer und quantitativer Sozialforschung und den jeweilig zugeschriebenen Vor- und Nachteilen findet sich in Flick (2007, S. 39-55.)

oder weniger relevante Ereignisse, obschon sich das Verhalten der zu beobachtenden Person nicht geändert hat. Nach Schmid-Atzert (2005) ist die Abnahme der beobachteten Ereignisse meist auf eine abnehmende Aufmerksamkeit des Beobachtenden zurückzuführen. Die Zunahme hängt dagegen oftmals mit mangelnder Übung zusammen: Die Beobachtenden sind zu Beginn der Beobachtung überfordert und werden im Verlaufe der Beobachtungssituation immer „ser“ und entdecken immer mehr Verhaltensweisen.

- **Halo-Effekt / logischer Denkfehler**

Damit ist gemeint, „dass die zentrale Eigenschaft eines Menschen seine anderen Eigenschaften überstrahlt“ (Schmid-Atzert, 2005, S. 79). Tritt eine Person zum Beispiel besonders charmant auf, werden ihr womöglich auch andere Eigenschaften wie etwa grosse Fach- und Sozialkompetenz, Kreativität etc. zugeschrieben. Der Beobachtende lässt sich also quasi blenden vom Charme der Person. Logischer Denkfehler könnte aber auch etwa sein, dass eine Person, die Brillenträgerin ist, automatisch als intelligent eingestuft wird.

- **Fundamentaler Attributionsfehler**

Ein fundamentaler Attributionsfehler liegt vor, wenn versucht wird, das Verhalten von Menschen mit ihrer Persönlichkeit zu erklären. Dabei werden die situationsbedingten Faktoren unterschätzt (Aronson et al., 2008). Diesem Fehler gilt es unbedingt im Rahmen von Assessments vorzubeugen, bei dem ja bekanntlich aufgrund von Verhaltensweisen in bestimmten Situationen auf Eigenschaften geschlossen wird. Fundamentale Attributionsfehler können im Rahmen von Assessments Folgen von Fehlern der dritten Art sein (vgl. Abschnitt 2.4.1), also Fehlern aufgrund unzulänglicher äusserer Rahmenbedingungen (z.B. hektische Atmosphäre im AC, unklare Übung etc.) (Obermann, 2013).

- **Primacy- und Recency-Effekt**

Primacy-Effekt: An die zu Beginn einer Beobachtungssituation wahrgenommenen Informationen kann man sich besser erinnern als an jene in der Mitte. Folglich wird man eine Person zum Schluss einer Beobachtungssituation vor allem nach dem Verhalten, das die Person zu Beginn der Situation gezeigt hat, beurteilen (Schmid-Atzert, 2005).

Recency-Effekt: Die abgespeicherten Informationen aus der zuletzt beobachteten Verhaltensweisen einer Beobachtungssituation sind im Gedächtnis verfügbarer

als andere. Die Beurteilung erfolgt oftmals auf Basis dieser Informationen. Um dem Primacy- und Recency-Fehler vorzubeugen, bietet es sich an, Ratings zu mehreren Zeitpunkten durchzuführen, z.B. indem das Ratingverfahren mit dem Verfahren der Zeitstichprobe kombiniert wird (Schmid-Atzert, 2005)

- **Zentrale Tendenz**

Damit ist das Phänomen gemeint, dass Personen bei mehrstufigen Skalen dazu neigen, die Skalenmitte anzuwählen. Oftmals ist dies eine Art Flucht des Beurteilenden, weil er etwa nicht weiss, welches Urteil angemessen ist. Die Tendenz zur Mitte kann beispielsweise die Folge einer Überforderung auf Seiten des Beurteilenden aufgrund von zu schwer zu beurteilenden Merkmalen sein. Um keine verfälschten Ergebnisse zu erhalten, empfiehlt es sich, z.B. eine weitere Antwortkategorie mit „nicht beurteilbar“ aufzuführen (Schmid-Atzert, 2005). Ferner ist es wichtig, einen guten Beobachtungsbogen zu entwickeln, der den Beobachtenden darin unterstützt, die Komplexität der Beobachtungssituation zu reduzieren. Mit letzterem beschäftigt sich der Abschnitt 2.3.4.

Laut Schmidt-Atzert (2005) sind sich die Beobachtenden den Fehlern oftmals gar nicht bewusst. Deshalb ist es wichtig, dass Beobachtende ein Beobachtertraining besuchen, bei dem sie die Beobachtungsmethodik kennenlernen, sensibilisiert werden hinsichtlich Beobachtungsfehlern und mit den Beobachtungseinheiten und folglich mit dem Beobachtungsleitfaden vertraut werden. Weitere Informationen zum Thema Beobachtertraining findest sich bei Swiss Assessment (2015). Im nächsten Abschnitt soll nun noch das Thema „Beobachtungsraster“ diskutiert werden.

2.3.4 Das Beobachtungsraster

Sowohl bei der Ereignis- und der Zeitstichprobe als auch beim Ratingverfahren verwendet man sogenannte Beobachtungsraster⁶. Beobachtungsraster sind eine Art Beobachtungsschema, das vorgibt, was und wie zu beobachten ist (Attestlander, 2010; Grümer, 1974). Sie dienen der Standardisierung der Beobachtung. Mit ihnen wird die Konzentration der beobachtenden Person auf die interessierenden Verhaltensmerkmale gelenkt. Beobachtungsraster beinhalten die relevanten Kategorien, die beobachtet werden sollen (z.B. Misserfolgstoleranz). Dabei werden diese Kategorien beispielhaft mit damit zusammenhängenden Beobachtungseinheiten (z.B. verhält

⁶ Grümer (1974) bezeichnet Beobachtungsraster als Beobachtungsleitfaden, womit in dieser Arbeit dasselbe gemeint ist.

sich aggressiv) konkretisiert. Beobachtungsraster haben aber auch die Funktion eines Klassifikationssystems, das soziale Vorgänge aus der Beobachtungssituation zerlegt und einer passenden Kategorie zuteilt (z.B. Ereignis- und Zeitstichprobenverfahren) und beurteilt werden (Ratingverfahren) (Grüner, 1974).

Wie schon im Abschnitt 2.3.2 erwähnt, hat das Ratingverfahren durchaus eine qualitative Komponente. Dies kann zu Beobachtungsverzerrungen oder Beobachtungsfehlern führen (vgl. Abschnitt 2.3.3). Greve und Wentura (1997) nennen dies Fehler zu Lasten der Beobachtung aufgrund eines falschen Beobachtungssystems. Um diese zu vermeiden, sollen die beim Ratingverfahren verwendeten Beobachtungsraster gewisse Kriterien erfüllen. Diese werden nun in Anlehnung an Grüner (1974) kurz beschrieben.

- **Konkretion** (Bestimmtheit)

Das beobachtbare Verhalten muss den Kategorien zugeordnet werden können. Das zu beobachtende Verhalten muss genau definiert werden (Grüner, 1974). Eine Hilfestellung dafür bieten die vier Kriterien zur Definition von Beobachtungseinheiten nach Margraf-Stiksrud (2005) (vgl. Abschnitt 2.3.2).

- **Ausschliesslichkeit**

Die verschiedenen Kategorien müssen sich voneinander abgrenzen. Eine bestimmte Verhaltensweise resp. Beobachtungseinheit darf nur in einer Kategorie vorkommen.

- **Vollständigkeit**

Alle Verhaltensweisen, die in der Situation auftreten können, müssen bei der entsprechenden Kategorie auf dem Beobachtungsraster erfasst sein. Damit sollen durch den Beobachtenden registrierte Verhaltensweisen nicht falschen Kategorien zugeordnet oder falsch beurteilt werden.

- **Begrenzung**

Beobachtende sollen anhand von drei bis maximal vier Kategorien beobachten. Mehr als vier Kategorien überfordern die Informationsverarbeitungsleistung des Beobachtenden.

2.4 Gütekriterien

In den vorangehenden Abschnitten wurde aufgezeigt, was unter einem Assessment Center zu verstehen ist. Anschliessend wurde auf die Messmethodik eingegangen, der sich das Assessment Center bedient – die Verhaltensbeobachtung. In diesem Zusammenhang wurde aufgezeigt, wie man zu einer systematischen Verhaltensbeobachtung gelangt und was es dabei zu berücksichtigen und zu vermeiden gilt. Diese differenzierte Darlegung der systematischen Verhaltensbeobachtung ist insbesondere deshalb wichtig, da der Grad der Systematisierung erheblichen Einfluss auf die Qualität einer Assessment-Aufgabe und somit Einfluss auf deren Güte hat. Mit den Gütekriterien beschäftigt sich nun dieser Abschnitt.

Nach Lienert (1996, zitiert nach Rauchfleisch, 2005) kann nach drei Hauptgütekriterien unterschieden werden: Objektivität, Reliabilität und Validität. Ferner gibt es die Nebengütekriterien Normierung, Zulänglichkeit, Vergleichbarkeit, Ökonomie und Nützlichkeit. Nachfolgend werden diese Gütekriterien näher ausgeführt. Der Fokus liegt dabei auf den Hauptgütekriterien. Diese werden hier stets in Anlehnung an das Messverfahren der systematischen Verhaltensbeobachtung erläutert. Die Nebengütekriterien werden in dieser Arbeit nicht behandelt, da dies den Rahmen der Arbeit sprengen würde.

2.4.1 Objektivität

Als ein wichtiges Kriterium für die Güte eines Tests gilt im Rahmen der klassischen Testtheorie die Objektivität. Lienert (1969, zitiert nach Rauchfleisch, 2005) versteht unter Objektivität den Grad, „in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind“ (S. 65). Kommen verschiedene Untersucher bei denselben Probanden zum identischen Testergebnis, wird von vollkommener Objektivität gesprochen. Oftmals ist deshalb auch von interpersoneller Übereinstimmung die Rede (Lienert & Raatz, 1998). Diese Arbeit lehnt sich an das Verständnis von Objektivität nach Lienert (1996, zitiert nach Rauchfleisch 2005) an. Demnach muss weiter nach drei Arten von Objektivität differenziert werden: Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität.

Durchführungsobjektivität

Damit ist der Grad der Unabhängigkeit des Testergebnisses vom Verhalten des Untersuchenden gemeint. Eine geringe Durchführungsobjektivität besteht dann, wenn der Untersuchende während der Testdurchführung systematisch oder zufällig sein

Verhalten variiert, was wiederum zu einer Verhaltensvariation aufseiten des Probanden führt und folglich das Ergebnis beeinflusst. Um die Durchführungsobjektivität zu erhöhen, muss der Untersuchende möglichst genaue (schriftliche) Instruktionen erhalten, wie er sich in der Untersuchungssituation zu verhalten hat. Die Untersuchungssituation muss so weit wie möglich standardisiert werden (Lienert & Raatz, 1998). Die Interaktion zwischen dem Untersuchenden und Probanden soll während der Testsituation auf ein Minimum reduziert werden (Lienert & Raatz, 1998). Ist die Durchführungsobjektivität während der Beobachtungssituation unzureichend, kann es zu sogenannten Beobachtungsfehlern zulasten äusserer Bedingungen nach Greve und Wentura (1997) kommen (vgl. Abschnitt 2.3.3). Ein Beispiel hierfür wäre reaktives Verhalten.

Auswertungsobjektivität

Mit Auswertungsobjektivität ist im Rahmen der systematischen Verhaltensbeobachtung die Beobachterübereinstimmung (Inter-Rater-Reliabilität) gemeint. Die Auswertungsobjektivität ist die wohl wichtigste Art der Objektivität bei Verhaltensbeobachtungen. Sie gibt an, „wie hoch die Übereinstimmung verschiedener Beobachter bezüglich der Registrierung eines oder mehrerer Verhaltensmerkmale ist“ (Bühner, 2005, S. 82). Dabei sollen die Registrierungen beider Beobachtender unabhängig voneinander d.h. ohne gegenseitige Absprache geschehen. Wie aus den Abschnitten 2.3.2 bis 2.3.4 herausgeht, ist das Ziel von Beobachtertrainings, Beobachtungsfehler zu senken. Fehler zu identifizieren und ihr Ausmass zu bestimmen, „ist die Aufgabe von Untersuchungen zur Beobachterübereinstimmung“ (Greve & Wentura, 1997, S. 96). Weist eine Verhaltensbeobachtung eine schlechte Auswertungsobjektivität auf, ist dies ein Hinweis darauf, dass die Beobachtung zu wenig strukturiert ist, die Beobachtenden zu wenig trainiert sind, die Beobachtungseinheiten falsch oder unzureichend definiert sind oder der Beobachtungsleitfaden inadäquat ist (Bühner, 2005). Um den Kennwert der Auswertungsobjektivität zu errechnen, gibt es verschiedene Möglichkeiten: Eine Möglichkeit besteht darin, dass man den Korrelationskoeffizienten berechnet, welcher den Grad des Zusammenhangs der Ergebnisse von mindestens zwei Beobachtenden berechnet. Weiter gibt es die Möglichkeit, den Übereinstimmungsgrad der Ergebnisse beider Beobachter prozentual zu berechnen, was als sogenannter Kappa-Wert bezeichnet wird. Welches Mass eingesetzt wird, ist abhängig vom Skalenniveau, welchem sich das Beobachtungsverfahren bedient (nominal

oder intervallskaliert) (Bortz & Döring, 1995; Bühner, 2005; Bortz, 1984, zitiert nach Rauchfleisch, 2005).

Interpretationsobjektivität

Unter Interpretationsobjektivität versteht sich nach Rauchfleisch (2005) der „Grad der Unabhängigkeit der Interpretation des Testergebnisses von der Person des interpretierenden Untersuchers“ (S. 66). Hier ist also der Prozess gemeint, der nach der Verhaltensbeobachtung resp. Registrierung von Verhaltensweisen stattfindet, z.B. die Interpretation von registrierten Häufigkeiten bei der Ereignis- oder Zeitstichprobe. Diesbezüglich ist es wichtig, in einem Manual festzuhalten, wie ein bestimmtes erhobenes Testverhalten zu interpretieren ist (Rauchfleisch, 2005). Wie den vorangehenden Darlegungen entnommen werden kann, erfolgt die Beobachtung und Beurteilung bei einem Ratingverfahren im Grunde genommen im selben Zuge – obschon die Beobachtenden dazu angehalten werden, zuerst zu beobachten und erst am Schluss die Beobachtungssituation zu beurteilen. Daraus kann gefolgert werden, dass bei Ratingverfahren mittels der Erhebung der Auswertungsobjektivität sogleich auch die Interpretationsobjektivität erhoben wird. Denn bei der Erhebung der Auswertungsobjektivität von Ratingverfahren werden die auf Ratingskalen eingetragenen Beurteilungen zweier Beobachter verglichen, bei denen schon eine Interpretation im Sinne der Beurteilung des beobachteten Verhaltens stattgefunden hat. Wie die Skalenwerte letztlich interpretiert werden, lässt sich selbstverständlich dem Thema der Interpretationsobjektivität zuordnen.

2.4.2 Reliabilität

Mit Reliabilität ist die Zuverlässigkeit oder der Grad der Genauigkeit eines Verfahrens gemeint, mit dem es ein bestimmtes Verhaltens- oder Persönlichkeitsmerkmal misst, unabhängig davon, ob es dieses Merkmal zu messen beabsichtigt (Lienert & Raatz, 1998). Ein Verfahren darf dann als reliabel bezeichnet werden, wenn es das Merkmal, das es zu messen gilt, exakt misst. Bei der Reliabilität geht es um die formale Exaktheit der Merkmalerfassung, und nicht um die inhaltlichen Aspekte (Rauchfleisch, 2005). Die Reliabilität wird durch den sogenannten Reliabilitätskoeffizienten bestimmt. Dieser drückt aus, inwieweit die Testergebnisse reproduzierbar sind. Wie schon bei der Objektivität muss auch bei der Reliabilität nach verschiedenen Formen unterschieden werden (Lienert & Raatz, 1998). Es lassen sich drei Arten der Reliabilität unterscheiden: Retestreliabilität, Paralleltestreliabilität und interne Konsistenz (A-

sendorpf, 2009). Die Retest- und Paralleltestreliabilität werden nachfolgend näher ausgeführt. Die interne Konsistenz wird in dieser Arbeit nicht weiter thematisiert. Diese wird im Rahmen von Verhaltensbeobachtungen normalerweise nicht erhoben. Zum Thema interne Konsistenz und Verhaltensbeobachtung sei bei Interesse auf Westemeyer und Nell (2005, S. 207) verwiesen.

Retestreliabilität

Bei der Retestreliabilität wird ein Proband in der Regel zweimal mit dem gleichen Verfahren untersucht, sodass der Grad der Übereinstimmung der ersten und zweiten Messung anhand einer Korrelation berechnet werden kann. Unterscheidet sich das Ergebnis der ersten Erhebung stark von der zweiten, wäre eine geringe Reliabilität die Folge. Somit wären die Ergebnisse der Messungen unzuverlässig und unbrauchbar (Rauchfleisch, 2005). Bei der Methode der Messwiederholung besteht die Gefahr, dass die erzielten Reliabilitätswerte überschätzt werden. Denn der Proband erinnert sich beim zweiten Durchlauf womöglich an die Aufgabe, wodurch er von einem „Übungseffekt“ profitiert. Die Wahrscheinlichkeit eines Erinnerungseffekts kann jedoch durch die Vergrößerung des Zeitabstandes zwischen den beiden Testzeitpunkten verringert werden. Unbrauchbar ist die Retest-Methode bei der Erhebung von zeitlich instabilen Merkmalen (z.B. Erhebung der Angespanntheit). Es wäre schwierig zu bestimmen, ob ein allfällig geringer Reliabilitätskoeffizient für einen unzuverlässigen Test oder ein unstabiles Kriterium stünde. Die Methode der Retestreliabilität ist zeitlich und untersuchungstechnisch sehr aufwendig, da derselbe Proband nach einem zeitlich festgelegten Intervall erneut zur Teilnahme am Test kontaktiert und motiviert werden muss (Bortz & Döring, 1995).

Paralleltest-Reliabilität

Wie die Retest-Reliabilitätsprüfung ist auch die Paralleltest-Reliabilität mit einem grossen untersuchungstechnischen Aufwand verbunden (Bortz & Döring, 1995). Dieses Verfahren, bei welchem zwei Parallelformen der ein und derselben Stichprobe dargeboten werden, gilt als bestes Verfahren der Testreliabilität (Lienert & Raatz, 1998). Beim Paralleltest wird davon ausgegangen, dass man zu mehreren Messungen ein und desselben Merkmals kommen kann, indem man verschiedene Tests einsetzt, die dieses gleiche Merkmal erfassen (Rauchfleisch, 2005). Die aus den zwei Tests herausgehenden Ergebnisse werden miteinander korreliert und ergeben den Reliabilitätskoeffizient (Fröhlich, 2008). Um die Paralleltestreliabilität durchführen zu

können, muss ein äquivalenter Paralleltest vorliegen. Äquivalente Parellelformen liegen gemäss Rauchfleisch (2005) jedoch selten vor. Auch stellt sich die Frage, ab wann die „Parallelität“ gewährleistet ist. Die Ansichten diesbezüglich divergieren in der Wissenschaft (Rauchfleisch, 2005).

Inter-Rater-Reliabilität

Gemäss Rentsch und Schütz (2009) hängt die Reliabilität bei der Verhaltensbeobachtung besonders stark mit der Objektivität zusammen. So wird sie denn auch meist mit der sogenannten Inter-Rater-Reliabilität resp. Beobachterübereinstimmung gemessen (Rentsch & Schütz, 2009; Greve & Wentura, 1997). Wie aus dem Abschnitt 2.4.1 herausgeht, ist die Beobachterübereinstimmung jedoch auch das Mass für die Auswertungsobjektivität. Dies verlangt nach einer Erklärung. Simon und Kreuzpointer (2008, zitiert nach Rentsch & Schütz, 2009) erachten die Interrater-Übereinstimmung als Methode zur Objektivitätsbestimmung und nicht Reliabilitätsbestimmung. Ihrer Meinung nach wird die Inter-Rater-Übereinstimmung heute als Indiz für die Reliabilität herangezogen, weil die Rater ursprünglich als Paralleltestformen aufgefasst wurden. Diese Auffassung ist jedoch insofern problematisch, als die Rater unterschiedliche Personen sind und somit auch in gewisser Weise unterschiedliche Instrumente darstellen, deren Äquivalenz fraglich ist (Rentsch & Schütz, 2009).

