

Grundlagen der Testtheorie

WS 2020/21

11. Validität

01.02.2021

Prof. Dr. Eunike Wetzel

Semesterplan

Sitzung	Termin	Thema
1	02.11.	Grundlagen & Gütekriterien
2	09.11.	Schritte der Testkonstruktion: Übersicht Konstruktdefinition & Itemgenerierung
3	16.11.	Erstellung eines Testentwurfs
4	23.11.	Klassische Testtheorie
5	07.12.	Item Response Theorie
6	14.12.	Exploratorische Faktorenanalyse
7	04.01.	Itemanalyse 1
8	11.01.	Itemanalyse 2, Itemselektion & Testrevision
9	18.01.	Objektivität
10	25.01.	Reliabilität
11	01.02.	Validität
12	08.02.	Normierung, Standards für psychologisches Testen

Validität

- Die Validität gibt an, ob der Test auch wirklich das misst, was er zu messen beansprucht.
- Validität bezieht sich auf die Angemessenheit der Schlussfolgerungen, die aus den Testergebnissen gezogen werden
- Validität ist Reliabilität und Objektivität übergeordnet

3 Strategien zur Testvalidierung

1. Inhaltsvalidität
2. Konstruktvalidität
3. Kriteriumsvalidität

1. Inhaltsvalidität

Zentrale Fragen

- Wie repräsentativ sind die Inhalte des Tests für das zu erfassende Konstrukt?
- Sind alle relevanten Inhalte vorhanden?
- Stehen die Inhalte in einem angemessenen Verhältnis zueinander?
- Sind keine Inhalte enthalten, die sich auf etwas Irrelevantes beziehen?

1. Inhaltsvalidität

- Bestimmung aufgrund logischer und fachlicher Überlegungen (meist unter Einbezug von Expertenurteilen)
- Da die Inhaltsvalidität theoretisch-argumentativ bestimmt wird und nicht empirisch, wird sie häufig vernachlässigt
- Die Inhaltsvalidität ist eng verknüpft mit den Schritten der Konstruktdefinition und Itemgenerierung
- Mängel in der Inhaltsvalidität sind kaum zu kompensieren

2. Konstruktvalidität

- Ziel der Konstruktvalidierung ist es zu überprüfen, ob die Testergebnisse im Sinne des interessierenden Konstrukts interpretiert werden können
- Zentrale Frage: Wird das interessierende Konstrukt gemessen?
- Vorgehen:
 1. Analysen auf Itemebene
 - Analyse der Zusammenhangsstruktur der Items (Dimensionalität) mit Methoden der exploratorischen und konfirmatorischen Faktorenanalyse und der Item Response Theorie
→ **faktorielle Validität**
 - Analyse der Antwortprozesse (lautes Denken, kognitive Interviews)
 2. Analysen auf Testebene

Empirische Überprüfung theoretischer Annahmen über Zusammenhänge latenter Konstrukte