2.4.3 Validität

Unter Validität versteht sich der Grad der Genauigkeit, mit welchem ein Verfahren dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise tatsächlich misst, unter Berücksichtigung dessen, was das Verfahren messen soll oder zu messen vorgibt. Es handelt sich um das wichtigste, aber zugleich empirisch am schwierigsten überprüfbare Gütekriterium (Rauchfleisch, 2005). Auch bei der Validität kann nach verschiedenen grundsätzlichen Aspekten unterschieden werden. Demnach ist von inhaltlicher Validität, Konstruktvalidität sowie kriteriumsbezogener Validität die Rede (Bortz & Döring, 1998).

Inhaltsvalidität

Inhaltsvalidität wird oftmals auch als Augenscheinvalidität bezeichnet. Auf die Verhaltensbeobachtung heruntergebrochen ist die Inhaltsvalidität dann gegeben, wenn das zu messende Konstrukt, z.B. Misserfolgstoleranz, in seinen wichtigsten Aspekten durch die Beobachtungseinheiten erfasst wird. Eine geringe Inhaltsvalidität bestünde

etwa dann, wenn man der Kategorie Misserfolgstoleranz Beobachtungseinheiten unterordnen würde, die mit Misserfolgstoleranz wenig zu tun hätten (Bortz & Döring, 1998). Das Konzept der Inhaltsvalidierung wird vor allem dann angewendet, wenn ein „neues“ Messverfahren untersucht werden soll, das Kriterien zu messen versucht, über die empirisch noch wenig bekannt ist (Asendorpf, 2009). Die Höhe der Inhaltsvalidität eines Tests beruht einzig auf der subjektiven Einschätzung und kann nicht numerisch geprüft werden. Diese Form der Validitätsüberprüfung stellt die schwächste Art dar und streng genommen handelt es sich dabei mehr um eine Zielvorgabe, die bei der Testkonstruktion bedacht werden soll und nicht um ein Testgütekriterium (Bortz & Döring, 1998).

Kriteriumsvalidität

Nach Bortz und Döring (1995) liegt Kriteriumsvalidität dann vor, „wenn das Ergebnis eines Tests zur Messung eines latenten Merkmals bzw. Konstrukts (z.B. Berufseignung) mit Messungen eines korrespondierenden manifesten Merkmals bzw. Kriteriums übereinstimmt (z.B. beruflicher Erfolg)“ (S. 185). Die Kriteriumsvalidität ergibt sich folglich aus einer Korrelation zwischen den Testwerten und den Kriteriumswerten einer Stichprobe. Oftmals handelt es sich bei den Kriteriumswerten um ein beobachtbares Verhalten, das erst zu einem späteren Zeitpunkt erhoben werden kann. Ob beispielsweise die in der Turmbaufaufgabe erhobene Planungsfähigkeit (Testwert) wirklich das Kriterium Berufserfolg erfasst, zeigt sich meist erst einige Zeit später. Die Kriteriumsvalidität wird auch als prognostische Validität bezeichnet. Häufig kann die Kriteriumsvalidierung nicht angewendet werden, da kein adäquates Aussenkriterium vorhanden ist (Bortz & Döring, 1995). Ausserdem ist es jeweils schwierig, den Korrelationskoeffizienten zu interpretieren. So kann aus einem kleinen Validitätswert nicht gefolgert werden, ob es sich um einen schlechten Fragebogen oder etwa um ein schlechtes Aussenkriterium handelt.

Konstruktvalidität

Die Eigenschaft der Methode der Konstruktvalidierung besteht darin, dass ein psychologisches Konstrukt - d.h. komplexe Fähigkeiten oder Eigenschaften - vor dem Hintergrund ihrer theoretischen Herleitung „auf ihre Zusammenhänge mit dem zu prüfenden Test untersucht werden“ (Rauchfleisch, 2005, S.73). Es wird folglich aufgrund von theoretischen Überlegungen und daran anknüpfenden empirischen Untersuchungen bestimmt, ob ein Konstrukt durch einen bestimmten Test gemessen wer-

den kann (Lienert & Raatz, 1998). Laut Bortz und Döring (1995) ist ein Test dann konstruktvalid, wenn Hypothesen aus dem zu messenden Zielkonstrukt ableitbar sind und diese durch die Testwerte bestätigt werden können. Das Prinzip der Konstruktvalidierung liegt darin, dass man ein Netz von Hypothesen über das Konstrukt formuliert, welche in Relation mit anderen manifesten und latenten Variablen stehen. Man gehe davon aus, ein Fragebogen zur subjektiven Einsamkeit müsste validiert werden. Aus der Einsamkeitstheorie wäre bekannt, dass ein geringer Selbstwert und soziale Ängstlichkeit mit Einsamkeit zusammenhängen und bei geschiedenen Ehepaaren ausgeprägter wären. Eine solche inhaltliche Hypothese zu prüfen, wäre das Ziel der Konstruktvalidierung (Bortz & Döring, 1995).

2.4.4 Wechselbeziehung zwischen den Gütekriterien

Nach Lienert und Raatz (1998) besteht zwischen den Gütekriterien eine Wechselbeziehung. Demzufolge hat die Objektivität direkten Einfluss auf die Reliabilität und indirekten Einfluss auf die Validität. Die Reliabilität kann demnach nicht höher sein als seine Objektivität. Wiederum ist der Grad der Genauigkeit direkt durch die Reliabilität beeinflusst. Somit lässt sich schlussfolgern, dass ein Test nicht valider ist, als er reliabel ist. Nach Margraf-Stiksrud und Stemmler (2015) stellt die Validität das wichtigste Gütekriterium dar. Um dieses angemessen beurteilen zu können, müssen jedoch auch die Objektivität und Reliabilität berücksichtigt werden.

2.5 Zwischenfazit

In den vorangegangenen theoretischen Darlegungen wurde aufgezeigt, dass die Popularität rund um das Assessment-Center zunehmend ist. Hingegen ist deren Qualität tendenziell abnehmend (Schuler, 2007). Der Grund dafür ist, dass die AC's zur Spielwiese der „Laiendiagnostiker“ verkommen sind, denen das Wissen fehlt, ein qualitativ hochstehendes AC zu konstruieren resp. durchzuführen. Swiss-Assessment will dem abhelfen. Neun Qualitätsstandards geben vor, wie man zu einem qualitativ guten Assessment gelangt. Gemäss dem neunten Qualitätsstandard ist es wichtig, dass das AC resp. die einzelnen AC-Aufgaben auf ihre Güte hin überprüft werden. Dieses Ziel hat sich denn auch *impuls* im Zusammenhang mit der Turmbaufaufgabe gesetzt. Um die Güte eines Verfahrens – der Turmbaufaufgabe – zu evaluieren, stellt sich die Frage, welches die Messmethode ist, der es sich bedient. Im Falle der AC-Aufgabe „Turmbau“ ist dies die Verhaltensbeobachtung. Mit der Ausführung der verschiedenen Arten von Verhaltensbeobachtung kann nun jene ange-

wandte Verhaltensbeobachtung von *impuls* theoretisch eingeordnet werden. Es liegt nahe, dass bei der Turmbauaufgabe die systematische Verhaltensbeobachtung und nicht die freie oder alltägliche angewendet wird. Denn es besteht im Vorfeld eine Hypothese darüber, welche Verhaltensweisen wichtig sind in der Beobachtungssituation und deshalb beobachtet werden sollen. Weiter wird bei der Turmbauaufgabe direkt beobachtet, d.h. ohne Videoaufzeichnung. Demgemäss handelt es sich auch um eine passiv-teilnehmende Beobachtung. Ausserdem erfolgt die Beobachtung bei *impuls* im Sinne einer Laborbeobachtung. Aufgrund der Turmbauaufgabe sollen bestimmte Verhaltensweisen hervorgerufen werden. Bezüglich Strukturiertheitsgrad liesse sich die systematische Verhaltensbeobachtung bei *impuls* passend als halbstrukturiert bezeichnen. Dies aufgrund des Ratingverfahrens, dem sich *impuls* bedient. Zwar wird die Beobachtungssituation bei *impuls* standardisiert und ein Leitfaden eingesetzt, der vorgibt, auf welche Verhaltensweisen die Beobachtenden zu achten haben und wie sie diese beurteilen müssen. Jedoch ist bei der Beobachtung und Beurteilung nach dem Ratingverfahren immer auch ein relativ grosser Grad an Subjektivität im Spiel. Die Anfälligkeit für Beobachtungs- und Beurteilungsfehler ist damit im Gegensatz zu stark strukturierten Verfahren grösser. Um die Fehlerquote so tief wie möglich zu halten, muss das Ratingverfahren im Rahmen des Möglichen strukturiert werden. Denn Beobachtungsfehler können zu einer schlechten Objektivität führen und sich negativ auf die Reliabilität und Validität auswirken. Wichtig ist, dass die Beobachtungseinheiten deshalb adäquat formuliert werden und den Beobachtenden ein Beobachtungsraster gegeben wird, das den im Abschnitt 2.3.4 erwähnten Kriterien gerecht wird. Zudem müssen die Beobachtenden ein Beobachtertraining absolvieren, bei dem sie hinsichtlich der möglichen Beobachtungsfehler sensibilisiert werden und unter anderem auch die Beobachtungseinheiten und das Beobachtungsraster „kennenzulernen“. Alle diese Punkte, wie gut die Beobachter trainiert und die Beobachtungseinheiten definiert sind, wie das Beobachtete quantifiziert und der Beobachtungsleitfaden gestaltet werden, haben Einfluss auf die Objektivität und somit auch auf die Validität der Turmbauaufgabe. Wie dem Abschnitt 2.4 entnommen werden kann, ist es wichtig, auch die Kriterien der Objektivität und Reliabilität zu erheben, um ein umfassendes Bild zur Validität zu erhalten. Deshalb wurden die drei Gütekriterien relativ detailliert ausgeführt. In diesem Zusammenhang wurde aufgezeigt, welche Methoden es gibt, die drei Hauptgütekriterien zu erheben. Wie das methodische Vorgehen in Bezug auf die Turmbauaufgabe aussieht, wird Thema des Kapitels 3

sein. An dieser Stelle der Arbeit muss das methodische Vorgehen hinsichtlich der Validitätsprüfung nun jedoch ein wenig vorweggenommen werden: Die Validität der Turmbauaufgabe wird in der vorliegenden Arbeit mittels Konstruktvalidierung geprüft. Daher wird nachfolgend vor dem Hintergrund der erläuterten Theorie zum Thema Verhaltensdisposition und unter Einbezug weiterer Studien versucht, Hypothesen herzuleiten, die einen Zusammenhang zwischen den vier Fähigkeitsmerkmalen nach MELBA und der Intelligenz sowie den Persönlichkeitsmerkmalen proklamieren.

2.6 Hypothesen

Lernen/Merken

Beginnen wir mit dem ersten Fähigkeitsmerkmal nach MELBA: der Fähigkeit zum Lernen/Merken. Unter dieser Fähigkeit nach MELBA versteht sich, „arbeitsrelevante Informationen auffassen, im Gedächtnis speichern und zu einem gegebenen Zeitpunkt verfügbar machen zu können“ (Föhres et al., 2011, S. 14). Mit dem Thema des Lernens/Merkens resp. mit den Prozessen rund um die Informationsverarbeitung beschäftigt sich die kognitive Psychologie (Fröhlich, 2008). Wie dem Abschnitt 2.2.1 entnommen werden kann, gilt die Merkfähigkeit in der Intelligenzforschung als eine über etliche Intelligenzstrukturmodelle hinweg konvergente Teilfähigkeit der Intelligenz. Demzufolge gibt es viele validierte Tests, mit denen sich dieses Konstrukt messen lässt.

Um herauszufinden, ob sich mit der Turmbauaufgabe die Lern-/Merkfähigkeiten der Teilnehmenden tatsächlich abbilden lassen, sollen dieselben Teilnehmenden nebst der Turmbauaufgabe eine validierte Merkfähigkeits-Testaufgabe bearbeiten. Die Beurteilungen hinsichtlich Lern-/Merkfähigkeit im Rahmen der Turmbauaufgabe sollen dann ganz im Sinne der Konstruktvalidierung mit den Ergebnissen der Merkfähigkeitsaufgaben des Intelligenztests verglichen und auf ihren Zusammenhang geprüft werden. Eine Person, die von den Assessoren bezüglich des Konstrukts Lernen/Merken als „stark“ beurteilt wird, sollte folglich auch in einer Merkfähigkeitsaufgabe eines beliebigen Intelligenztests gut abschneiden. Daraus leitet sich folgende Hypothese ab.

H₁: Zwischen dem Konstrukt Lernen/Merken der Turmbauaufgabe und der allgemeinen Merkfähigkeit aus der Intelligenzforschung besteht ein positiver Zusammenhang.

H₀: Zwischen dem Konstrukt Lernen/Merken der Turmbauaufgabe und der allgemeinen Merkfähigkeit aus der Intelligenzforschung besteht kein positiver Zusammenhang.

Arbeitsplanung

Föhres et al. (2011) verstehen unter der Fähigkeit zur Arbeitsplanung das Vermögen einer Person, „eine Arbeitsaufgabe unter Berücksichtigung der technischen, administrativen und personellen Bedingungen im Hinblick auf ein optimales Zusammenwirken des einzelnen Elements zu einem effektiven Ergebnis zu gliedern und strukturieren zu können“ (S. 1). Nach Föhres et al. (2011) ist die Planungsfähigkeit eine Strategie der Problemlösefähigkeit. Die Problemlösefähigkeit ist laut den Autoren wiederum die Fähigkeit, „neuartige Situationen in ihrer Bedeutung rechtzeitig erkennen und innerhalb einer angemessenen Frist durch den Einsatz verschiedener Strategien (Kreativität, Analyse, Planung etc.) sach- und personengerecht lösen zu können“ (S. 41). Nach Leighton (2004) hängt die Problemlösefähigkeit weiter eng mit dem Konstrukt Schlussfolgerndes Denken zusammen. Demnach ist das Schlussfolgernde Denken ein der Problemlösefähigkeit vorgelagerter kognitiver Prozess. Gehen wir davon aus, wir entdecken eine Pflanze in einem übergrossen Blumentopf, der in einer schattigen Ecke steht und deren Blätter nach und nach abfallen. Wir ziehen den Schluss, dass die Blätter aufgrund mangelnder Sonnenstrahlen abfallen (Schlussfolgerndes Denken). Der nächste Schritt bestünde nun darin, eine Lösung für das Problem zu finden, z.B. muss der übergrosse Blumentopf an einen sonnigen Platz verrückt werden (Problemlösung). Da der Topf jedoch übergross und folglich sehr schwer ist, braucht es einen Plan dazu, wie die Pflanze von A nach B verschoben werden kann (Planungsfähigkeit). Um den Plan zu erarbeiten, bräuchte es womöglich eine erneute Schlussfolgerung (Beispiel in Anlehnung an Leighton, 2004). Laut der Definition von Föhres et al (2011) muss die Planungsfähigkeit eng mit der Problemlösefähigkeit zusammenhängen. Durch die Ausführungen zum Thema schlussfolgerndes Denken in Anlehnung an Leighton (2004) wird zudem ersichtlich, dass die Problemlösefähigkeit und Planungsfähigkeit mit schlussfolgerndem Denken in Verbindung stehen müssen. Schlussfolgerndes Denken ist ein über die Intelligenzforschung hinweg konvergentes Teilkonstrukt der Intelligenz. Deshalb lautet die zu prüfende Hypothese zur Planungsfähigkeit:

H₂: Das Konstrukt Planungsfähigkeit hängt positiv mit dem Konstrukt der schlussfolgernden Denkfähigkeit zusammen.

H₀: Das Konstrukt Planungsfähigkeit hängt nicht positiv mit dem Konstrukt der schlussfolgernden Denkfähigkeit zusammen.

Kritische Kontrolle

„Kritische Kontrolle ist die Fähigkeit, mit dem Arbeitsprozess in Zusammenhang stehende eigene Verhaltensweisen und selbsterbrachte Arbeitsergebnis auf sachbezogene Richtigkeit hin zu prüfen und bewerten zu können“ (Föhres et al. 2011, S. 12). Personen mit einer hohen kritischen Kontrollfähigkeit können sich eigene Fehler eingestehen, kontrollieren die eigene Arbeit selbstständig und fortlaufend. Sie brauchen keine ständige Kontrolle von aussen und können die eigene Leistung sehr gut einschätzen. Das Konstrukt kritische Kontrolle wird deshalb mit folgender Hypothese überprüft.

H₃: Das Konstrukt kritische Kontrolle hängt positiv mit dem Konstrukt Gewissenhaftigkeit zusammen.

H₀: Das Konstrukt kritische Kontrolle hängt nicht positiv mit dem Konstrukt Gewissenhaftigkeit zusammen

Wie aus dem Abschnitt 2.2.2 hervorgeht, zeichnen sich gewissenhafte Personen dadurch aus, dass sie sich sorgfältig und beharrlich verhalten und über ein grosses Mass an Zuverlässigkeit, Selbstdisziplin und Verantwortungsbewusstsein verfügen. Die Definition der kritischen Kontrollfähigkeit nach Föhres et al. (2011) und jene der Gewissenhaftigkeit führen vor Augen, dass die beiden Konstrukte in Verbindung gebracht werden können.

Misserfolgstoleranz

Kann man sich einer Arbeitsaufgabe stetig zuwenden, auch wenn ihre Erfüllung nicht gesichert ist, dann hat man die Fähigkeit zur Misserfolgstoleranz. Eine Person, die über eine geringe Misserfolgstoleranz verfügt, ist quasi bereits bei blossem Anblick der Arbeitsaufgabe blockiert und hat Mühe, mit der Bearbeitung der Aufgabe zu beginnen. Ausserdem gerät sie in Rage oder wird deprimiert, wenn sie die Aufgabe nicht erfolgreich beenden kann oder es auf dem Weg zum Ziel „Holpersteine“ gibt. Eine Person mit einer geringen Misserfolgstoleranz hat Mühe, mit der Aufgabe noch

einmal von vorne zu beginnen, konnte sie zuvor nicht erfolgreich abgeschlossen werden (Föhres et al., 2011). Aus dieser Definition kann gefolgert werden, dass - im vorliegenden Fall die Turmbauaufgabe - wahrscheinlich eine stressauslösende Situation dazu führt, dass sich eine Person im Sinne einer geringen oder grossen Misserfolgstoleranz verhält. Ob und wie eine Person Stress erlebt, hängt wiederum von der Art und Weise ab, wie sie eine potentiell stressige Situation bewertet resp. welchen Coping-Stil sie anwendet. Nach Nerdinger, Blickle und Schaper (2008) gibt es zwei Arten von Stressbewältigungs- resp. Coping-Strategien: das problembezogene und das emotionsbezogene Coping. Personen die einer Stresssituation mit problembezogenem Coping begegnen, zeichnen sich dadurch aus, dass sie konkrete Aktionen unternehmen (z.B. gezielte Informationssuche, Handlungen gegen die Bedrohung etc.), um die potentielle Bedrohung abzuwenden (Nerdinger et al., 2008). Daraus kann gefolgert werden, dass Personen, die eine hohe Misserfolgstoleranz haben, bessere Fähigkeiten im Bereich der Arbeitsplanung aufweisen. Folgende Hypothese wird deshalb aufgestellt.

H₄: Das Konstrukt Misserfolgstoleranz hängt positiv mit dem Konstrukt Fähigkeit zur Arbeitsplanung zusammen.

H₀: Das Konstrukt Misserfolgstoleranz hängt nicht positiv mit dem Konstrukt Fähigkeit zur Arbeitsplanung zusammen.

Da das Konstrukt Arbeitsplanung, wie in H₂ postuliert, positiv mit der Schlussfolgernden Denkfähigkeit zusammenhängen sollte, ist auch ein positiver Zusammenhang zwischen der Misserfolgstoleranz und dem Schlussfolgernden Denken zu vermuten.

H₅: Das Konstrukt Misserfolgstoleranz hängt positiv mit dem Konstrukt schlussfolgerndes Denken zusammen.

H₀: Das Konstrukt Misserfolgstoleranz hängt nicht positiv mit dem Konstrukt schlussfolgerndes Denken zusammen.

Dem problemorientierten Coping steht die emotionsbezogene Coping-Strategie gegenüber. Sie zeichnet sich darin aus, dass aktuelle Bedrohungen resp. Stresssituationen mit Emotionsregulationen entschärft versucht werden (z.B. durch Bagatellisierung, Ablenkung usw.). Zwar bringt emotionsbezogenes Coping nach Nerdinger et al. (2008) eine vorübergehende Entlastung. Da die Turmbauaufgabe jedoch dreissig

Minuten lange dauert, kann davon ausgegangen werden, dass das emotionsbezogene Coping aufgrund der permanenten „Aussetzung“ vor zwei Beobachtenden und des damit verbundenen sozialen Drucks in erster Linie nicht funktionieren wird. Die Person kann sich der Aufgabe nicht entziehen und muss diese zu lösen versuchen. Dadurch wird es wahrscheinlich, dass die Person früher oder später beispielsweise emotional oder frustriert im Sinne von misserfolgsintolerantem Verhalten reagieren wird. Das misserfolgsintolerante Verhalten seinerseits kann quasi wiederum als eine Coping-Strategie im zweiten Durchgang verstanden werden, wenn beispielsweise die Arbeitsplanung nicht erfolgreich war. Verschiedene Autoren konnten belegen, dass das Persönlichkeitsmerkmal Neurotizismus besonders stark mit emotionsorientiertem Coping sowie mit Stressempfinden zusammenhängt (Alarcon, Eschlemann & Bowling, 2009; Hooker et al., 1994, McCrae & Costa, 1986, O'Brien & DeLongis, 1996, zitiert nach Penley & Tomaka, 2002; Penley & Tomaka, 2002). Deshalb wird folgende Hypothese aufgestellt.

H₆: Das Konstrukt Misserfolgstoleranz hängt negativ mit dem Konstrukt Neurotizismus zusammen.

H₀: Das Konstrukt Misserfolgstoleranz hängt nicht negativ mit dem Konstrukt Neurotizismus zusammen.

Bezüglich des Zusammenhangs zwischen Coping und Extraversion sind die Meinungen unterschiedlich. O'Brien & DeLongis (1996, zitiert nach Penley & Tomaka, 2002) konnten in ihrer Studie „*The interactional context of problem-, emotion-, and association-focused coping: the role of the big five personality factors*“ einen positiven Zusammenhang zwischen Extraversion und problemorientiertem Coping bzw. einen negativen Zusammenhang zwischen Extraversion und emotionsorientiertem Coping aufzeigen. Penley und Tomaka (2002) konnten diese Ergebnisse hingegen in ihrer Studie nicht bestätigen. Auch bezüglich des Persönlichkeitsmerkmals Offenheit und Verträglichkeit kommen Penley und Tomaka (2002) zu abweichenden Ergebnissen im Vergleich zu O'Brien & DeLongis (1996, zitiert nach Penley & Tomaka, 2002). Da mittels Konstruktvalidierung Gewissheit darüber geschaffen werden soll, ob die Turmbaufgabe bestimmte Verhaltensweisen abzubilden vermag, soll kein ungewisser Zusammenhang als Grundlage für die Validierung verwendet werden. Denn nach Bortz und Döring (1995) ist eine Konstruktvalidierung nur dann erfolgsversprechend, wenn die Hypothesen Gültigkeit besitzen. Aus diesem Grund fließen die Kriterien

Extraversion, Offenheit und Verträglichkeit nicht in die Konstruktvalidierung ein, weshalb dazu keine Hypothesen formuliert werden.

Einigkeit bezüglich des Zusammenhangs zwischen der Coping-Strategie und den Persönlichkeitsmerkmalen herrscht hingegen hinsichtlich des Merkmals Gewissenhaftigkeit. So konnten Penley und Tomaka (2002) vorgängige Untersuchung bestätigen (z.B. O'Brien & DeLongis, 1996; Watson & Hubbard, 1996, zitiert nach Penley & Tomaka, 2002), wonach gewissenhafte Personen aufgrund ihrer selbstdisziplinierten und zielorientierten Art eher problemorientiert als emotionsorientiert mit einer Stresssituation umgehen. Daraus leitet sich die letzte zu prüfende Hypothese ab

H₇: Das Konstrukt Gewissenhaftigkeit hängt positiv mit dem Konstrukt Misserfolgstoleranz zusammen.

H₀: Das Konstrukt Gewissenhaftigkeit hängt nicht positiv mit dem Konstrukt Misserfolgstoleranz zusammen.

3. Methodisches Vorgehen

Die Fragestellung, ob die Turmbaufaufgabe die vier Fähigkeitsmerkmale abzubilden vermag, soll wie erwähnt anhand der Konstruktvalidierung beantwortet werden. Demnach ist die Turmbaufaufgabe dann valide, wenn die aus dem Zielkonstrukt (Arbeitsplanung, kritische Kontrolle, Lernen/Merken, Misserfolgstoleranz) im Abschnitt 2.6 beschriebenen und aus der Theorie abgeleiteten Hypothesen geprüft und bestätigt werden können (Bortz & Döring, 1995). Das wirft die Frage auf, wie die in den Hypothesen enthaltenen Konstrukte gemessen werden sollen. Mit dieser Thematik befasst sich der Abschnitt 3.1. In den Abschnitten 3.2 - 3.3 wird dargelegt, wie bei der Vorbereitung der Datenerhebung vorgegangen und wie diese durchgeführt wurde. Der Abschnitt 3.4 beschäftigt sich mit der Auswertungsmethodik.

3.1 Operationalisierung

Wie der Theorie zu den Gütekriterien im Abschnitt 2.4 entnommen werden kann, hängt die Validität eng mit der Objektivität zusammen. Um ein ganzheitliches Bild zur Validität zu erhalten, muss dieses Kriterium ebenfalls operationalisiert werden.

3.1.1 Objektivität

Durchführungsobjektivität

Mit Durchführungsobjektivität ist gemäss dem Abschnitt 2.4.1 der Grad der Unabhängigkeit des Testergebnisses vom Verhalten der Untersuchenden gemeint. Im Falle der Turmbaufaufgabe handelt es sich um die Unabhängigkeit der Ergebnisse von den Beobachtenden und derjenigen Person, welche die Instruktionen zur Turmbaufaufgabe gibt. Da in dieser Arbeit die Erhebung der Validität im Zentrum steht, wurde die Durchführungsobjektivität nur grob erfasst. Den Probanden wurde im Anschluss an die Turmbaufaufgabe ein Fragebogen ausgeteilt. Mit diesem Fragebogen konnten die Teilnehmenden auf einer fünf-Punkte-Likert-Skala von *1=stimmt ganz und gar nicht* bis *5=stimmt voll und ganz* rückmelden, wie sie die Turmbaufaufgabe erlebt haben. Der Fragebogen wurde in Anlehnung an den Akzept!-Fragebogen von Kersting (2010) erstellt. Beim Akzept!-Fragebogen handelt es sich um ein hochwertiges mehrdimensionales Instrument, mit dem sich die Akzeptanz diagnostischer Verfahren messen lässt. Es gibt je nach zu evaluierendem Verfahren vier Varianten des Fragebogens. In der vorliegenden Arbeit wurden Fragen des Akzept!-Bogens für das Assessment Center als Referenz genommen. Der Akzept!-Bogen fürs AC umfasst gesamthaft 23

Items, mit denen sich verschiedene Skalen messen lassen, die mit der Akzeptanz gegenüber diagnostischen Verfahren zusammenhängen. Obschon sich mit dem Akzept!-Fragebogen gut die Testfairness und somit ein wichtiges Nebengütekriterium messen liesse, wurde der Fragebogen nicht zu diesem Zweck eingesetzt. Zusätzlich die Nebengütekriterien zu überprüfen, würden den Rahmen dieser Arbeit sprengen. Vielmehr wurden daher einzelne Items des Akzept!-Fragebogens verwendet, die nach Augenschein Aufschluss über Kriterien geben könnten, die möglicherweise die Gesamtergebnisse beeinflussen. Somit wurden mit dem Fragebogen quasi Kontrollvariablen erhoben. Die verwendeten Items finden sich ausgeschrieben in der Tabelle 7. Der Fragebogen findet sich im Anhang B. Der rechten Spalte ist zu entnehmen, welchen Aufschluss die einzelnen Items bringen sollen.

Tabelle 7: Eingesetzte Items des Akzept!-Fragebogens und deren Zweck

Nr.	Item	Gibt Aufschluss über
1	Die Turmbaufaufgabe war klar und verständlich.	Verständlichkeit der Aufgabeninstruktion resp. Anleitung
2	Die Turmbaufaufgabe war zu schwer für mich.	Schwierigkeitsgrad der Aufgabe
3	Den Turm zu bauen, empfand ich als belastend.	Belastungsfreiheit der Turmbaufaufgabe
4	Die Atmosphäre während der Turmbaufaufgabe war – im Rahmen des Möglichen – sehr positiv.	Durchführungsobjektivität, Einfluss der Beobachtenden auf die Situation

Auswertungs- und Interpretationsobjektivität

Die Auswertungsobjektivität wurde mit der im Abschnitt 2.4.1 beschriebenen Beobachterübereinstimmung bzw. Inter-Rater-Reliabilität erhoben. Folglich wurden die von den beiden Beobachtenden auf dem Beobachtungsraster eingetragenen Beurteilungen auf einer Ratingskala von eins bis fünf miteinander verglichen. Da der Beobachtende bei einem Rating im Gegensatz zur Ereignis- oder Zeitstichprobe sogleich eine Interpretation bezüglich Intensität resp. Ausprägung des objektiv registrierten Verhaltens vornimmt, wird die Interpretationsobjektivität mittels der Beobachterübereinstimmung gewissermassen mitgemessen (vgl. Abschnitt 2.4.1).

Es gibt vor allem zwei Faktoren, die bei der systematischen Verhaltensbeobachtung einen Einfluss auf die Auswertungs- resp. Interpretationsobjektivität haben. Dies sind der Grad und die Qualität der Strukturierung der Beobachtung. Wie dem Abschnitt 2.3.4 entnommen werden kann, ist es wichtig, dass den Beobachtenden – vor allem bei sogenannten Ratingverfahren - ein qualitativ hochstehendes Beobachtungsraster

vorliegt, das die genannten Kriterien erfüllt. Das Beobachtungsraster wird zum Schluss der Arbeit im Kapitel 6 noch einmal aufgegriffen.

3.1.2 Reliabilität

Wie aus dem Abschnitt 2.4.2 ersichtlich wird, gibt es zwei Möglichkeiten, die Reliabilität bei Verfahren, die sich der systematischen Verhaltensbeobachtung bedienen, zu prüfen: mittels Retest und Paralleltest. Aufgrund beschränkter Ressourcen auf Seiten von *impuls* war es nicht möglich, den Retest durchzuführen. Die hohe Fluktuation der Klientel von *impuls* hätte einen grossen Stichprobenverlust bei der zweiten Testung herbeigeführt (Personen finden einen Job, haben einen Termin bei einem Arzt, sind max. 3 Monate bei *impuls*, fallen aufgrund von Krankheit aus usw.). Ein entsprechender Paralleltest konnte ferner nicht gefunden werden resp. existiert nicht. Aufgrund dessen wurde entschieden, der Reliabilitätsmessung mit der Erhebung der Inter-Rater-Reliabilität zu genügen (vgl. Abschnitt 3.1.1).

3.1.3 Validität

Es gibt drei Arten von Validität. Im Zentrum dieser Arbeit steht die Erhebung der Konstruktvalidität. Folglich gilt es zu beschreiben, wie die Konstrukte operationalisiert wurden, mit denen die Turmbaufaufgabe auf ihre Validität überprüft wurde. Die Kriterienvalidität kann zur Prüfung des Sachverhalts nicht verwendet werden. Die Inhaltsvalidität, welche eng mit den Kriterien an ein gutes Beobachtungsraster zusammenhängt, wurde nicht erhoben, zumal gemäss Kleffmann et al. (1997) das Beobachtungsraster in Bezug auf die Augenscheinvalidität geprüft wurde. Auch dieser Punkt wird jedoch noch einmal in der Schlussfolgerung im Kapitel 6 kurz aufgegriffen.

KONSTRUKTVALIDITÄT

Fähigkeitsmerkmale nach MELBA

Die vier Fähigkeitsmerkmale *Lernern/Merken*, *Arbeitsplanung*, *kritische Kontrolle* und *Misserfolgstoleranz* wurden mittels der Turmbaufaufgabe erhoben. Probanden bearbeiteten die im Abschnitt 2.1.3 beschriebene Turmaufgabe und wurden dabei von zwei Mitarbeitenden von *impuls* beobachtet und beurteilt. Die beiden Beobachtenden haben eine fünftägige Beobachterschulung bei der Miro GmbH, dem Herausgeber des MELBA-Katalogs, durchlaufen. Zur Dokumentation der Ergebnisse und im Sinne eines Beobachtungsleitfadens verwendeten sie das bereits beschriebene von *impuls* entwickelte Beobachtungsraster. Dieses orientiert sich an dem von Kleffmann et al.

(1997) entwickelten Merkmalskatalog MELBA. Ausserdem verwendeten die Beobachtenden zusätzlich das Handbuch III von Föhres et al. (2011), in dem sich detaillierte Definitionen der zu messenden Fähigkeiten sowie einschlägige Fragen und Hilfestellungen finden. Diese sollen es dem Beobachtenden erleichtern, bestimmte Verhaltensweisen zu entdecken, dem entsprechenden Fähigkeitsmerkmal zuzuordnen und auf seine Ausprägung hin zu beurteilen.

Merkfähigkeit aus Intelligenzforschung

Nach Bortz und Döring (1995) ist eine Konstruktvalidierung nur dann erfolgsversprechend, wenn neben dem zu prüfenden Verfahren (Turmbau) ein ausschliesslich gut gesichertes Instrument eingesetzt wird. Ein Wunsch, der von *impuls* geäussert wurde, bestand darin, dass das Verfahren nicht zu viel Zeit beanspruchen durfte, einerseits um die Belastung auf Seiten der Probanden in Grenzen zu halten (bspw. aufgrund von psychischer Beeinträchtigung der Probanden etc.) und andererseits um den Betrieb von *impuls* zu garantieren (Probanden fehlten während den Erhebungen am Arbeitsplatz).

Wie schon erwähnt beschäftigt sich die Intelligenzforschung unter anderem mit der Merkfähigkeit (siehe Abschnitt 2.2.1). Folglich wurde nach einem passenden Intelligenztest gesucht, der die Merkfähigkeit misst, der ökonomisch ist und der normiert und auf die Güte hin überprüft wurde. Ein passendes Instrument, das alle diese Kriterien erfüllt, fand sich in dem von Liepmann et al. (2007) entwickelten Intelligenzstrukturtest 2000 R (I-S-T 2000 R). Dem I-S-T 2000 R unterliegt ein theoretisch fundiertes und empirisch begründetes Strukturkonzept. Intelligenz *besteht nach Liepmann et al. (2007)* aus 11 Teilkonstrukten resp. 11 Fähigkeiten: verbale Intelligenz, figural-räumliche Intelligenz, numerische Intelligenz, Merkfähigkeit, schlussfolgerndes Denken mit Wissensanteil, verbales Wissen, figural-bildhaftes Wissen, numerisches Wissen sowie fluide und kristallisierte Intelligenz (Liepmann et al., 2007; testzentrale.de, 2015;). Der I-S-T 2000 R ist modular aufgebaut. Der Merkfähigkeitstest kann unabhängig von den restlichen Modulen durchgeführt werden. Die Merkfähigkeitsaufgabe des I-S-T 2000 R „erfasst die Fähigkeit zum aktiven Einprägen und kurzfristigen Wiedererkennen von Informationen“ (Liepmann et al., 2007, S. 82). Weitere bedeutsame Spektren der Gedächtnisfähigkeit werden mit der Merkfähigkeitsaufgabe des I-S-T 2000 R zugunsten der Ökonomie nicht erfasst. Die Merkfähigkeitsaufgabe im Rahmen der Turmbauaufgabe ist jedoch ebenfalls „nur“ zur Prüfung der Leistung des

Kurzzeitgedächtnisses konzipiert (vgl. Beobachtungsraster im Abschnitt 2.1.3). Der Proband wird nach 15 Minuten Pause danach gefragt, was er noch weiss von der Instruktion, womit die kurzfristige Gedächtnisleistung geprüft wird. Deshalb scheint die Merkfähigkeitsaufgabe des I-S-T 2000 R zur Konstruktvalidierung adäquat zu sein. Sie besteht aus zwei Aufgabengruppen - einer verbalen und einer figuralen Aufgabengruppe. Dadurch wird die Merkfähigkeit nach Liepmann et al. (2007) weitgehend unabhängig von materialspezifischen Aspekten erfasst, was ebenfalls von Vorteil ist, da die Probanden verschiedene Bildungsniveaus haben (mehr dazu in später in dieser Arbeit). Die Durchführung der Merkfähigkeitsaufgabe nimmt gesamthaft 11 Minuten in Anspruch (sieben Minuten Bearbeitungszeit und vier Minuten Instruktionszeit).

Schlussfolgerndes Denken

Eine Besonderheit des I-S-T 2000 R von Liepmann et al. (2007) liegt darin, dass sich - etwas vereinfacht ausgedrückt - vier Konstrukte höherer Ordnung erheben lassen. Dies ist zum einen die kristallisierte Intelligenz. Damit ist nach Horn (1988, zitiert nach Liepmann et al., 2007) das Wissen und somit die Fähigkeiten gemeint, die auf die kulturelle Prägung und vor allem auf die Bildung zurückzuführen sind. Weiter gibt es das Schlussfolgernde Denken auf Basis der kristallisierten Intelligenz. Mit dem dem I-S-T 2000 R können ferner das schlussfolgernde Denken mit Wissensanteilen, das nicht gänzlich unabhängig, aber nicht stark vom kristallisierten Wissen beeinflusst ist und das schlussfolgernde Denken ohne Wissen, das auch als fluide Intelligenz bezeichnet wird, erfasst werden. Die fluide Intelligenz ist unabhängig vom Vorwissen des Probanden. Am sinnvollsten wäre es, in dieser Arbeit, die schlussfolgernde Denkfähigkeit ohne Vorwissen zu erheben. Die Probanden weisen, wie schon erwähnt, kulturell unterschiedliche Hintergründe auf und erfuhren somit auch unterschiedliche Bildung oder gar keine Bildung. Um mit dem I-S-T 2000 R die Fähigkeit des schlussfolgernden Denkens zu erfassen, dauert der Test gesamthaft 92 Minuten. Diese Bearbeitungszeit ist für die Stichprobe ungeeignet (zu grosse Belastung und zu langes Fehlen am Arbeitsplatz). Die Probanden müssen neben dem Test zur Erhebung der Merkfähigkeit und dem schlussfolgerndem Denken auch noch weitere Tests resp. Fragebogen bearbeiten. Deswegen wurde der wesentlich ökonomischere Intelligenz-Struktur-Test Screening von Liepmann, Beauducel, Brocke und Nettlenstroth (2012) eingesetzt. Das Screening Verfahren erfasst die verbale, figurale und

numerische Intelligenz und misst aus deren Summierung relativ gut fundiert das schlussfolgernde Denken. Das Verfahren ist nicht dazu gedacht, das mit dem zum I-S-T 2000 R abgebildete Strukturmodell vollständig zu erfassen. So werden beispielsweise die Facetten der Intelligenz im höheren Leistungsbereich nicht differenziert erfasst. Vielmehr dient das Screening Verfahren dazu, Hinweise auf die Fähigkeiten von Extrem- oder Risikopopulationen zu geben resp. Leistungsdefizite zu eruieren. Mit dem Screening wird also quasi eine Gesamtübersicht zu den Fähigkeiten ermittelt. Es gilt zu beachten, dass das mit dem Screening-Test erfasste schlussfolgernde Denken zwar weitgehend unbeeinflusst vom Wissen und somit von Akkulturation und Bildung ist. Möchte man jedoch die schlussfolgernde Denkfähigkeit gänzlich unabhängig von kristallisiertem Wissen erfassen, empfiehlt sich der Einsatz des I-S-T 2000 R. Da jedoch auch die erfolgreiche Planungsfähigkeit im Zusammenhang mit der Turmbaufaufgabe nicht gänzlich unabhängig vom Wissen sein dürfte, macht es durchaus Sinn, das Konstrukt des schlussfolgernden Denkens mit Wissensanteil zwecks Konstruktvalidierung zu verwenden (Liepmann et al., 2012). Der I-S-T 2000 wurde auf die Güte hin überprüft und verfügt über Reliabilitätsschätzungen zwischen .72 und .90. Im Rahmen der Validierung konnte festgestellt werden, dass das Verfahren das schlussfolgernde Denken mit den verbalen, numerischen und figuralen Aufgaben breit erfassen kann. Das Verfahren ist ferner normiert worden. Die Bearbeitungszeit beträgt 26 Minuten.