→ **nomologisches Netz**

2. Konstruktvalidität: Itemebene

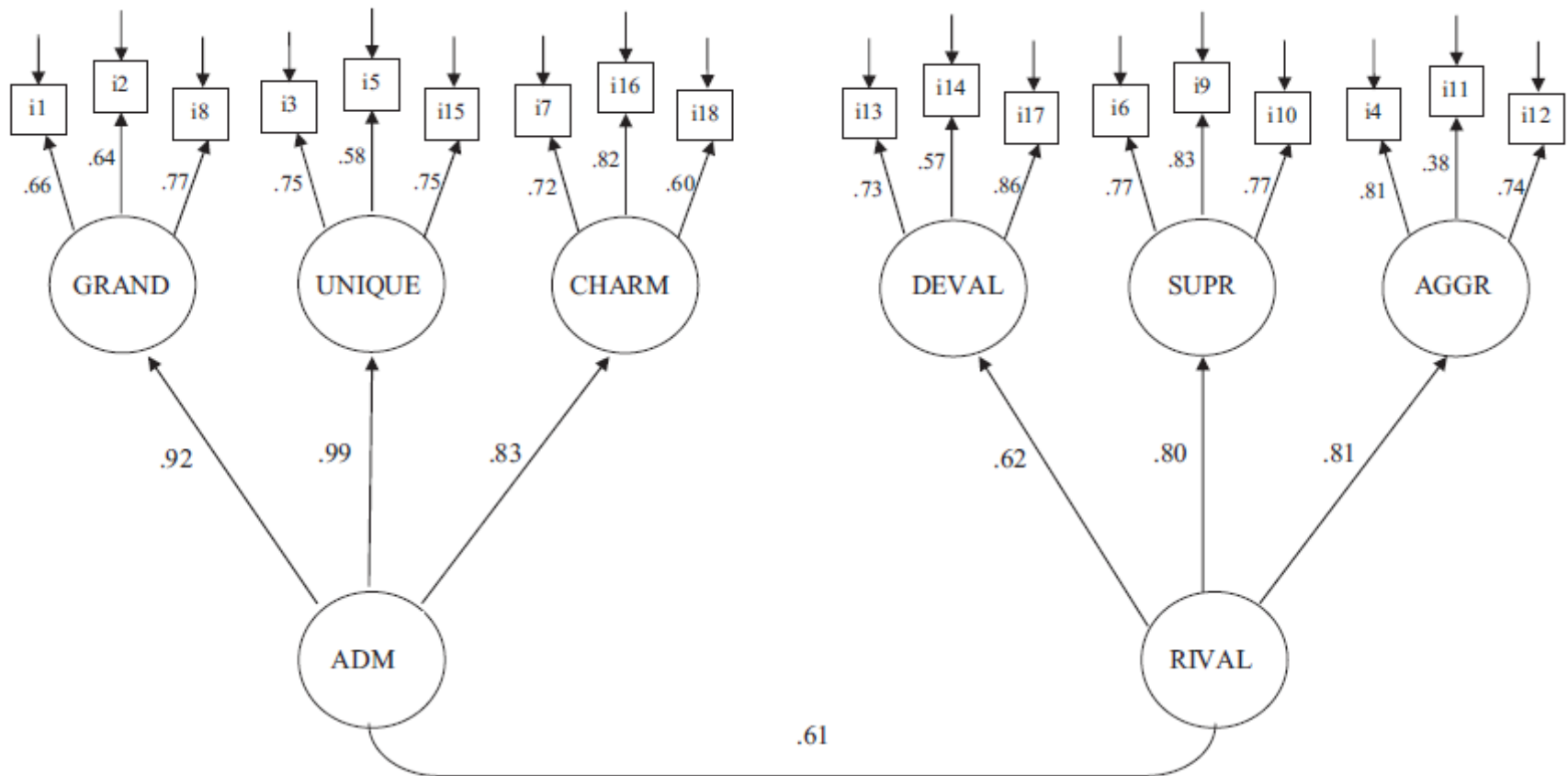


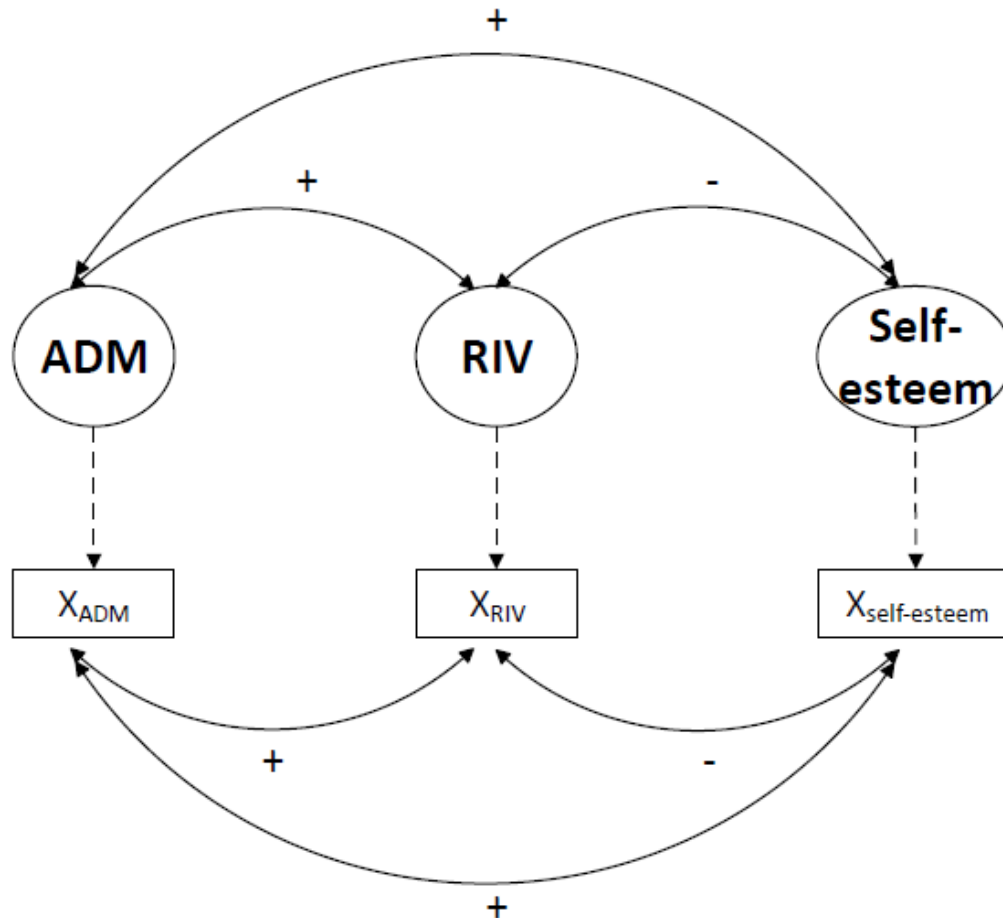
Figure 2. Confirmatory factor analysis model of the Narcissistic Admiration and Rivalry Questionnaire. $N = 953$. See Table 2 for item wordings. All loadings are standardized. ADM = narcissistic admiration; RIVAL = narcissistic rivalry; GRAND = grandiosity; UNIQUE = striving for uniqueness; CHARM = charmingness; DEVAL = devaluation; SUPR = striving for supremacy; AGGR = aggressiveness.

2. Konstruktvalidität: Testebene

- Beschreibung von latenten Konstrukten und den Interdependenzen (Zusammenhängen) zwischen den Konstrukten
- Formulierung von Annahmen, welche latente Konstrukte in Verbindung mit welchen beobachtbaren Testwerten stehen (Korrespondenzregeln)
- Aus den theoretischen Zusammenhängen lassen sich für die beobachtbaren Testwerte konkrete Zusammenhänge vorhersagen
- Die Gesamtheit der Interdependenzen, der Korrespondenzregeln und der vorhergesagten Zusammenhänge bildet das *nomologische Netz*
- Ziel der Konstruktvalidierung auf Testebene: Schrittweise Überprüfung des nomologischen Netzes
- Stimmen die Vorhersagen und empirischen Beobachtungen überein, spricht das dafür, dass die Testwerte als individuelle Ausprägungen auf dem latenten Konstrukt interpretiert werden können