Persönlichkeitsmerkmale

Zur Erfassung der Dimensionen der Persönlichkeitsstruktur werden heute verschiedene Persönlichkeitsfragebögen eingesetzt, bei denen per Selbsteinschätzung eine mehrdimensionale Beschreibung der Persönlichkeit gemacht wird (Hossiep & Schulz, 2013). Sehr bekannte Vertreter dieser Fragebögen sind die Big-Five-Fragebögen, die auf dem Big-Five-Modell basieren. Auch in dieser Arbeit sollen die Persönlichkeitsmerkmale Gewissenhaftigkeit und Neurotizismus mit einem auf dem Big-Five Modell basierenden Fragebogen erhoben werden, nicht zuletzt deshalb, weil diese gut auf ihre psychometrische Güte hin geprüft wurden. Zu diesem Schluss kommen auch Rammstedt et al. (2012), gemäss denen das Kriterium der Ökonomie jedoch häufig unberücksichtigt bleibt. Persönlichkeitsfragebögen nehmen oftmals viel Zeit in Anspruch, was nicht unproblematisch ist. Der Faktor Zeit ist in den meisten sozialwissenschaftlichen Untersuchungen ein kritischer Kostenfaktor. Auch *impuls* stellt aus

den bereits genannten Gründen den Anspruch, dass die „Testerhebungen“ ökonomisch sein sollten. Das von Rammstedt et al. (2012) entwickelte und validierte Big-Five-Inventory-10 (BFI-10) scheint daher ein geeignetes Instrument zu sein, um die Persönlichkeitsmerkmale zu operationalisieren. Der BFI-10 wurde auf der Grundlage des Big Five-Inventory von John, Donahue und Kentle (1991, zitiert nach Rammstedt et al., 2012) entwickelt. Es misst die fünf Persönlichkeitsmerkmale mit gesamthaft 10 Items resp. 2 Items pro Persönlichkeitsmerkmal mit einer Fünf-Punkte-Likert-Skala. Die Durchführungszeit beträgt 108 Sekunden. Der Fragebogen wurde auf die Gütekriterien überprüft. Nach Rammstedt et al. (2012) sprechen die empirischen Belege dieser Validierungsstudie dafür, „dass der BFI-10 nicht nur eine ökonomische, sondern auch eine reliable und valide Erfassung der Big Five erlaubt“ (S. 21). Einzig im Bereich der Retest-Reliabilität sind Einschränkungen bei den Kriterien Verträglichkeit und Neurotizismus zu verzeichnen. Gemäss Rammstedt et al. (2012) sind diese Einschränkungen jedoch wahrscheinlich auf die grosse Heterogenität der Stichprobe zurückzuführen und nicht zu überbewerten. Denn bei den Kriterien Verträglichkeit und Neurotizismus konnte ein guter empirischer Validitätskoeffizient verzeichnet werden. Mehr zur Diskussion bezüglich der Güte dieses Instrumentes findet sich in Rammstedt et al. (2012). Der BFI-10 weist insgesamt eine ausreichende psychometrische Güte auf und erlaubt eine grobe Messung der individuellen Persönlichkeitsstruktur. Deshalb und weil der Test sehr ökonomisch ist, wurde in dieser Arbeit der BFI-10 zur Erhebung der Persönlichkeitsmerkmale verwendet.

Gemäss Kemper, Beierlein, Bensch, Kovaleva und Rammstedt (2012) hat das sozial erwünschte Verhalten einen beträchtlichen Einfluss auf die Messung von Persönlichkeitsmerkmalen. Soziale Erwünschtheit ist als Tendenz zu verstehen, die darin liegt, dass Personen aufgrund sozialer Normen dazu neigen, überwiegend positive Beschreibungen über die eigene Person abzugeben. Dadurch kann es zu einer Verfälschung der Antworten kommen (Bortz & Döring, 1995; Paulus, 2002, zitiert nach Kemper et al., 2012). Der Aspekt des sozial erwünschten Antwortverhaltens sollte in dieser Arbeit kontrolliert werden. Deshalb wurde die von Kemper et al. (2012) entwickelte Kurzsкала zu Erfassung des sozial erwünschten Antwortverhaltens eingesetzt (KSE-G). Mit dem KSE-G werden mittels je drei Items einerseits das sozial erwünschte Antwortverhalten *Übertreibung positiver Qualitäten* (PQ+) und andererseits *Minimierung negativer Qualitäten* (NQ-) erfasst. Die Skalen PQ+ und NQ- sind Kon-

trollskalen, die aus je drei Items bestehen, die besonders sensibel auf die Tendenz zu sozial erwünschten Antworten reagieren (Bortz & Döring, 1995). Mit den Items werden Eigenschaften oder Verhaltensweisen erfasst, die von der Allgemeinheit resp. Gesellschaft als negativ oder positiv beurteilt werden wie etwa das Beispiel-Item aus der KSE-G „Im Streit bleibe ich stets sachlich und objektiv“ (Bortz & Döring, 1996; Kemper et al., 2012, S. 14). Beantwortet eine Person diese Frage etwa mit „1=Nein, trifft überhaupt nicht zu“, so erscheint diese Antwort eher unglaubwürdig, weshalb hier sozial erwünschtes Verhalten im Sinne von Minimierung negativer Qualität vorläge. Auch die KSE-G von Kemper et al. (2012) wurde auf die Reliabilität und Validität empirisch überprüft. Das Verfahren verfügt über gute Werte in diesem Bereich. Die Durchführungszeit der KSE-G beträgt zwischen einer und eineinhalb Minuten. Die Items wurden im Rahmen dieser Arbeit mit den Items des BFI-10 gemischt. Zum Schluss lag ein Fragebogen mit 16 Items vor, mit dem sich die Big-Five und das sozial erwünschte Antwortverhalten messen liessen (vgl. Anhang C).

3.2 Vorbereitende Datenerhebung und Pretest

1. Treffen

Im Vorfeld der Erhebung haben zwei Besprechungen mit Martin Birchler, Ansprechperson bei *impuls* und AC-verantwortliche Person, stattgefunden. Bei diesen Treffen vor Ort in Lachen SZ ging es darum zu prüfen, ob das im Abschnitt 3.1.1 vorgestellte und ausgearbeitete Vorgehen bezüglich der Operationalisierung adäquat ist. Alle Tests und Fragebögen wurden deshalb ausgedruckt zum Treffen mitgebracht. In diesem Zusammenhang wurde der Praxispartner gefragt, ob die Tests zielgruppengerecht sind. Denn aus einem früheren Kick-off-Meeting mit M. Birchler war bekannt, dass einige Klienten von *impuls* und somit die spätere Stichprobe teilweise über schlechte Deutschkenntnisse verfügt. M. Birchler beurteilte die Testaufgaben und Fragebögen als zielgruppengerecht. Jedoch schlug er vor, für die Bearbeitung der Testaufgaben und die Beantwortung der Fragebögen drei Gruppen zu bilden. Diese Gruppen unterschieden sich nach den Sprachkenntnissen. In die Gruppe 1 wurde eingeteilt, wer über schlechte Deutschkenntnisse verfügt, in Gruppe 2 jene mit mittelmässigen Deutschkenntnissen und in Gruppe 3 Personen mit Deutsch als Muttersprache. Dadurch konnten zielgruppengerechte Instruktionen zu den Testaufgaben gemacht werden.

2. Treffen - Pretest

Beim zweiten Treffen im Rahmen der Vorbereitungen wurde ein Pretest der Turmbauaufgabe durchgeführt. Dadurch sollten eventuelle Mängel der Turmbauaufgabe im Hinblick auf die definitive Datenerhebung behoben werden. Proband des Pretests war der Autor dieser Masterarbeit selbst. Während der Proband die Turmbauaufgabe bearbeitete, wurde er von zwei geschulten Beobachtenden von *impuls* anhand des von *impuls* entwickelten Beobachtungsrasters auf die Fähigkeiten nach MELBA beurteilt. Im Anschluss an den Pretest wurde mit den Beobachtenden über den soeben durchgeführten Pretest reflektiert und in diesem Zusammenhang die „Regeln“ besprochen, nach denen die Assessoren bei der definitiven Erhebung beobachten und beurteilen würden. Dadurch sollte eine bestmögliche Durchführungsobjektivität, Auswertungs- und Interpretationsobjektivität erzielt werden. Die Beobachtenden wurden z.B. dazu angehalten, die Beurteilungen erst am Ende der Beobachtungssituation vorzunehmen und möglichst keinen Einfluss auf die Beobachtungssituation zu nehmen. Ausserdem wurde den Beobachtenden empfohlen, im Vorfeld der Beobachtungen das Handbuch III von Föhres et al. (2011) und die darin enthaltenen Definitionen, Leitfragen und Hilfestellungen zu den vier Kriterien zu studieren. Denn auf dem von *impuls* entwickelten Beobachtungsraster sind nur die Leitfragen zu den jeweiligen Fähigkeitsmerkmalen aufgeführt und keine Hilfestellungen zur Beurteilung der Ausprägung des beobachteten Verhaltens (mehr dazu in Kapitel 6). Die Terminkoordination, das Aufbieten von Probanden und das Erstellen des Zeitplans zum Ablauf der Erhebung wurde von *impuls* organisiert. Die Teilnahme an der Erhebung als Testperson war freiwillig.

3.3 Stichprobe und Datenerhebung

Stichprobe

An der Turmbauaufgabe haben N=42 Personen resp. Klienten von *impuls* teilgenommen. Von den N=42 Teilnehmenden verfügen N=6 Personen über Deutschkenntnisse auf dem Niveau A1⁷, N=15 auf dem Niveau A2, N=3 auf dem Niveau von

⁷ Definition der Deutschniveaustufen (Europäischer Referenzrahmen, 2016):

- Niveau A1 - Anfänger: Versteht und verwendet alltägliche Ausdrücke und einfache Sätze.
- Niveau A2 - Grundlegende Kenntnisse: Versteht häufig gebrauchte Ausdrücke. Kann sich in routinemässigen Situation verständigen
- Niveau B1 – Fortgeschritten: Kann die meisten Situationen in Arbeit, Schule und Freizeit bewältigen
- Niveau B2 – Selbstständige Sprachanwendung: Kann Hauptinhalte komplexer Themen verstehen und versteht unter anderem auch Fachdiskussionen
- Niveau C1-C2: Werden nicht beschrieben, da keiner der Probanden auf diesem Niveau war.

B1-B2 und bei N=18 Personen ist Deutsch die Muttersprache. Jene Personen mit dem Deutschniveau A1 haben keine Schulbildung erfahren. Alle 15 Personen auf dem Niveau A2 haben mindestens die Grundschule in ihrem Herkunftsland besucht. Von den total 21 Personen ab dem Deutschniveau B1 verfügt eine Person über den obligatorischen Grundschulabschluss in der Schweiz, 17 Personen über einen Lehrabschluss mit Fähigkeitszeugnis und drei Personen über eine höhere Ausbildung. Weitere Eigenschaften wurden von den Teilnehmenden nicht erhoben. Dadurch sollte die Anonymität gewährleistet und die Glaubwürdigkeit gegenüber den Probanden gestärkt werden, dass die Daten auch wirklich anonym behandelt werden. Ferner sollte damit das Vertrauen gewonnen werden. Wie bereits erwähnt handelt es sich bei der Stichprobe um Personen ohne Erwerbstätigkeit. Die Gründe dafür sind vielfältig, z.B. Burn-out, psychische Beeinträchtigungen, Alkoholismus, Jobverlust aus wirtschaftlichen Gründen, Unfall etc.

Ablauf der Turmbauaufgabe

Die Turmbauaufgaben wurden während einer Woche durchgeführt. Pro Tag konnten zwischen acht und neun Personen an der Aufgabe teilnehmen. Pro Teilnehmer dauerte die Turmbauaufgabe eine Stunde. Während einer Woche haben N=42 Personen die Turmbauaufgabe bearbeitet, d.h. pro Tag zwischen acht und neun Personen. Dabei wurden sie von denselben Assessoren aus dem Pretest beobachtet und beurteilt. Die Instruktion zur Turmbauaufgabe und das Abfragen der Merkfähigkeit bei den Probanden wurde immer durch denselben Moderator vorgenommen (den Autor dieser Arbeit). Vom Ablauf her gestaltete sich die Turmbauaufgabe wie im Abschnitt 2.1.3 beschrieben. Zur Veranschaulichung wird der Ablauf noch einmal in Tabelle 8 dargestellt.

Tabelle 8: Ablauf der Turmbauaufgabe

Dauer (Minuten)	Was	Beschreibung
5	Instruktion	Der Moderator instruiert den Probanden mithilfe der Anleitung und anhand eines standardisierten Leitfadens zur Turmbauaufgabe (vgl. Anhang D). Der Proband hat die Möglichkeit, Verständnisfragen zu stellen. Die Instruktion findet zu zweit in einem anderen Zimmer statt als jenes, in dem der Proband den Turm bauen wird.
10	Merkaufgabe	Der Proband hat Zeit, sich den Auftrag so gut wie möglich einzuprägen und die Möglichkeit, allfällige Fragen zu stellen.
15	Pause	Der Proband geht in die Pause.
5	Merken/Lernen	Der Moderator bittet den Probanden, das soeben Gelernte so gut wie möglich wiederzugeben, womit das Konstrukt Lernen/Merken nach MEL-BA geprüft wird. Die Beobachtenden machen sich Notizen.
30	Turmbau (Arbeitsplanung, Kritische Kontrolle, Misserfolgstoleranz)	Dem Probanden wird die Anleitung wieder ausgehändigt. Die Beobachtenden geben die Zeit vor und der Proband beginnt mit dem Bau des Turms resp. mit der Planung des Turmbaus. Der Moderator verlässt das Zimmer. Die Beobachtenden teilen dem Probanden fünf Minuten vor Schluss die verbleibende Zeit mit. Während der gesamten Dauer beobachten die Assessoren den Probanden auf die Fähigkeiten und nehmen nach Ablauf der Dauer unabhängig voneinander eine Beurteilung der einzelnen Fähigkeiten auf einer Skala von 1-5 vor.
5	Fragebogen	Der Proband beantwortet den im Abschnitt 3.1.1 beschriebenen in Anlehnung an Kersting (2010) erstellten Fragebogen.

Der Autor dieser Arbeit und die beiden Beobachtenden haben sich am Ende des Tages jeweils kurz ausgetauscht, um allfällige Unklarheiten oder Probleme im Zusammenhang mit der Turmbauaufgabe zu besprechen und Anpassungen vorzunehmen.

Stichprobe bei der Testdurchführung und Fragebogenerhebung

Von den N=42 Personen aus der Turmbauaufgabe konnten noch N=32 Personen am zweiten Teil der Erhebung - der Testdurchführung und Fragebogenerhebung - teilnehmen. Die Ausfallzahl von N=10 ist auf Gründe wie Arzttermin, Vorstellungsgespräch, Job gefunden und deshalb nicht mehr Klient von *impuls* etc. zurückzuführen. Von N=32 verfügen N=6 Personen über Deutschkenntnisse auf dem Niveau A1. Die-

se Personen weisen keine schulische Bildung auf. Weiter befinden sich N=8 auf dem Deutschniveau A2, die allesamt mindestens die Grundschulbildung in ihrem Herkunftsland absolviert haben. N=3 Personen sind auf dem Deutschniveau zwischen B1 und B2. Deutsch als Muttersprache sprechen N=15 Personen. Von den N=18 Personen ab dem Niveau B1 hat eine Person die obligatorische Schulbildung abgeschlossen, 14 Personen haben eine Lehre absolviert und 3 Personen eine höhere Ausbildung.

Ablauf der Testdurchführung und Fragebogenerhebung

Eine Woche nach der Durchführung der Turmbauaufgaben wurden die Testaufgaben resp. Fragebogenerhebungen durchgeführt. Wie im Abschnitt 3.2 angesprochen, wurden die Tests und Fragebogenerhebungen in drei Gruppen durchgeführt. Die Teilnehmenden wurden von *impuls* je nach Sprachniveau in die entsprechende Gruppe eingeteilt. Für die Durchführung der Tests und Fragebogenerhebungen wurden pro Gruppe maximal eineinhalb Stunden einkalkuliert. Nachdem die Probanden über den Ablauf der Erhebung orientiert wurden, bearbeiteten sie zuerst den Merkfähigkeitstest des I-S-T-2000 R. Dieser besteht wie bereits erwähnt aus einer verbalen und figuralen Merkfähigkeitsaufgabe (siehe Abschnitt 3.1.3). Bevor sich die Probanden an die Lösung der Aufgaben machten, wurde ihnen der Aufgabentyp jeweils zuerst anhand von Beispielen des I-S-T 2000 R mit Powerpoint-Slides Schritt für Schritt erklärt. Um es für die Probanden möglichst einfach zu halten, wurde ihnen nicht ein separater Antwortbogen ausgehändigt, wie dies üblich wäre. Stattdessen wurden die Probanden gebeten, alle Lösungen direkt ins Aufgabenblatt einzutragen. Dadurch sollten Fehler aufgrund falschen Übertragens in den Lösungsbogen vermieden werden. Einige der Aufgabenblätter wurden daher angepasst, damit die Lösungen direkt darin eingetragen werden konnten. Nach dem Merkfähigkeitstest wurde das Screening-Verfahren zur Erhebung des schlussfolgernden Denkens durchgeführt. Auch bei diesem Verfahren wurde, bevor die Probanden die einzelnen Aufgabentypen lösten, ein Beispiel gegeben. Sowohl bei der Durchführung der Merkfähigkeitsaufgaben als auch bei den Aufgaben zum schlussfolgernden Denken orientiert man sich an den Vorgaben der Testtherausgebenden. Den Gruppen 1 und 2 wurde aufgrund schlechterer Deutschkenntnisse etwas mehr Zeit zugestanden. Diese Gruppen benötigten denn auch mehr Instruktion und Unterstützung, um die Aufgaben zu lösen.

3.4 Datenaufbereitung

Die Ergebnisse aus der Turmbauaufgabe und dem Fragebogen, mit welchem die Teilnehmenden eine Rückmeldung dazu gaben, wie sie die Turmbauaufgabe empfunden hatten, wurden allesamt mit dem Statistikprogramm von IBM SPSS 22.00 aufbereitet. Der Datensatz wurde auf sogenannte Missing-Values und unbrauchbare Ergebnisse bereinigt. Auch die Ergebnisse des zweiten Teils der Erhebung (Intelligenztest und Persönlichkeitsfragebogen) wurden auf diese Art aufbereitet. Einige Items aus den einzelnen Fragebögen wurden gemäss den Anweisungen der Autoren der einzelnen Instrumente umgepolt und ausserdem zu Skalen zusammengefasst. Informationen dazu, welche Skalen je Erhebungsinstrument umzupolen oder zu Skalen zusammenzufassen sind, findet sich in der genannten Literatur der Test- und Fragebogenherausgeber.

3.5 Datenauswertung

Alle Daten wurden mit SPSS 22.00 mittels folgender Auswertungsmethoden ausgewertet.

Konstruktvalidität

Um die Hypothesen zu prüfen und somit die Konstruktvalidierung durchzuführen, wurde die sogenannte Multitrait-Multimethod-Methode (MTMM) nach Campbell und Fiske (1959, zitiert nach Bortz & Döring, 1995) angewendet. Die MTMM ist eine besondere Variante der Konstruktvalidierung, die erfordert, dass mit mehreren Erhebungsmethoden (Multi-Method) mehrere Konstrukte (Multi-Trait) erfasst werden. Mit der MTMM wird die wechselseitige Beziehung zwischen Konstrukten und Methoden systematisch und nach bestimmten Regeln analysiert. Dadurch wird es möglich, die Höhe der Konstruktvalidität abzuschätzen. Bei der MTMM wird gemäss Bortz und Döring (1995) nach zwei Arten von Konstruktvalidität unterschieden – der konvergen-ten und divergenten Validität.

Konvergente Validität liegt vor, wenn es eine Korrelation zwischen den Messergebnissen verschiedener Methoden zum selben Konstrukt gibt. Oder anders ausgedrückt, wenn die verschiedenen Operationalisierungen der theoretisch begründeten, durch Hypothesen zueinander in Verbindung gebrachten Konstrukte zu ähnlichen Ergebnissen führen, kann von konvergenter Validität gesprochen werden (Bortz & Döring, 1995; Häder, 2006).

Diskriminante Validität besteht dann, wenn sich die Messergebnisse durch dieselbe Methode verschiedener Konstrukte voneinander unterscheiden. Folglich dürften verschiedene Konstrukte, die theoretisch nicht miteinander in Verbindung gebracht werden können, nicht oder nur geringfügig zusammenhängen. Korrelieren die Konstrukte miteinander, würde beispielsweise die Turmbauaufgabe redundante Informationen liefern. Folglich müsste die Aufgabe angepasst oder das Beobachtungsraster resp. das Zielkonstrukt präzisiert werden (Bortz & Döring, 1995).

Die MTMM bedient sich zur systematischen Abschätzung der konvergenten und diskriminanten Validität eines Zusammenhangsmasses wie der Korrelation. Da es sich in dieser Arbeit um nichtnormalverteilte Daten handelt, wurden die Zusammenhänge im Rahmen der MTMM-Analyse mittels der Korrelation nach Spearman gerechnet (Bühl, 2008). Nach Cohen (1992) können Korrelationskoeffiziente (r) wie folgt interpretiert werden.

$r = 0.10 \rightarrow$ kleine Effektstärke
$r = 0.30 \rightarrow$ mittlere Effektstärke
$r = 0.50 \rightarrow$ grosse Effektstärke

Die Ergebnisse werden bei der MTMM zur Interpretation in eine speziellen Korrelationsmatrix, in die sogenannte Multitrait-Multimethod-Matrix, eingetragen (mehr dazu im Ergebnisteil im Kapitel 4) (Bortz & Döring, 1995).

Die Ergebnisse aus der Korrelationsberechnung wurden ferner teilweise mit der Statistik-Software G*Power (<http://www.gpower.hhu.de/>) auf ihre Teststärke untersucht. Ausserdem wurden in Anlehnung an die Ergebnisse - wo angebracht - mit derselben Software optimale Stichprobenumfänge berechnet – dies um ein differenziertes Bild über die Ergebnisse und damit nicht Fehler der 1. Art resp. 2. Art zu erhalten.

Nebst den Korrelationen nach Spearman wurde eine multiple lineare Regression gerechnet, um die Ergebnisse in der MTMM-Matrix interpretieren zu können. Als abhängige Variable wurden alle Fähigkeiten resp. Konstrukte der Turmbauaufgabe in SPSS zu einer Variable zusammengefasst. Der Grund, weshalb die multiple lineare Regression eingesetzt wurde, wird später in dieser Arbeit erläutert (siehe Kapitel 5). Da die multiple Regression intervallskalierte Daten voraussetzt, wurden die Ratingskalen wie intervallskalierte Skalen behandelt (Bühl, 2008; Porst, 2008).

Wie vorangehend beschrieben wurde, kommt es bei Fragebogenerhebungen oftmals zu sozial erwünschtem Antwortverhalten. Um zu kontrollieren, dass die Ergebnisse

des Big-Five-Inventary nicht verfälscht sind und dadurch die Ergebnisse im Rahmen der MTMM verzerren, wurde die Kurzskala zur Erhebung sozial erwünschten Antwortverhaltens von Kemper et al. (2012) eingesetzt. Zur Auswertung dieses Fragebogens wurden Mittelwerte und dazugehörige Standardabweichungen berechnet. Diese wurden mit den empirisch erarbeiteten Referenzwerten von Kemper et al. (2012) verglichen. Dies sollte Aufschluss darüber geben, ob die Probanden sozial erwünscht geantwortet haben.

Die Probanden, welche an der Turmbauaufgabe und bei der Test- und Fragebogendurchführung teilgenommen hatten, verfügen über ein unterschiedliches Deutsch- und Bildungsniveau. Daher wurden diese beiden Kriterien im Sinne von Kontrollvariablen erhoben. Um zu kontrollieren, dass die Ergebnisse im Rahmen der Konstruktvalidierung nicht zufällig resp. durch andere Faktoren bedingt zustande gekommen waren, flossen die beiden Kriterien in die Datenauswertung mit ein.

Auswertungsobjektivität und Reliabilität

Als Mass zur Überprüfung der Auswertungsobjektivität resp. Beobachterübereinstimmung wurde der sogenannte Intra-Klassen-Korrelationskoeffizient verwendet. Denn verwendet man bei Beobachtungen Ratingskalen, wird nach Greve und Wentura (1997) die Beobachterübereinstimmung mittels dieses Koeffizienten ermittelt. Der Intraclass Correlation Coefficient (ICC) ist nach Wirtz und Caspar (2002) „ein Mass für die Stärke des Zusammenhangs der Urteile zweier Rater, die dieselben Personen oder Objekte beurteilt haben“ (S. 189). Laut Greve und Wentura (1997) gibt es verschiedene Auffassungen darüber, ab wann ein Wert als gut zu betrachten ist. Fisseni (1990, zitiert nach Nerdinger et al., 2008) erachtet in Bezug auf die Auswertungsobjektivität Werte von $p=.60$ als niedrig, zwischen $p=.70$ und $p=.90$ als mittelmässig und Werte über $p=.90$ als hoch. Gemäss Bakemann und Gottmann (1986, zitiert nach Greve und Wentura, 1997) werden in der Forschung jedoch meist Werte über $p=.70$ als gut bezeichnet. Folglich werden in dieser Arbeit alle Werte unter $p=.70$ als schlecht, Werte ab $p=.70$ bis $p=.90$ als gut und Werte ab $p=.90$ als ausgezeichnet bezeichnet.

Der ICC setzt mindestens intervallskalierte Daten voraus. Deshalb wurden auch hier die Ratingskalen in Anlehnung an Porst (2008), wonach Ordinal- resp. Ratingskalen als Intervallskalen betrachtet werden können, als solche behandelt. Der Intra-Klassen-Korrelationskoeffizient wurde in SPSS mit dem sogenannten Two-Way Mixed Modell

resp. mit dem justierten ICC berechnet, da alle Probanden jeweils durch dieselben beiden Beobachtenden beobachtet wurden. Mehr zu den verschiedenen Arten des ICC findet sich in Wirtz und Caspar (2002).

Zur Auswertung der Daten aus der Erhebung mit dem Akezept!-Fragebogen wurden Häufigkeitsrechnungen vorgenommen.

4. Ergebnisse

Aus den Erhebungen resultierte ein Klassenkorrelationskoeffizient, der als gut bezeichnet werden kann. Im Rahmen der Konstruktvalidierung kann vereinzelt konvergente Validität festgestellt werden. Diskriminante Validität ist hingegen nicht feststellbar. Diese Ergebnisse werden nun detailliert dargelegt. Zunächst werden im Abschnitt 4.1 die Ergebnisse aus der Erhebung zur Prüfung der Objektivität resp. Reliabilität erläutert. Danach werden die Ergebnisse zu den Hypothesen vorgestellt.

4.1 Objektivität und Reliabilität

In der Tabelle 9 finden sich die Intraklassenkorrelationskoeffizienten, die mit SPSS berechnet wurden. Bei 42 Probanden und jeweils vier zu beurteilenden Fähigkeitsmerkmalen hätten die Beobachtenden je 168 Ratings vorgenommen. Bei 8 Probanden konnte jedoch aufgrund sehr schlechter Deutschkenntnisse die Fähigkeit *Lernen/Merken* nicht beurteilt werden, weshalb die beiden Beobachtenden stattdessen 160 Ratings gemacht haben. In der Tabelle 9 sind einerseits die Intraklassenkorrelationskoeffizienten zu allen Ratings über alle Konstrukte hinweg und andererseits zu den einzelnen Fähigkeitsmerkmalen ersichtlich.

Tabelle 9: Intraklassen-Korrelationskoeffiziente Turmbauaufgabe

Konstrukt	N	Durchschnittliche Korrelation innerhalb der Klasse ^a	95% Konfidenzintervall	
			Untergrenze	Obergrenze
Alle vier Konstrukte	160 ^b	,809	,737	,861
Lernen/Merken	34 ^b	,772	,549	,885
Arbeitsplanung	42	,767	,583	,880
Kritische Kontrolle	42	,786	,604	,884
Misserfolgstoleranz	42	,775	,580	,879

- Intraklassen-Korrelationskoeffizienten des Typs A mit einer Definition einer absoluten Übereinstimmung. Die Schätzung wird unter der Annahme berechnet, dass kein Interaktionseffekt vorliegt, da andernfalls keine Schätzung durchgeführt werden könnte.
- Bei N=8 Personen Lernen/Merken nicht beurteilbar aufgrund sehr schlechter Deutschkenntnisse

In Anlehnung an die im Abschnitt 3.5 beschriebenen Richtwerte, wonach ein Wert ab $p=.70$ als gut und ab $p=.90$ als hervorragend gilt, kann gefolgert werden, dass die Beobachterübereinstimmung resp. die Inter-Rater-Reliabilität über alle Fähigkeitsmerkmale hinweg gut ist. Betrachtet man nun die Konfidenzintervalle beim ICC über alle Ratings hinweg, so wird sichtbar, dass 95% der Fälle einen Intraklassenkorrelationskoeffizienten zwischen $p=.737$ und $p=.861$ aufweisen. Dies ist zufrieden-

stellend. Schaut man die Konfidenzintervalle der einzelnen Kategorien resp. Fähigkeiten an, so liegen einige der Fälle unter der als gut geltenden Grenze von $p=.70$.

Akzept!-Fragebogen und Durchführungsobjektivität

Ohne ins Detail zu gehen, sollen hier kurz die Ergebnisse aus der Erhebung mit dem Akzept!-Fragebogen dargelegt werden. Von insgesamt 42 befragten Personen waren die Ergebnisse zwischen 30 und 31 Personen verwertbar. Bei 11 Personen besteht aufgrund des Antwortverhaltens die Vermutung, dass sie die Fragen sprachlich nicht verstanden haben, weshalb diese Ergebnisse nicht in die Auswertung miteinbezogen wurden.

Item 4 wurde dazu eingesetzt, um Aufschluss über die Qualität der Durchführungsobjektivität zu erhalten. Das Item lautete: *„Die Atmosphäre während der Turmbauaufgabe war – im Rahmen des Möglichen – sehr positiv“*. Von 31 Personen haben 26 Personen die Frage mit dem Skalenwert *4=trifft zu* und *5=trifft voll und ganz zu* beantwortet. Zwei Personen haben den Wert *3=geht so* und drei Personen den Wert *2=trifft nicht zu* angewählt. Keiner der Befragten hat die Frage mit *1=trifft überhaupt nicht zu* beantwortet.

Die weiteren drei verwendeten Items des Akzept!-Fragebogens wurden nicht zur Überprüfung der Objektivität eingesetzt. Trotzdem werden sie an dieser Stelle kurz erwähnt. Auf die Frage, ob die Turmbauaufgabe verständlich war, haben die meisten Personen mit *5=trifft voll und ganz zu* und *4=trifft zu* und nur eine Person *3=geht so* angegeben. Zwei Personen haben die Aufgabe als zu schwierig empfunden, 9 Personen als mittelmässig schwierig und 20 Personen als gar nicht schwierig oder nicht schwierig. Drei Testpersonen fühlten sich belastet durch die Turmbauaufgabe. 8 Personen verspürten eine mittelstarke Belastung und 19 Personen beurteilten die Aufgabe als gar nicht belastend oder kaum belastend (detaillierte Ergebnisse im Anhang E).

4.2 Konstruktvalidität

Auf der Folgeseite ist die sogenannte Multitrait-Multimethod Matrix nach Campbell und Fiske (1959, zitiert nach Bortz & Döring, 1995) abgebildet (siehe Tabelle 10). Darin zu erkennen sind die Korrelationskoeffizienten nach Spearman, das jeweilige Signifikanzniveau und die Grösse der Stichprobe. Um das Prinzip hinter der Multitrait-Multimethod-Matrix nachzuvollziehen, bedarf es nun noch eines kleinen theoretischen Einschubs, bevor die Ergebnisse vorgestellt werden können.

Die MTMM-Matrix kann nach Bortz und Döring (1995, S. 188) in drei unterschiedliche „Blöcke“ unterteilt werden:

- **Monotrait-Heteromethod-Block** (rot umrandet in Tabelle 10)
Ein Konstrukt (Mono-Trait) wird durch mehrere Methoden (Hetero-Method) gemessen.
- **Heterotrait-Monomethod-Block** (gelb schattiert)
Dieselbe Methode misst unterschiedliche Konstrukte und die Messwerte werden korreliert.
- **Heterotrait-Heteromethod-Block** (blau schattiert)
Die Messwerte von mehreren Konstrukten (Hetero-Trait) werden mit unterschiedlichen Methoden (Hetero-Method) korreliert.

Nach Campbell und Fiske (1959, zitiert nach Bortz und Döring, 1995) müssen wie schon im Abschnitt 3.5 beschrieben, konvergente und divergente Validität vorliegen, damit von Konstruktvalidität gesprochen werden kann. Konvergente Validität liegt vor, wenn dasselbe Konstrukt mit mehreren Methoden gemessen wird und die Mittelwerte bedeutsam signifikant grösser sind als Null (rot umrandet in Tabelle 10). Diskriminante Validität liegt vor, wenn die die Monotrait-Heteromethod-Korrelation (rot) grösser ist als die Werte im Heterotrait-Heteromethod-Block in der gleichen Zeile und Spalte (gelb). Zudem sollte die Monotrait-Heteromethod-Korrelation (rot) grösser sein als die Werte der Heterotrait-Monomethod-Korrelation (blau).

Vor diesem theoretischen Hintergrund können die Ergebnisse nun beschrieben werden.

Tabelle 10: Multitrait-Multimethod-Matrix mit Korrelation nach Spearman

		Turmbauaufgabe				Tests & Fragebogen			
		Lernen/Merken	Arbeitsplanung	Kritische Kontrolle	Misserfolgstoleranz	Merkfähigkeit	Schlussfolgerndes Denken	Gewissenhaftigkeit	Neurotizismus
Turmbauaufgabe									
Lernen/Merken	r_s Sig. (2-seitig) N	1							
Arbeitsplanung	r_s Sig. (2-seitig) N	,531** ,001 34	1						
Kritische Kontrolle	r_s Sig. (2-seitig) N	,252 ,150 34	,490** ,001 42	1					
Misserfolgstoleranz	r_s Sig. (2-seitig) N	,363* ,035 34	,555** ,000 42	,193 ,220 42	1				
Tests & Fragebogen									
Merkfähigkeit	r_s Sig. (2-seitig) N	,392 ,071 22	,630** ,002 22	,582** ,005 22	,370 ,090 22	1			
Schlussfolgerndes Denken	r_s Sig. (2-seitig) N	,430 ,052 21	,446* ,043 21	,473* ,030 21	,394 ,077 21	,459* ,042 20	1		
Gewissenhaftigkeit	r_s Sig. (2-seitig) N	-,367 ,093 22	-,032 ,888 22	-,116 ,606 22	,052 ,817 22	,095 ,673 22	-,128 ,590 20	1	
Neurotizismus	r_s Sig. (2-seitig) N	,304 ,169 22	,336 ,126 22	,190 ,398 22	,201 ,369 22	,170 ,451 22	,173 ,465 20	-,281 ,205 22	1

** . Korrelation ist bei Niveau 0.01 signifikant (zweiseitig)

* . Korrelation ist bei Niveau 0.05 signifikant (zweiseitig)

Hinweis zum Stichprobenumfang

Bei der Betrachtung der Ergebnisse muss beachtet werden, dass die Grösse der Stichprobe aufgrund unterschiedlicher Deutschkenntnisse jeweils variiert. Wie schon erwähnt, verfügten die Probanden über unterschiedliche Deutschniveaus, weshalb einige der Verfahren nicht mit allen Probanden durchgeführt werden resp. einige Ergebnisse einzelner Personen nicht verwertbar waren. Bei der Turmbauaufgabe haben alle Personen (N=42) teilgenommen. Bei 8 Personen konnte die Fähigkeit Ler-

nen/Merken im Rahmen der Turmbauaufgabe aufgrund sehr schlechter Deutschkenntnisse nicht beurteilt werden. Diese Probanden konnten das Gelernte nicht in verbalisierter Form wiedergeben oder verstanden die Turmbauanleitung gar nicht erst. Ferner konnte die Merkfähigkeitsaufgabe des I-S-T, der Test zum schlussfolgerndem Denken und die Fragebogenerhebung mit einigen Personen aufgrund schlechter Deutschkenntnisse oder sehr tiefem Bildungsniveau nicht durchgeführt resp. die Daten nicht verwendet werden. Um Verzerrungen der Ergebnisse zu vermeiden, wurden daher nur die Werte der Personen, die Deutsch als Muttersprache sprechen (N=14), berücksichtigt. Hinzu kamen die Werte der Personen, die über das Deutschniveau B1-B2 verfügen (N=3) sowie einige Werte von Personen, die sich auf dem Deutschniveau A2 befinden (N=6). Bei den N=6 Personen auf dem Niveau A2 wurde aufgrund der guten Testergebnisse entschieden, die Daten für die Ausrechnungen zu verwenden. Die Werte einer Person mit Deutsch als Muttersprache aus dem Test zur Erfassung des schlussfolgernden Denkens waren nicht verwertbar.

H₁: Zusammenhang Lernen/Merken und Merkfähigkeit I-S-T 2000 R

Betrachtet man die rot umrandeten Werte in der MTMM-Matrix, wird erkennbar, dass die Ergebnisse aus der Turmbauaufgabe im Bereich Lernen/Merken mit den Ergebnissen aus dem Merkfähigkeitstest des I-S-T 2000 R mit $r_s=.392$ (N=22) korrelieren, was nach Cohen (1992) eine mittelgrosser Zusammenhang ist. Der Wert liegt mit einem Signifikanzniveau von 7.1% jedoch über dem als gut geltenden 5%-Niveau. Zwar scheint es eine Tendenz zu einem positiven Zusammenhang zu geben. Dennoch muss die H_1 , wonach es zwischen dem Konstrukt Lernen/Merken mit der allgemeinen Merkfähigkeit aus der Intelligenzforschung ein positiver Zusammenhang gibt, verworfen werden. Auch die dazugehörige Nullhypothese muss verworfen werden, da die Berechnung der Teststärke mit G-Power unter 80% liegt. Laut Döring und Bortz (2016) kann die Nullhypothese nur dann angenommen werden, wenn die Teststärke mindestens 80% beträgt. Die Poweranalyse mit der Software G*Power ergab weiter, dass es bei diesem Zusammenhang 36 Testpersonen gebraucht hätte, damit die Ergebnisse signifikant geworden wären. Oder anders herum: bei einem N=22 hätte der Korrelationskoeffizient mindestens $r_s=.556$ sein müssen, damit der Zusammenhang auf dem 5%-Niveau signifikant geworden wäre.

Zusammengefasst kann festgehalten werden, dass es keine konvergente Validität gibt. Auch divergente Validität liegt nicht vor. So ist der Korrelationskoeffizient von

$r_s=.392$ (N=22) kleiner als die Korrelation zwischen Lernen/Merken und schlussfolgerndem Denken von $r_s=.430$ (N=21) auf einem Signifikanzniveau von 5.2% oder kleiner wie der Zusammenhang zwischen Lernen/Merken mit der Fähigkeit zur Arbeitsplanung aus der Turmbauaufgabe $r_s=.531$ ($p<.01$, N=34) und der Misserfolgstoleranz von $r_s=.363$ ($p<.05$, N=34). Auch der I-S-T Merkfähigkeitstest korreliert stärker mit der Planungsfähigkeit aus der Turmbauaufgabe als mit Lernen/Merken der gleichen Aufgabe. Objektiv betrachtet kann folglich nicht von Konstruktvalidität die Rede sein.