2. Konstruktvalidität: Testebene

Bsp.: Aufgrund des theoretischen Modells postulierte Zusammenhänge zwischen Admiration, Rivalry und Self-esteem auf der latenten Konstruktebene und der manifesten Ebene der beobachteten Testwerte



2. Konstruktvalidität: Testebene

Table 5

Relations to the Big Five and Self-Esteem (Study 4)

Trait correlate	NARQ	
	r/β	
	ADM	RIV
Neuroticism	−.16/−.25	.19/.28
Extraversion	.31/.39	−.11/−.24
Openness	.25/.31	−.08/−.18
Agreeableness	−.04/.11	−.42/−.46
Conscientiousness	.08/.16	−.19/−.25
Self-esteem	.33/.49	−.23/−.42

2. Konstruktvalidität: Testebene

Methoden

- Experimentelle Ansätze
 - Konstrukt als AV
 - Konstrukt als UV
- Korrelative Ansätze
 - Konvergente Validität
 - Diskriminante Validität

2. Konstruktvalidität: Testebene

Experimentelle Ansätze: Konstrukt als AV

- Theoretisch begründete Annahme, dass bestimmte Faktoren einen Effekt auf das interessierende Konstrukt haben
- Experimentelle Variation der Faktoren sollte einen Effekt auf die manifesten Testwerte des Konstrukts haben
- Sind die Effekte hypothesenkonform spricht das für die Konstruktvalidität des Tests

2. Konstruktvalidität: Testebene

Experimentelle Ansätze: Konstrukt als AV

State-Trait-Angstinventar (STAI; Laux, Glanzmann, Schaffner & Spielberger, 1981)