H₂: Zusammenhang Arbeitsplanung und schlussfolgerndes Denken

Die Ergebnisse aus der Turmbauaufgabe betreffend des Kriteriums Fähigkeit zur Arbeitsplanung nach MELBA korrelieren positiv mit der schlussfolgernden Denkfähigkeit aus dem I-S-T Screening. Hier liegt konvergente Validität von $r_s=.446$ ($p<.05$, N=21). H_2 , wonach das Konstrukt Planungsfähigkeit mit dem Konstrukt der schlussfolgernden Denkfähigkeit zusammenhängt, kann folglich angenommen und H_0 verworfen werden. Auch hier liegt jedoch keine divergente Validität vor. So korrelieren die Ergebnisse der Fähigkeit zur Arbeitsplanung etwa stärker mit der Fähigkeit zur kritischen Kontrolle und vor allem stark mit der Fähigkeit Lernen/Merken aus der Turmbauaufgabe sowie der Misserfolgstoleranz. Konstruktvalidität nach Campbell und Fiske (1959, zitiert nach Bortz & Döring, 1995) ist auf den ersten Blick auch hier nicht gegeben.

H₃: Zusammenhang Kritische Kontrolle und Gewissenhaftigkeit

Es wurde ein Zusammenhang vermutet, dass die Fähigkeit zur kritischen Kontrolle positiv mit der Gewissenhaftigkeit einer Person zusammenhängt. Wie der Tabelle 10 entnommen werden kann, gibt es einen nicht signifikanten negativen Zusammenhang. Folglich muss H_3 verworfen werden. Bei einem Zusammenhang von $r_s=-.116$ hätte es N=581 Personen benötigt, damit das Ergebnis signifikant ausgefallen wäre. Hier wäre bei N=22 ein $r_s=.556$ nötig gewesen, damit der Zusammenhang auf dem 5%-Niveau signifikant geworden wäre. Konvergente Validität ist somit nicht gegeben. Eine Poweranalyse ergab, dass die Teststärke bei .116 liegt, weshalb auch die dazugehörige H_0 verworfen werden muss. Auch diskriminante Validität liegt nicht vor. Daher kann keine Konstruktvalidität angenommen werden.

H₄: Zusammenhang Misserfolgstoleranz und Arbeitsplanung

Der Zusammenhang zwischen der Fähigkeit zur Misserfolgstoleranz und Arbeitsplanung ist hochsignifikant positiv und beträgt $r_s=.555$ ($p<.01$, $N=42$). H_4 kann demnach bestätigt werden. Jedoch ist der Zusammenhang zwischen der Fähigkeit zur Arbeitsplanung und zum Lernen/Merken mit $r_s=.630$ ($p<.01$, $N=22$) grösser, weshalb keine diskriminante Validität vorliegt.

H₅: Zusammenhang Misserfolgstoleranz und schlussfolgerndes Denken

Die Misserfolgstoleranz hängt positiv mit einem Wert von $r_s=.394$ ($N=21$) mit der schlussfolgernden Denkfähigkeit zusammen. Die Ergebnisse sind jedoch nicht signifikant, weshalb H_5 verworfen werden muss. Gemäss Poweranalyse hätten 48 Personen untersucht werden müssen, um ein signifikantes Ergebnis nachweisen zu können. Die Nullhypothese kann jedoch ebenfalls nicht bestätigt werden. Die errechnete Teststärke mit G-Power mit rund 60% liegt unter dem kritischen Wert von 80%. Auch diskriminante Validität liegt nicht vor.

H₆: Zusammenhang Misserfolgstoleranz und Neurotizismus

Zwischen Misserfolgstoleranz und Neurotizismus wurde ein negativer Zusammenhang vermutet. Der MTMM-Matrix kann entnommen werden, dass ein Zusammenhang von $r_s=.201$ ($N=22$) zustande gekommen ist. Dieser ist jedoch nicht signifikant. Die Hypothese muss somit abgelehnt werden. Die Teststärkenanalyse mit G-Power ergibt jedoch, dass auch die Nullhypothese nicht zutrifft.

H₇: Zusammenhang Misserfolgstoleranz und Gewissenhaftigkeit

Die Ergebnisse sind nicht signifikant. Deshalb müssen der vermutete positive Zusammenhang zwischen der Misserfolgstoleranz und der Gewissenhaftigkeit und somit auch H_7 verworfen werden. Auch die Nullhypothese kann abgelehnt werden.

Weitere Ergebnisse: Zusammenhang Sprache und Bildung mit Konstrukten

Zusätzlich wurden Koeffiziente gerechnet, um einen allfälligen Zusammenhang zwischen den zu validierenden Konstrukten und dem Sprach- und Bildungsniveau aufzudecken – im Sinne von Kontrollvariablen. Die Ergebnisse finden sich abgebildet in der Korrelationstabelle nach Spearman in der Tabelle 11.

Tabelle 11: Zusammenhänge Sprachen und Konstrukte (Korrelation nach Spearman)

		Deutsch- kenntnisse	Bildungs- niveau	Arbeits- planung	Kritische Kontrolle	Misserfolg stoleranz	Lernen/ Merken	Merkfähig- keit gesamt	Schlussfol- gerndes Denken
Deutschkenntnisse	r	1.000	,837**	,562**	.159	.246	,601**	.199	,488*
	Sig. (2-seitig)		.000	.000	.313	.116	.000	.374	.025
	N	42	42	42	42	42	34	22	21
Bildungsniveau	r	,837**	1.000	,519**	.197	,313*	,512**	,443*	,586**
	Sig. (2-seitig)	.000		.000	.211	.044	.002	.039	.005
	N	42	42	42	42	42	34	22	21
Arbeitsplanung	r	,562**	,519**	1.000	,490**	,555**	,531**	,630**	,446*
	Sig. (2-seitig)	.000	.000		.001	.000	.001	.002	.043
	N	42	42	42	42	42	34	22	21
Kritische Kontrolle	r	.159	.197	,490**	1.000	.193	.252	,582**	,473*
	Sig. (2-seitig)	.313	.211	.001		.220	.150	.005	.030
	N	42	42	42	42	42	34	22	21
Misserfolgstoleranz	r	.246	,313*	,555**	.193	1.000	,363*	.370	.394
	Sig. (2-seitig)	.116	.044	.000	.220		.035	.090	.077
	N	42	42	42	42	42	34	22	21
Lernen/Merken	r	,601**	,512**	,531**	.252	,363*	1.000	.392	.430
	Sig. (2-seitig)	.000	.002	.001	.150	.035		.071	.052
	N	34	34	34	34	34	34	22	21
Merkfähigkeit gesamt	r	.199	,443*	,630**	,582**	.370	.392	1.000	,459*
	Sig. (2-seitig)	.374	.039	.002	.005	.090	.071		.042
	N	22	22	22	22	22	22	22	20
Schlussfolgerndes Denken	r	,488*	,586**	,446*	,473*	.394	.430	,459*	1.000
	Sig. (2-seitig)	.025	.005	.043	.030	.077	.052	.042	
	N	21	21	21	21	21	21	20	21

** Korrelation ist bei Niveau 0,01 signifikant (zweiseitig).

* Korrelation ist bei Niveau 0,05 signifikant (zweiseitig).

Auffällig ist, dass es zwischen der Fähigkeit zur Arbeitsplanung mit $r_s=.562$ ($p<.01$, $N=42$) und Lernen/Merken mit $r_s=.601$ ($p<.01$, $N=43$) grosse Zusammenhänge gibt. Des Weiteren gibt es bei der schlussfolgernden Denkfähigkeit gibt es mit $r_s=.488$ ($p<.05$, $N=21$) einen positiven Zusammenhang. Das Bildungsniveau korreliert in ähnlicher Höhe signifikant positiv mit der Fähigkeit zur Arbeitsplanung, Lernen/Merken und der schlussfolgernden Denkfähigkeit. Im Gegensatz zum Deutschniveau kann beim Bildungsniveau ausserdem ein signifikant positiver Zusammenhang von $r_s=.443$ ($p<.05$, $N=22$) mit der Merkfähigkeit festgestellt werden.

Einfluss einzelner Turmbaufähigkeiten auf die Gesamtleistung Turmbau

Neben allen soeben beschriebenen Resultaten verschiedener Ausrechnungen, werden nun auch die Ergebnisse aus der multiplen linearen Regressionsanalyse ausgeführt. Die Fähigkeiten Arbeitsplanung, Kritische Kontrolle, Misserfolgstoleranz und Lernen/Merken wurden wie schon erwähnt zu einer Variable zusammengefasst und bilden damit die abhängige Variable. Betrachtet man den Beta-Koeffizienten in der Tabelle 12, wird deutlich, dass die unabhängige Variable Arbeitsplanung die Varianz

der abhängigen Variable *Turmbau gesamt* am besten aufzuklären vermag, gefolgt von der UV Kritische Kontrolle. Die unabhängige Variable Misserfolgstoleranz und Lernen/Merken klären beide gleichviel Varianz auf.

Tabelle 12: Ergebnisse der multiplen linearen Regressionsanalyse *Turmbau*

	Koeffizienten ^a												
	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.	Konfidenzintervall für B (95,0%)		Korrelationen			Kollinearitätsstatistik		
	B	Standardfehler	Beta			Untergrenze	Obergrenze	Nullte Ordnun	Partiell	Teil	Toleranz	VIF	
(Konstante)	.069	.064		1.074	.292	-.063	.201						
Arbeitsplanung	.297	.022	.416	13.525	.000	.252	.342	.891	.929	.281	.458	2.184	
Kritische Kontrolle	.240	.017	.353	14.212	.000	.205	.274	.685	.935	.296	.701	1.427	
Misserfolgstoleranz	.224	.020	.284	11.448	.000	.184	.264	.632	.905	.238	.703	1.422	
Lernen/Merken	.215	.018	.289	12.012	.000	.178	.251	.677	.912	.250	.748	1.336	

a. abhängige Variable: *Turmbau gesamt*

Soziale erwünschtes Antwortverhalten

Mit je drei Items wurden bei N=22 Personen mittels Fragebogen mit Antwortskalen von *1=stimmt überhaupt nicht* bis *5=stimmt voll und ganz* die Tendenz zum verzerrten Antwortverhalten erhoben.

Aus den Antworten zur Skala *Minimieren von negativen Qualitäten* resultiert ein Mittelwert von M=2.45 (SD = 1.00). Dieser Wert liegt stark unter dem Referenzwert von M=3.19 (SD = .80) der 35-65-Jährigen von Kemper et al. (2012) aus einer Zufallsstichprobe von N=566. Die Probanden haben also Fragen, die ein gesellschaftlich unerwünschtes Verhalten suggerieren, eher verneint. Demgegenüber resultierte aus der Befragung mit den Items zur Skala *Übertreibung positiver Qualitäten* ein M=3.83 (SD = .61), was deutlich über dem Referenzwert von M=2.73 (SD = .74) ist.

Anmerkung Persönlichkeitsmerkmale

Mit dem eingesetzten Fragebogen BFI-10 wurden alle fünf Persönlichkeitsmerkmale erhoben. Aufgrund der in der Theorie unklaren Zusammenhänge der zu prüfenden Zielkonstrukte mit den Persönlichkeitsmerkmalen Extraversion, Offenheit und Verträglichkeit wurden diese nicht in die obigen Ergebnisse miteinbezogen (vgl. Abschnitt 3.1.3). Eine Korrelationstabelle, in der auch diese Persönlichkeitsmerkmale den verschiedenen Konstrukten gegenübergestellt werden, findet sich in Anhang F.

5. Diskussion

Die Ergebnisse zu den einzelnen Hypothesen 1-7 werden nun noch einmal zusammengefasst. Danach werden die Ergebnisse interpretiert. Die Interpretation verfolgt unterteilt in drei Schritten. Die Ergebnisse im Zusammenhang mit Merken/Lernen und Arbeitsplanung werden gemeinsam diskutiert. Danach folgt die Diskussion der Ergebnisse zum Konstrukt kritische Kontrolle und schliesslich zur Misserfolgstoleranz.

- 1.) H_1 wird verworfen. Positiver Zusammenhang Lernen/Merken und Merkfähigkeit jedoch tendenziell vorhanden. Keine diskriminante Validität.
- 2.) H_2 wird bestätigt. Positiver Zusammenhang Arbeitsplanung und Schlussfolgerndes Denken vorhanden, jedoch keine divergente Validität.
- 3.) H_3 wird verworfen. Positiver Zusammenhang kritische Kontrolle und Gewissenhaftigkeit nicht bestätigt.
- 4.) H_4 wird bestätigt. Positiver Zusammenhang Misserfolgstoleranz und Arbeitsplanung vorhanden.
- 5.) H_5 wird verworfen. Positiver Zusammenhang Misserfolgstoleranz und schlussfolgerndes Denken jedoch tendenziell vorhanden.
- 6.) H_6 wird verworfen. Kein negativer Zusammenhang zwischen Misserfolgstoleranz und Neurotizismus.
- 7.) H_7 wird verworfen. Kein positiver Zusammenhang zwischen Misserfolgstoleranz und Gewissenhaftigkeit.

Lernen/Merken und Arbeitsplanung

Der Zusammenhang zwischen Lernen/Merken der Turmbaufaufgabe und der Merkfähigkeit ist knapp nicht signifikant. Konvergente Validität ist damit nicht vorhanden, weshalb H_1 abgelehnt wird. Dies dürfte auf die kleine Stichprobe und der damit verbundenen kleinen Teststärke zurückzuführen sein. Auch ist die divergente Validität nicht gegeben. Daher ist die Turmbaufaufgabe als nicht konstruktvalide bezüglich der Facette Lernen/Merken einzustufen. Interessant ist, dass es innerhalb der Turmbaufaufgabe zwischen dem Konstrukt Lernen/Merken und dem Konstrukt Arbeitsplanung einen signifikanten und grossen Zusammenhang gibt. Die Fähigkeit zur Arbeitsplanung wiederum hängt mit der schlussfolgernden Denkfähigkeit zusammen. Dadurch kann H_2 angenommen werden. Jedoch besteht ein noch grösserer und hochsignifi-

kanter Zusammenhang zwischen der Arbeitsplanung und der Merkfähigkeit, der Fähigkeit zur kritischen Kontrolle und der Misserfolgstoleranz. Der Zusammenhang zwischen der Misserfolgstoleranz und der Fähigkeit zur Arbeitsplanung ist in Anlehnung an die Theorie folgerichtig. H_4 kann daher ebenfalls angenommen werden. Dies könnte auch den signifikanten Zusammenhang zwischen Lernen/Merken und der Misserfolgstoleranz erklären. Dieser ist wahrscheinlich die Folge des theoretisch begründeten Zusammenhangs zwischen der Arbeitsplanung und der Misserfolgstoleranz. Auch der Zusammenhang zwischen der Arbeitsplanung mit der kritischen Kontrollfähigkeit ist einleuchtend. Der grosse Zusammenhang zwischen Arbeitsplanung und Lernen/Merken sowie zwischen der Arbeitsplanung und Merkfähigkeit bleibt jedoch nach wie vor das grosse Fragezeichen. Eine mögliche Erklärung könnte sein, dass die Probanden bereits während dem Einprägen der Turmbau-Anleitung schlussfolgernd denken. So fiel dem Autor als aktiv-teilnehmender Beobachter während den Instruktionen zum Turmbau auf, dass sich einige Personen während dem Einprägen der Anleitung - teilweise lautstarke - Gedanken machten, wie sie den Turm bauen könnten. Dies würde den tendenziell grossen Zusammenhang zwischen Lernen/Merken und darüber hinaus den signifikanten Zusammenhang zwischen der Merkfähigkeit und schlussfolgerndem Denken erklären. Die Korrelation zwischen Lernen/Merken mit der Arbeitsplanung ist jedoch nach wie vor grösser als mit dem schlussfolgernden Denken. Die Verbindung zwischen Lernen/Merken und Arbeitsplanung ist deshalb nicht gänzlich erklärt und es stellt sich die Frage der Trennschärfe.

Interessant ist, dass das Deutschniveau sowohl mit Lernen/Merken und der Arbeitsplanung der Turmbauaufgabe in Verbindung steht. Daraus könnte gefolgert werden, dass Personen, die über gute Deutschkenntnisse verfügen, von den Ratern besser im Bereich Lernen/Merken und Arbeitsplanung beurteilt wurden. Dies könnte unter Umständen auf einen Beobachtungsfehler zurückzuführen sein, wonach einer Person, die über geringe Deutschkenntnisse verfügt, auch schlechtere Fähigkeiten zugeschrieben werden (Halo-Effekt). Näher liegt jedoch die Begründung, dass Personen mit schlechteren Deutschkenntnissen schlicht und einfach Mühe haben, die Aufgabe zu verstehen (Anleitung in Deutsch mit nicht alltäglichen Wörtern etc.) und deshalb schlechter im Bereich des Fähigkeit Merken/Lernen und als Folge davon schlechter bei der Arbeitsplanung und kritischen Kontrolle abschneiden. Fakt ist nun jedoch, dass *impuls* die Fähigkeiten im Hinblick auf eine Wiedereingliederung in den

Arbeitsmarkt abklärt. Auch wenn die Ergebnisse womöglich durch die Deutschkenntnisse verzerrt werden, ergibt sich daraus dennoch eine praktische Implikation. So werden es Personen bei der Arbeit vermutlich schwer haben, wenn sie innerhalb des damit verbundenen Rahmens aufgrund von schlechten Deutschkenntnissen ihre Fähigkeiten nicht ausschöpfen können.

Kritische Kontrolle

Zwischen der kritischen Kontrolle und der Gewissenhaftigkeit konnte entgegen den Erwartungen kein signifikanter Zusammenhang festgestellt werden. Dies könnte wahrscheinlich auf die kleine Stichprobe bei der Fragebogenerhebung zurückzuführen sein. Auch wenn die Ergebnisse aus der Fragebogenerhebung nicht signifikant und die Ergebnisse somit mit Vorsicht zu geniessen sind, resultierte aus der Korrelationsberechnung zwischen der Fähigkeit zur kritischen Kontrolle und Gewissenhaftigkeit überraschenderweise ein negativer Zusammenhang. Dies steht im Widerspruch zur Theorie. Gründe dafür könnten in einer Unzulänglichkeit beim Big-Five-Inventory von Rammstedt et al. (2012) liegen. Der BFI-10 misst die Gewissenhaftigkeit mit zwei Items: 1.) *Ich erledige Aufgaben gründlich* und 2.) *Ich bin bequem, neige zu Faulheit*. Ob mit diesen beiden Items in dieser Arbeit das Konstrukt Gewissenhaftigkeit ganzheitlich repräsentiert wird, ist – trotz der von Rammstedt et al. (2012) dem Fragebogen zugeschriebenen Güte – fraglich. Den Ergebnissen in Kapitel 4 kann aufgrund des Mittelwertvergleichs mit dem Referenzwert von Kemper et al. (2012) entnommen werden, dass die Probanden zu verzerrtem Antwortverhalten tendieren. Es kann daher vermutet werden, dass einige Probanden auf die Frage „Ich bin bequem, neige zu Faulheit“, welche ein sozial unerwünschtes Verhalten suggeriert, mit sozial erwünschtem Antwortverhalten reagiert haben. Dies könnte wiederum zu einer Verzerrung der Ergebnisse in Bezug auf die Gewissenhaftigkeit geführt haben.

Kritische Kontrolle hängt gemäss den Ergebnissen wie schon erwähnt mit der Planungsfähigkeit zusammen, grenzt sich im Gegensatz zur Fähigkeit zur Arbeitsplanung jedoch von dem Lernen/Merken ab. Somit besteht Trennschärfe zum Konstrukt Lernen/Merken. Die kritische Kontrolle korreliert ferner mit schlussfolgerndem Denken, hingegen nicht mit der Misserfolgstoleranz. Dadurch wird klar, dass es zwar einen Zusammenhang zwischen der kritischen Kontrolle und der Arbeitsplanung gibt, mit diesen beiden Konstrukten jedoch nicht dasselbe gemessen wird.

Misserfolgstoleranz

Die Misserfolgstoleranz wurde anhand des Persönlichkeitsmerkmals Neurotizismus geprüft. Hier kristallisierte sich jedoch kein signifikant negativer Zusammenhang heraus. Es zeichnete sich entgegen der Vermutung ein positiver Zusammenhang ab. Dieses Resultat könnte ebenfalls in der Tendenz zur Antwortverzerrung im Sinne der sozialen Erwünschtheit begründet sein, nicht zuletzt aber auch mit der kleinen Stichprobe. Auch H_7 , wonach die Gewissenhaftigkeit positiv mit der Misserfolgstoleranz zusammenhängt, konnte nicht bestätigt werden. Dies dürfte ebenfalls auf das sozial erwünschte Antwortverhalten zurückzuführen sein.

Spannend ist wie schon erwähnt, dass die Misserfolgstoleranz stark mit der Fähigkeit zur Arbeitsplanung zusammenhängt. Geht man von der Theorie aus, dass Personen mit einer grossen Misserfolgstoleranz eher dazu neigen, besser planen zu können oder es aufgrund besserer Planung gar nicht zu einem Verhalten im Sinn einer schlechten Misserfolgstoleranz kommt, so könnte durchaus gefolgert werden, dass die Turmbauaufgabe diese Fähigkeit abzubilden vermag.

Zusammenfassend kann festgehalten, dass Konstruktvalidität nach den Kriterien von Campbell und Fiske (1959, zitiert nach Bortz & Döring, 1995) gesamthaft nicht erzielt wurde. Jedoch gibt es klare Tendenzen, dass die Turmbauaufgabe die Fähigkeiten abzubilden vermag. Nach Leighton (2012) ist schlussfolgerndes Denken als ein der Problemlösefähigkeit vorgelagerter Prozess und Föhres et al. (2011) beschreiben die Planungsfähigkeit als Strategie der Problemlösefähigkeit. Verbindet man diese Ansichten mit der Theorie des Copings, machen die Ergebnisse durchaus Sinn. Die Zusammenhänge könnten etwa wie folgt interpretiert werden: Personen mit einer grossen Misserfolgstoleranz sind eher in der Lage, auch in stressigen Situationen problembezogenes Coping anzuwenden. Personen, die problembezogen „copen“, wenden eine konkrete „Bedrohung“ mit konkreten Aktionen ab. Dies tun sie, indem sie einen Plan erstellen, wie sie das Problem lösen können (Arbeitsplanung). Damit die Personen ihren erarbeiteten „Plan“ erfolgreich umsetzen können, braucht es eine stetige Kontrolle der eigenen Arbeitsschritte (kritische Kontrolle). Dadurch kann das Problem resp. die Aufgabe erfolgreich gelöst werden und es kommt nicht zu den mit Misserfolgstoleranz negativ assoziierten Verhaltensweisen.

Der Zusammenhang zwischen Lernen/Merken und Arbeitsplanung könnte sich wie schon erwähnt dadurch erklären, dass bereits während dem Auswendiglernen der

Turmbauanleitung geschlussfolgert und geplant wird, wie die Aufgabe gelöst werden kann. Da es keinen Zusammenhang zwischen Lernen/Merken und der kritischen Kontrolle gibt, kann darauf geschlossen werden, dass die Fähigkeit zur kritische Kontrolle erst später im Prozess während dem Bau des Turms sichtbar wird. Der Zusammenhang zwischen Lernen/Merken und der Misserfolgstoleranz könnte ausserdem darin begründet liegen, dass Personen, die bereits während der Einprägung der Aufgabe mit Planen beginnen, die Arbeit erfolgreicher lösen und es deshalb nicht zu Verhalten im Sinne von Misserfolgsintoleranz kommt. Aus den Ergebnissen der multiplen linearen Regression geht hervor, dass die Arbeitsplanung am meisten Varianz der gesamten Turmbauaufgabe zu erklären vermag und dies folglich ein zentrales Kriterium für die erfolgreiche Bearbeitung der Turmbauaufgabe ist.

6. Schlussfolgerung und Ausblick

In dieser Arbeit wurde versucht herauszufinden, ob die Turmbauaufgabe die vier Fähigkeiten Arbeitsplanung, kritische Kontrolle, Lernen/Merken und Misserfolgstoleranz abzubilden vermag. Dazu wurde die entsprechende Fragestellung „Lassen sich mit der Turmbauaufgabe die vier Fähigkeiten Arbeitsplanung, kritische Kontrolle, Lernen/Merken und Misserfolgstoleranz abbilden?“ formuliert. Alles in Allem muss die Frage vor dem Hintergrund der diskutierten Ergebnisse mit Nein beantwortet werden. Zwar bildet die Turmbauaufgabe bestimmte Verhaltensweisen ab, aber aufgrund der teils widersprüchlichen Ergebnisse kann kein eindeutiger Schluss gezogen werden, was abgebildet wird resp. was mittels Ratingskala durch die Beobachtenden gemessen wird.

Zu Beginn der Arbeit wurde zuerst die Fragestellung theoretisch eingebettet. In diesem Zusammenhang wurde die Assessmentaufgabe „Turmbau“ vorgestellt, deren Ablauf erklärt und mit der Anleitung veranschaulicht. Auch wurden die vier Fähigkeiten samt den dazugehörigen Definitionen, Leitfragen und Hilfestellungen nach MEL-BA erörtert, die es mit dem Turmbau zu messen gilt und die Turmbau-Anleitung abgebildet. In diesem Zusammenhang wurde denn auch das von *impuls* entwickelte Beobachtungsraster vorgestellt. Der Abschnitt 2.1.3 beschäftigte sich mit dem Problem der zunehmend sinkenden Qualität von AC's. Swiss-Assessment versucht diesem Problem mit neun Qualitätsstandards entgegenzuwirken. In dieser Arbeit wurde folglich versucht, Qualitätsstandard Nummer neun resp. die Güte einer AC-Aufgabe zu prüfen. Zwar kann aufgrund der Datenlage nicht gesagt werden, dass die Turmbauaufgabe valide ist. Jedoch können nun aufgrund der Untersuchung Empfehlungen zuhanden *impuls* im Hinblick auf das weitere Vorgehen gemacht werden.

Da keine eindeutige Konstruktvalidität aufgezeigt werden konnte, stellt sich die Frage, ob die Turmbauaufgabe überhaupt geeignet ist, die vier Fähigkeiten zu erheben. Der Autor der Arbeit ist der Ansicht, dass die Turmbauaufgabe sich dazu gut eignet, jedoch noch weitere Nachforschungen notwendig sind, um dies zu verifizieren. Die Ergebnisse zeigen einen Zusammenhang zwischen der Fähigkeit Arbeitsplanung und Lernen/Merken. Es wird vermutet, dass dieser Wert die Folge davon ist, dass bereits während dem Lernen/Merken schlussfolgerndes Denken und Arbeitsplanung stattfinden und dass die individuelle Ausprägung im Bereich der Misserfolgstoleranz bestimmt, wie eine Situation oder ein Problem bewertet wird und wiederum entschei-

dend dafür sein könnte, wie eine Arbeit geplant wird. Damit die Arbeitsplanung erfolgreich ist, braucht es kritische Kontrolle. Dies alles entscheidet darüber, ob ein Problem erfolgreich gelöst wird. Dies wiederum hat Einfluss darauf, ob es zu einem Misserfolgserlebnis oder Erfolgserlebnis kommt. Aufgrund der Ergebnisse und der in dieser Arbeit dargelegten Theorie gibt es Tendenzen, wonach es diese soeben geschilderten Zusammenhänge gibt. So konnten beispielsweise hochsignifikant positive Zusammenhänge zwischen Misserfolgstoleranz und Arbeitsplanung aufgezeigt werden. Um dies alles eindeutig zu verifizieren, bräuchte es jedoch eine Folgestudie, mit welcher der Sachverhalt empirisch noch einmal überprüft wird und dabei theoretisch noch umfangreicher untermauert wird. Dabei wäre insbesondere eine grössere Stichprobe notwendig aus Personen mit guten Deutschkenntnissen. Auch wäre es sinnvoll, das MELBA-Merkmal Problemlösefähigkeit in die Untersuchung miteinzuziehen, da dieses gemäss Leighton (2012) eng mit der Arbeitsplanung zusammenhängt (Definition Problemlösefähigkeit in Anhang G). Würden die Zusammenhänge empirisch eindeutig belegt werden, könnte daraus gefolgert werden, dass die Turmbaufgabe die Fähigkeiten abzubilden vermag. Um sich darüber zu vergewissern, dass diese Zusammenhänge jedoch nicht aufgrund einer schlechten Trennschärfe der einzelnen Fähigkeitsmerkmale und aufgrund verfälschter Ratings der Beobachtenden zustande kämen, müssten noch einmal die Fähigkeitsdefinitionen, Leitfragen, Hilfestellungen nach MELBA und in diesem Zusammenhang das Beobachtungsrasster von *impuls* unter die Lupe genommen werden – ganz im Sinne der dargelegten Theorie in dieser Arbeit, welche sich mit der Systematisierung der Verhaltensbeobachtung befasst.

Im Abschnitt 2.3.2 wurde beschrieben, dass zwei Beobachtende gewisse Fähigkeiten oder Verhaltensweisen völlig unterschiedlich interpretieren können. Die Zusammenhänge zwischen Lernen/Merken, Arbeitsplanung, kritische Kontrolle und Misserfolgstoleranz könnten durchaus auf diese unterschiedlichen Interpretationen zurückzuführen sein. Zwar konnte in dieser Studie mit der Erhebung der Intraklassenkorrelation aufgezeigt werden, dass die Raterübereinstimmung grundsätzlich gut ist. Dies ist jedoch kein Garant dafür, dass beide Rater wirklich dieselben Verhaltensweisen beurteilt haben. Auch ist dadurch nicht gewährleistet, dass die Beobachtenden unterschiedliche Fähigkeiten beurteilt haben. Die teilweise grossen Zusammenhänge zwischen Lernen/Merken, Arbeitsplanung, kritischer Kontrolle und Misserfolgstoleranz könnten zum Beispiel auch damit begründet werden, dass in Wahrheit gar nicht

die einzelnen Fähigkeiten beurteilt wurden. Vielleicht wurde die Problemlösefähigkeit beurteilt, welche sich in Anlehnung an Leighton (2012) darin äussert, ob eine Aufgabe erfolgreich gelöst wurde oder nicht, und daraus gefolgert die einzelnen Fähigkeiten bewertet (ein weiterer Grund, weshalb die Problemlösefähigkeit in die Turmbaufaufgabe miteinbezogen werden sollte). Es ist daher wichtig, die Beobachtungseinheit so genau und differenziert wie möglich zu definieren. Insbesondere bei Ratingverfahren kann es durch ungenau definierte Beobachtungseinheiten zu Beobachtungsfehlern und damit zu Ergebnisverzerrungen kommen. Deshalb sollen, wie aus dem Theorieteil der Abschnitte 2.3.2 bis 2.3.4 herausgeht, Beobachtungsraster in Anlehnung an Grümer (1974) sowie Beobachtungseinheiten in Anlehnung an Margraf-Stiksrud (2005) gewisse Kriterien erfüllen. Betrachtet man das Beobachtungsraster von *impuls* (vgl. Abschnitt 2.3.1), fällt auf, dass pro Fähigkeiten jeweils nur die Leitfragen aufgeführt sind. Es fehlen Beschreibungen für die Skalenwerte. Auch fehlen genau definierte Beobachtungseinheiten, mit denen für die Fähigkeiten typische und konkrete Verhaltensweisen identifiziert werden können. Ob in der Beobachtungssituation für die Beobachtenden genug Zeit besteht, sich zu überlegen, in welcher konkreten Verhaltensweise sich die mit den Leitfragen verbalisierten Sachverhalte zeigen, ist fraglich. Dem Praxispartner *impuls* wird daher empfohlen, die Fragestellungen auf dem Beobachtungsraster durch konkrete Beobachtungseinheiten zu ersetzen resp. konkrete Verhaltensweisen aus den Fragestellungen und Hilfestellungen abzuleiten und verbalisiert auf dem Beobachtungsraster aufzuführen. Denn sowohl die Leitfragen als auch die Hilfestellungen, die Föhres et al. (2011) im Handbuch III beschreiben, lassen viel Interpretation zu. Dies kann dazu führen kann, dass die Rater unterschiedliche Verhaltensweisen beurteilen. Wichtig wäre, die Beobachtungseinheiten möglichst angepasst auf die Turmbaufaufgabe zu formulieren. Bei der Planungsfähigkeit könnte eine konkrete Verhaltensweise z.B. „Misst zuerst alle Holzstücke ab, bevor mit dem Turmbau begonnen wird“ lauten. Die Beobachtungseinheiten sollten möglichst trennscharf formuliert sein, sodass wenn eine bestimmte Verhaltensweise beobachtet wird, diese nur einer Fähigkeit zugeordnet werden kann. Dies wäre insbesondere auch dann wichtig, wenn das Konstrukt Problemlösefähigkeit gemäss obigem Vorschlag zur Beurteilung in die Aufgabe integriert würde. Denn betrachtet man die Definitionen von Föhres et al. (2011) zur Arbeitsplanung und zur Problemlösefähigkeit, so fällt es schwer, einen Unterschied festzustellen.

Zusammengefasst wird also empfohlen, die Zusammenhänge zwischen den vier Fähigkeiten noch einmal empirisch zu untersuchen und dabei die Problemlösefähigkeit miteinzubeziehen. Damit die Ergebnisse jedoch nicht durch fehlende Trennschärfe verzerrt werden, müsste das Beobachtungsraster, wie soeben vorgeschlagen, angepasst werden und die Leitfragen durch konkrete Beobachtungseinheiten ersetzt werden. Bortz und Döring (1995) sprechen in diesem Zusammenhang, wie im Abschnitt 3.5 erwähnt, von der Konkretisierung des Zielkonstrukts. Wie bereits ausgeführt, wäre es für die erneute Untersuchung der Zusammenhänge sinnvoll, eine Stichprobe aus Personen mit guten Deutschkenntnissen zu verwenden.

Nichtsdestotrotz wird die Turmbaufaufgabe von *impuls* auch zukünftig zur Feststellung von Fähigkeiten bei Personen mit schlechten Deutschkenntnissen eingesetzt. Diese machen einen grossen Teil der Klientel von *impuls* aus. Da es für diese Personen schwierig ist, die Anleitung zu verstehen, wird nahe gelegt, die Anleitung für diese Zielgruppe stärker zu verbildlichen. Ob es sinnvoll ist, die Merkfähigkeit bei Personen mit schlechten Deutschkenntnissen mittels Turmbaufaufgabe zu erheben, ist fraglich. Vielleicht wäre hier ein sprachunabhängiges Verfahren besser geeignet.

7. Literaturverzeichnis

- Alarcon, A., Eschleman, K.J. & Bowling, N.A. (2009). Relationships between personality variables and burnout: A meta-analysis. *Work & Stress. An International Journal of Work, Health & Organisations*, 28 (2), 244-263.
- Aronson, E., Wilson, T.D. & Akert, R.M. (2008). *Sozialpsychologie* (6. Aufl.). München: Pearson Studium.
- Asendorpf, J.B. (2009). *Persönlichkeitspsychologie*. Heidelberg: Springer Medizin Verlag.
- Atteslander, P. (2010). *Methoden der empirischen Sozialforschung* (13. Aufl.). Berlin: Erich Schmidt Verlag.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Berlin Heidelberg: Springer-Verlag.
- Bühl, A. (2008). *SPSS 16: Einführung in die moderne Datenanalyse* (11. Aufl.). München: Pearson Studium.
- Bühner, M. (2005). Gütekriterien von Verhaltensbeobachtungen. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 81-92). Lengerich: Pabst Science Publishers.
- Cohen, J. (1992). A power Primer. *Psychological Bulletin*, 112, 155-159.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin Heidelberg: Springer-Verlag.
- Europäischer Referenzrahmen (2016). *Gemeinsamer Europäischer Referenzrahmen für Sprachen (GER): Die Niveaustufen des GER*. Verfügbar unter <http://www.europaeischer-referenzrahmen.de/> [09.01.2016].
- Fassnacht, G. (1995). *Systematische Verhaltensbeobachtung: Eine Einführung in die Methodologie und Praxis*. München: Ernst Reinhardt.
- Föhres, F., Kleffmann, A., Sturtz, A. & Weinmann, S. (2011). *Melba Handbuch III: Fähigkeitsanalyse* (8. Aufl.). Lich: Miro GmbH.
- Flick, U. (2007). *Qualitative Sozialforschung: Eine Einführung*. Hamburg: Rowohlt Verlag.
- Fröhlich, W.D. (2008). *Wörterbuch Psychologie* (26. Aufl.). München: Deutsch Taschenbuch Verlag.
- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung*. Weinheim: Psychologie Verlags Union.

- Grüner, K.W. (1974). *Beobachtung: Techniken der Datensammlung*. Stuttgart: B.G. Teubner.
- Gutknecht S. P., Semmer, N.K. & Hubert, A. (2005). Prognostische Validität eines Assessment Centers für den Studien- und Berufserfolg von Berufsoffizieren der Schweizer Armee. In *Zeitschrift für Personalpsychologie*, 4 (4), 170-180. Göttingen: Hogrefe Verlag.
- Häder, M. (2006). *Empirische Sozialforschung: Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hossiep & Schulz (2013). Persönlichkeitsinventare, -strukturtest. In A. Frey, U. Lissmann & B. Schwarz (Hrsg.), *Handbuch Berufspädagogischer Diagnostik* (S. 196-215). Weinheim und Basel: Beltz Verlag.
- Hülshager, U.R. & Maier, G.W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf: Eine Bestandsaufnahme internationaler und nationaler Forschung. *Psychologische Rundschau*, 59 (2), 108-122.
- Impuls (2015). *Impuls – Verein für bessere Chancen im Beruf*. Verfügbar unter <http://www.impuls-sz.ch/Home.9.0.html> [09.01.2016]
- Kanning, U. P. & Schuler, H. (2014). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler & U.P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl.) (S. 215-251). Göttingen: Hogrefe Verlag.
- Kemper, C.J., Beierlein, D., Kovalev, A. & Rammstedt, B. (2012). *Eine Kurzskala zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens: Die Kurzskala Soziale Erwünschtheit-Gamma (KSE-G)*. Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.
- Kemper, J., Ziegler, M., Krumm, S., Heene, M. & Bühner, M. (2015). Testkonstruktion. In G. Stemmler & J. Margraf-Stiksrud (Hrsg.), *Lehrbuch Psychologische Diagnostik* (S. 157-221). Bern: Verlag Hans Huber.
- Kersting, M. (2010). *Akzept!-Fragebogen zur Messung der Akzeptanz diagnostischer Verfahren*. Verfügbar unter <http://kersting-internet.de/testentwicklungen/akzeptfragebogen/> [09.01.2016].
- Kleffmann, A., Weinmann, S., Föhres, F. & Müller, B. (1997). *Melba - Psychologisches Merkmalprofil zur Eingliederung Behinderter in Arbeit* (2. Aufl.). Siegen: Universität-Gesamthochschule Siegen.
- Leighton, J.P. (2004). Defining and Describing Reason. In J.P. Leighton & R.J. Sternberg (Hrsg.), *Nature of Reasoning* (S. 3-11). Cambridge: Cambridge University Press.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R: Manual* (2. Aufl.). Göttingen: Hogrefe Verlag.

- Liepmann, D., Beauducel, A., Brocke, B. & Nettelstroth, W. (2012). *Intelligenz-Struktur-Test Screening: Manual*. Göttingen: Hogrefe Verlag.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie Verlag Union.
- Lissmann, U. (2013). Kombinierte Verfahren. In A. Frey, U. Lissmann & B. Schwarz (Hrsg.), *Handbuch Berufspädagogischer Diagnostik* (S. 117-123). Weinheim und Basel: Beltz Verlag.
- Margraf-Stiksrud, J. (2005). Definition und Abgrenzung von Beobachtungseinheiten. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 163-208). Lengerich: Pabst Science Publishers.
- Margraf-Stiksrud, J. & Stemmler, G. (2015). Testverfahren. In G. Stemmler & J. Margraf-Stiksrud (Hrsg.), *Lehrbuch Psychologische Diagnostik* (S. 223-319). Bern: Verlag Hans Huber.
- Moosbrugger, H. & Rauch, W. (2005). Psychometrische Grundlagen der Eignungsbeurteilung. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 163-208). Lengerich: Pabst Science Publishers.
- Myers, D.G. (2008). *Psychologie* (2. Aufl.). Heidelberg: Springer Medizin Verlag.
- Nerdinger, F., Blickle, G. & Schaper, N. (2008). *Arbeits- und Organisationspsychologie*. Heidelberg: Springer Medizin Verlag.
- Obermann, Ch. (2013). *Assessment Center: Entwicklung, Durchführung, Trends* (5. Aufl.). Wiesbaden: Springer Gabler.
- Penley, J.A. & Tomaka, J. (2002). Associations among the Big Five, emotional responses, and coping with acute stress. *Personality and Individual Differences*, 32 (7), 1215-1228.
- Porst, R. (2008). *Fragebogen: Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rammstedt, B., Kemper, C.J., Klein, M.C., Beierlein, D. & Kovalev, A. (2012). *Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI)*. Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.
- Rauchfleisch, U. (2005). *Testpsychologie: Eine Einführung in die Psychodiagnostik* (4. Aufl.). Göttingen: Vandenhoeck & Ruprecht.

- Reimann, G. (2005). Einführung in die DIN 33430. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 163-208). Lengerich: Pabst Science Publishers.
- Rentzsch, K. & Schütz, A. (2009). *Psychologische Diagnostik: Grundlagen und Anwendungsperspektiven*. Stuttgart: Verlag W. Kohlhammer.
- Sarges, W. (2015). *Warum Assessment Center häufig zu kurz greifen*. Barnitz: Prof. Sarges & Partner – Institut für Management-Diagnostik. Verfügbar unter: <http://www.sarges-partner.de/management-diagnostik.php> [09.01.2016].
- Schmidt-Atzert, L. (2005). Registrieren und Dokumentation der Beobachtung. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 59-61). Lengerich: Pabst Science Publishers.
- Schmid-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (5. Aufl.). Berlin Heidelberg: Springer-Verlag.
- Schuler, H. (2007). Assessment Center als multiples Verfahren zur Potenzialanalyse: Einleitung und Überblick. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 3-33). Göttingen: Hogrefe Verlag.
- Schnell, R., Hill, B.H., Esser, E. (2011). *Methoden der empirischen Sozialforschung* (9. Aufl.). München: Oldenbourg Wissenschaftsverlag.
- Stemmler, G. (2005). Operationalisierung von Eignungsmerkmalen. In K. Westhoff, L.J. Hellfritsch, L.F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33420* (2. Aufl.) (S. 44-54). Lengerich: Pabst Science Publishers.
- Stemmler, G., Hagmann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7. Aufl.). Stuttgart: Verlag W. Kohlhammer.
- Stemmler, G. & Margraf-Stiksrud, J. (2015). Verhaltensbeobachtung. In G. Stemmler & J. Margraf-Stiksrud (Hrsg.), *Lehrbuch Psychologische Diagnostik* (S. 13-76). Bern: Verlag Hans Huber.
- Stockmann, R. (2007). *Handbuch zur Evaluation: Eine praktische Handlungsanleitung*. Münster: Waxmann Verlag.
- Swiss-Assessment (2015). *Swiss Assessment: Das Kompetenzzentrum für Assessment Center-Verfahren in der Schweiz*. Verfügbar unter: <http://www.swissassessment.ch/deutsch/home.html> [09.01.2016].
- Testzentrale.de (2015). *Intelligenz-Struktur-Test 2000 R*. Verfügbar unter: <http://www.testzentrale.de/programm/intelligenz-struktur-test-2000-r.html> [09.01.2016].

- Thunsdorff, C. & Schmitt, M. (2013). Berufseignung und Personenmerkmale. In A. Frey, U. Lissmann & B. Schwarz (Hrsg.), *Handbuch Berufspädagogischer Diagnostik* (S. 126-136). Weinheim und Basel: Beltz Verlag.
- Vedel, A. (2014). The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71, 66-76.
- Westmeyer, H. & Nell, V. (2005). Verhaltensbeobachtung. In F. Petermann & H. Reinecker (Hrsg.), *Handbuch der Klinischen Psychologie und Psychotherapie* (S. 200-208). Göttingen: Hogrefe Verlag.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystem und Ratingskalen*. Göttingen: Hogrefe Verlag.

Abbildungsverzeichnis

Abbildung 1: Anleitung Turmbau von <i>impuls</i>	S. 11
Abbildung 2: Beobachtungsraster Turmbau von <i>impuls</i>	S. 13

Tabellenverzeichnis

Tabelle 1: Schlüsselfähigkeiten nach MELBA (Kleffmann et al., 1997, S. 10)	S. 6
Tabelle 2: Schlüsselqualifikation Kritische Kontrolle (Föhres et al., 2011, S. 13) .S.	8
Tabelle 3: Schlüsselqualifikation Arbeitsplanung (Föhres et al., 2011, S. 2)	S. 8
Tabelle 4: Schlüsselqualifikation Misserfolgstoleranz (Föhres et al., 2011, S. 16)S.	9
Tabelle 5: Schlüsselqualifikation Lernen/Merken (Föhres et al., 2011, S. 14)	S. 10
Tabelle 6: Big-Five-Persönlichkeitsfaktoren und ihnen zugeordnete Bedeutungsinhalte (Eigene Darstellung, 2015 in Anlehnung an Hossiep & Mühlhaus, 2005, zitiert nach Hossiep & Schulz, 2013, S. 202)	S. 18
Tabelle 7: Eingesetzte Items des Akzept!-Fragebogens und deren Zweck	S. 44
Tabelle 8: Ablauf Turmbauaufgabe	S. 53
Tabelle 9: Intraklassen-Korrelationskoeffiziente Turmbauaufgabe	S. 59
Tabelle 10: Multitrait-Multimethod-Matrix mit Korrelation nach Spearman	S. 62
Tabelle 11: <i>Zusammenhänge Sprachen und Konstrukte</i>	S. 66
Tabelle 12: Ergebnisse der multiplen linearen Regressionsanalyse Turmbau.....	S. 67

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Master-Arbeit selbständig, ohne Mithilfe Dritter und nur mit den angegebenen Quellen, Hilfsmitteln und Hilfestellungen erstellt habe und dass Zitate kenntlich gemacht sind.

Aarau, 11. Januar 2016

Wolfgang Meier

Anhang B

Eingesetzter Fragebogen in Anlehnung an den Akzept!-Fragebogen von Kersting (2010)



Wolfgang Meier
Student in Angewandter Psychologie
wolfgang.meier@students.fhwn.ch

Nr.

Sie haben gerade die Turmbauaufgabe bearbeitet – vielen Dank! Uns ist es wichtig, Ihre Meinung zur Turmbauaufgabe zu erfahren. Deshalb bitten wir Sie, die nachstehenden Fragen zu beantworten. Bitte kreuzen Sie an, inwieweit Sie den einzelnen Aussagen zustimmen oder nicht zustimmen. Machen Sie dies bitte bei jeder Aussage durch Ankreuzen des entsprechenden Smileys kenntlich.

	stimmt voll und ganz	stimmt	Geht so	stimmt nicht	stimmt ganz und gar nicht
1. Die Turmbau-Aufgabe war klar und verständlich.	☺☺	☺	☺	☹	☹☹
2. Die Turmbauaufgabe war zu schwer für mich.	☹☹	☹	☺	☺	☺☺
3. Den Turm zu bauen, empfand ich als belastend.	☹☹	☹	☺	☺	☺☺
4. Die Atmosphäre während der Turmbauaufgabe war – im Rahmen des Möglichen – sehr positiv	☺☺	☺	☺	☹	☹☹

Bemerkung:

Ihre Antworten und Daten werden anonym und sensibel behandelt. Es werden keine Rückschlüsse auf Sie gemacht werden können.

Bei Fragen stehe ich Ihnen gerne zur Verfügung. Besten Dank für das Beantworten der Fragen.

Anhang C

Fragebogen zur Erhebung der Big-Five und des sozial erwünschten Antwortverhaltens



Wolfgang Meier
Student in Angewandter Psychologie
wolfgang.meier@students.fhwn.ch

		stimmt überhaupt nicht	stimmt eher nicht	weder noch	stimmt eher	stimmt voll und ganz
1	Ich bin eher zurückhaltend, reserviert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Manchmal helfe ich jemandem nur, wenn ich eine Gegenleistung erwarten kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Ich bin bequem, neige zu Faulheit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Im Streit bleibe ich stets sachlich und objektiv.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Ich habe nur wenige künstlerische Interessen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Ich habe schon mal Müll einfach in die Landschaft oder auf die Straße geworfen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Ich gehe aus mir heraus, bin gesellig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Ich neige dazu, andere zu kritisieren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Wenn ich mich mit jemandem unterhalte, höre ich ihm immer aufmerksam zu.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	Ich erledige Aufgaben gründlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Es ist schon mal vorgekommen, dass ich jemanden ausgenutzt habe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	Ich werde leicht nervös und unsicher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Auch wenn ich manchmal gestresst bin, behandle ich andere immer freundlich und zuvorkommend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Anhang D

Leitfaden Instruktion Turmbauaufgabe



Fachhochschule Nordwestschweiz
Hochschule für Angewandte Psychologie

Wolfgang Meier
Tel. 079 487 20 07
wolfgang.meier@students.fhnw.ch

Aarau, 25. Oktober 2015

Leitfaden Instruktion Turmbau

In Anlehnung an die Anleitung Turmbau

Die unten aufgeführte Instruktion soll ca. 5 Minuten dauern. Danach hat der Teilnehmende 10 Minuten Zeit, sich den Auftrag in aller Ruhe einzuprägen und Rückfragen zu stellen

Guten Tag...

- Dank fürs Mitmachen aussprechen / Ziel der Erhebung / „das Ganze ist anonym“ / dem Teilnehmenden die Anleitung aushändigen
- In 2-3 Minuten grob erläutern, um was es geht (Punkte der Anleitung kurz zusammenfassen):
 - o Ziel der Aufgabe ist es, dass Sie mit Hilfe dieser Anleitung einen Holzturm aus verschiedenen Holzstücken bauen (auf Stückliste zeigen).
 - o Der Turm soll vom Grundriss her 25 cm auf 25 cm sein. Der Turm soll 80.8 cm hoch werden (vgl. Skizze)
 - o Ihnen wird verschiedenes Material vor, mit dem Sie den Holzturm bauen können: Holzteile gemäss der Stückliste, eine Grundrissplatte und ein Massstab (Meterstab).
 - o Der Turm soll nach bestimmten Kriterien gebaut werden. Er soll...
 - nach innen und aussen keine Hohlräume haben
 - die Holzteile sollen nicht hervorstehen
 - Der Turm soll in sich stabil sein
 - Gemäss Anleitung...
- Ablauf erläutern:

Sie haben nun 10 Minuten Zeit, diese Anleitung zu studieren und sich einzuprägen. Sie können mir Fragen stellen, sollte etwas unklar sein. Während dem Turmbau haben Sie die Möglichkeit nicht mehr, Fragen zu stellen. Nach diesen 10 Minuten haben Sie 15 Minuten Pause. Nach der Pause werden ich Sie fragen, was Sie noch vom Auftrag noch wissen. Sie sollten sich also die Anleitung so gut wie möglich einprägen. Danach werden wir Ihnen die Anleitung wieder aushändigen, und Sie haben 30 Minuten Zeit, den Turm zu bauen. Dabei werden Sie von zwei Personen beobachtet, welche Sie dann nach den Kriterien (auf Punkt 1 in Anleitung verweisen) beurteilen, wie Sie den Turm bauen.
- Noch einmal darauf hinweisen, dass das ganze anonym ist
- Fragen, ob der Auftrag klar ist
- Auffordern, sich die Anleitung während 10 Minuten einzuprägen
- In die Pause schicken und mitteilen, um welche Zeit sich der Teilnehmende wo einfinden soll

Anhang E

Ergebnisse Items aus Akzept!-Fragebogen nach Kersting (2010)

Statistiken

		Atmosphäre während der Aufgabe	Verständlichkeit der Aufgabe	Schwierigkeitsgrad Aufgabe	Belastungsfreiheit der Aufgabe
N	Gültig	31	31	31	30
	Fehlend	11	11	11	12
Mittelwert		4.1935	4.7419	2.1290	2.1333
Standardabweichung		.94585	.51431	1.02443	1.19578

Verständlichkeit der Aufgabe

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	geht so	1	2.4	3.2	3.2
	trifft zu	6	14.3	19.4	22.6
	trifft voll und ganz zu	24	57.1	77.4	100.0
	Gesamtsumme	31	73.8	100.0	
Fehlend	999,00	11	26.2		
Gesamtsumme		42	100.0		

Anmerkung. Item-Nr. 1: Die Turmbauaufgabe war klar und verständlich. Skalenwerte von 1 = *trifft ganz und gar nicht zu* bis 5 = *trifft voll und ganz zu*.

Schwierigkeitsgrad der Aufgabe

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	trifft ganz und gar nicht zu	10	23.8	32.3	32.3
	trifft nicht zu	10	23.8	32.3	64.5
	geht so	9	21.4	29.0	93.5
	trifft zu	1	2.4	3.2	96.8
	trifft voll und ganz zu	1	2.4	3.2	100.0
	Gesamtsumme	31	73.8	100.0	
Fehlend	999,00	11	26.2		
Gesamtsumme		42	100.0		

Anmerkung. Item-Nr. 2: Die Turmbauaufgabe war zu schwer für mich. Skalenwerte von 1 = *trifft ganz und gar nicht zu* bis 5 = *trifft voll und ganz zu*.

Belastungsfreiheit der Aufgabe

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	trifft ganz und gar nicht zu	12	28.6	40.0	40.0
	trifft nicht zu	7	16.7	23.3	63.3
	geht so	8	19.0	26.7	90.0
	trifft zu	1	2.4	3.3	93.3
	trifft voll und ganz zu	2	4.8	6.7	100.0
	Gesamtsumme	30	71.4	100.0	
Fehlend	999,00	12	28.6		
Gesamtsumme		42	100.0		

Anmerkung. Item-Nr. 3: Den Turm zu bauen, empfand ich als belastend. Skalenwerte von 1 = *trifft ganz und gar nicht zu* bis 5 = *trifft voll und ganz zu*.

Atmosphäre während der Aufgabe

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	trifft nicht zu	3	7.1	9.7	9.7
	geht so	2	4.8	6.5	16.1
	trifft zu	12	28.6	38.7	54.8
	trifft voll und ganz zu	14	33.3	45.2	100.0
	Gesamtsumme	31	73.8	100.0	
Fehlend	999,00	11	26.2		
Gesamtsumme		42	100.0		

Anmerkung. Item-Nr. 4: Die Atmosphäre während der Turmbaufaufgabe war – im Rahmen des Möglichen – sehr positiv. Skalenwerte von 1 = *trifft ganz und gar nicht zu* bis 5 = *trifft voll und ganz zu*.

Anhang F

Korrelationsmatrix mit Korrelationskoeffiziente nach Spearman

	Arbeitsplanung	Kritische Kontrolle	Misserfolgstoleranz	Lernen/Merken	Merkfähigkeit gesamt	Schlussfolgerndes Denken	Neurotizismus	Extraversion	Offenheit	Verträglichkeit	Gewissenhaftigkeit
Arbeitsplanung	r Sig. (2-seitig) N	1.000 .490** 42	.555** .000 42	.531** .001 34	.630** .002 22	.446* .043 21	.336 .126 22	.020 .930 22	.274 .217 22	-.221 .322 22	-.032 .888 22
Kritische Kontrolle	r Sig. (2-seitig) N	.490** .001 42	1.000 .193 42	.252 .150 34	.582** .005 22	.473* .030 21	.190 .398 22	.132 .558 22	.402 .064 22	-.628** .002 22	-.116 .606 22
Misserfolgstoleranz	r Sig. (2-seitig) N	.555** .000 42	1.000 .193 42	.363* .035 34	.370 .090 22	.394 .077 21	.201 .369 22	.150 .505 22	.328 .137 22	-.076 .737 22	.052 .817 22
Lernen/Merken	r Sig. (2-seitig) N	.531** .001 34	.252 .150 34	1.000 .363* 34	.392 .071 22	.430 .052 21	.304 .169 22	.189 .399 22	.258 .246 22	-.140 .535 22	-.367 .093 22
Merkfähigkeit gesamt	r Sig. (2-seitig) N	.630** .002 22	.582** .005 22	.392 .071 22	1.000 .370 22	.459* .042 20	.170 .451 22	.446* .037 22	.484* .022 22	-.188 .402 22	.095 .673 22
Schlussfolgerndes Denken	r Sig. (2-seitig) N	.446* .043 21	.473* .030 21	.430 .052 21	.392 .071 22	1.000 .173 20	.173 .465 20	-.030 .901 20	.430 .058 20	-.132 .578 20	-.128 .590 20
Neurotizismus	r Sig. (2-seitig) N	.336 .126 22	.190 .398 22	.304 .169 22	.201 .369 22	.173 .465 20	1.000 .373 22	-.200 .373 22	-.239 .283 22	.115 .609 22	-.281 .205 22
Extraversion	r Sig. (2-seitig) N	.020 .930 22	.132 .558 22	.189 .399 22	.150 .505 22	-.030 .901 20	1.000 .373 22	1.000 .014 22	.518* .014 22	-.092 .685 22	.479* .024 22
Offenheit	r Sig. (2-seitig) N	.274 .217 22	.402 .064 22	.252 .150 34	.370 .090 22	.394 .077 21	.201 .369 22	.150 .505 22	1.000 .014 22	-.284 .201 22	.212 .343 22
Verträglichkeit	r Sig. (2-seitig) N	-.221 .322 22	-.628** .002 22	-.076 .737 22	-.188 .402 22	-.132 .578 20	.115 .609 22	-.092 .685 22	1.000 .201 22	1.000 .353 22	.353 .107 22
Gewissenhaftigkeit	r Sig. (2-seitig) N	-.032 .888 22	-.116 .606 22	.052 .817 22	.095 .673 22	-.128 .590 20	-.281 .205 22	.479* .024 22	.212 .343 22	1.000 .107 22	1.000 .353 22

*. Korrelation ist bei Niveau 0,05 signifikant (zweiseitig).

** . Korrelation ist bei Niveau 0,01 signifikant (zweiseitig).

Anhang G

Schlüsselqualifikation Problemlösen (Föhres et al., 2011, S. 18)

Problemlösen

Fähigkeitsdefinition:

Problemlösen ist die Fähigkeit, neuartige Fragestellungen und Situationen in ihrer Bedeutung rechtzeitig erkennen und innerhalb der angemessenen Frist durch den Einsatz verschiedener Strategien (Kreativität, Analyse, Planung etc.) sach- und personengerecht lösen zu können.

Anmerkung: Derartige Fragestellungen können bestehen in neuartigen Aufgaben, interpersonalen Konflikten, dem Ausfall von Mitteln, die bisher zur Erfüllung des Arbeitsauftrages genutzt wurden usw.

Fragen zur Fähigkeitsanalyse:

- Kann die Person nicht vorhersehbare tätigkeitsbezogene Probleme selbstständig lösen oder ist sie immer auf Unterstützung angewiesen?
 - Wird sie aktiv um auftretende Probleme anzugehen?
 - Ist sie in der Lage, geeignete Mittel für die Lösung eines Problems zu benennen?
 - Kann sie vorhandene Mittel kreativ nutzen um neuartige Situationen zu meistern?
-

Hilfestellung zur Beurteilung der Merkmalsausprägung:

1	<ul style="list-style-type: none">• kann sehr deutliche Veränderungen im Arbeitsablauf, die eine Abweichung von der Routine erfordern, erkennen• reagiert mit Irritation, Aggression oder Angst auf ein erkanntes Problem• kann vorhandene Mittel unter Anleitung, zur Lösung einer neuartigen, einfachen Arbeitsaufgabe einsetzen
2	<ul style="list-style-type: none">• kann Veränderungen im Arbeitsablauf, die eine Abweichung von der Routine erfordern, erkennen• kann in unsystematischer Form durch Ausprobieren, vorhandene Mittel zur Lösung einer überschaubaren, neuartigen Situation einsetzen
3	<ul style="list-style-type: none">• kann eine neuartige Fragestellung oder Situation in ihrer Bedeutung erkennen und innerhalb einer angemessenen Frist sach- und personengerecht bearbeiten• überlegt selbstständig alternative Möglichkeiten für die Lösung, wenn mit vorhandenen Mitteln und Gegebenheiten ein Problem nicht gelöst werden kann
4	<ul style="list-style-type: none">• kann bei komplexen neuartigen Fragestellungen und Aufgaben personen- und sachgerecht reagieren• erkennt durch systematisches Analysieren verschiedener Lösungsmöglichkeiten schnell, ob die vorhandenen Mittel zur Lösung ausreichend sind• entwickelt schnell und kreativ effektive Lösungen
5	<ul style="list-style-type: none">• kann auch unter sehr schlechten Bedingungen, wie z.B. grossem Zeitdruck, kreative und effektive Lösungen für hochkomplexe Probleme erarbeiten