Trait-Angst							State-Angst						
Männliche Studenten			Weibliche Studenten				Männliche Studenten			Weibliche Studenten			
M	s	Alpha	M	s	Alpha		M	s	Alpha	M	s	Alpha	

Neutrale Situation

Neutrale Situation
(nach 1 Stunde)

Erste Klausur
(nach 24 Tagen)

Zweite Klausur
(nach 59 Tagen)

Neutrale Situation
(nach 63 Tagen)

Dritte Klausur
(nach 73 Tagen)

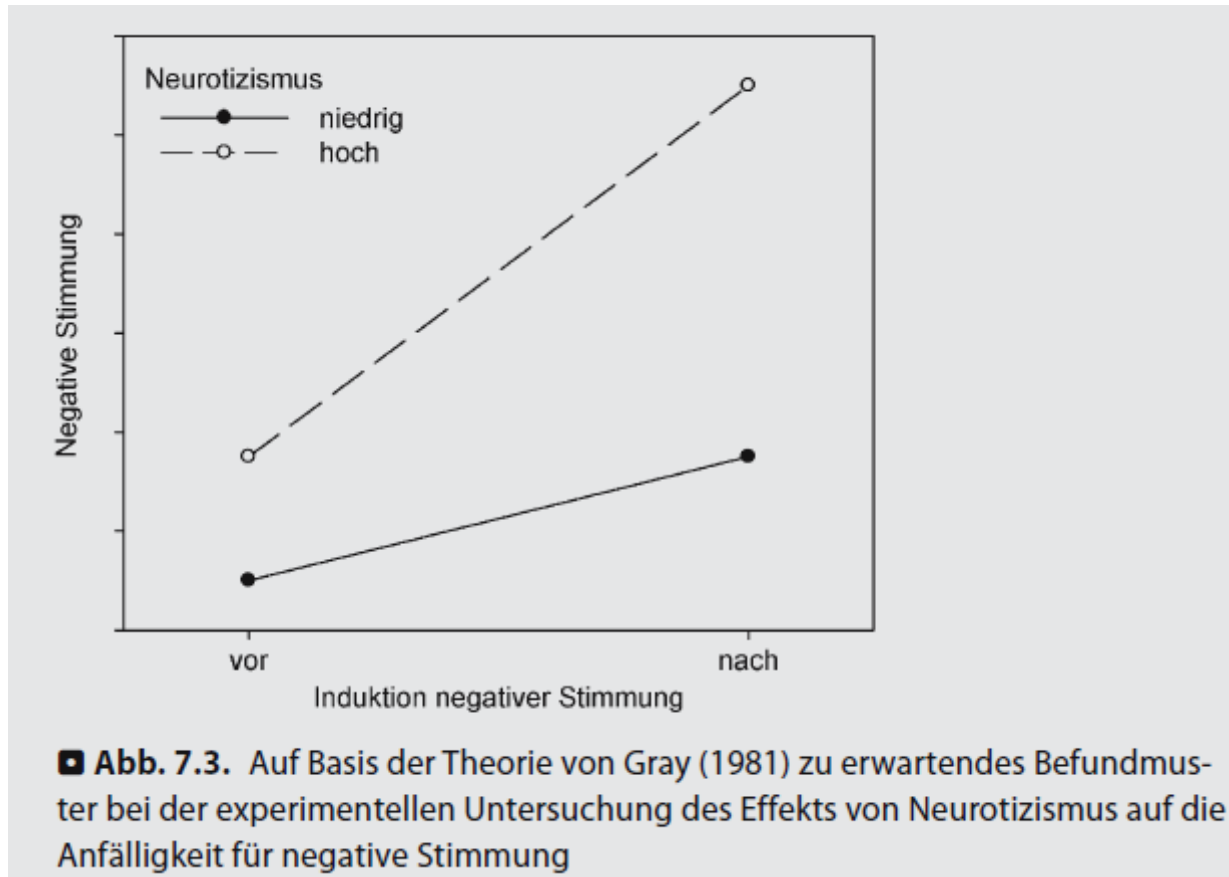
2. Konstruktvalidität: Testebene

Experimentelle Ansätze: Konstrukt als UV

- Theoretisch begründete Annahme, dass das interessierende Konstrukt einen Effekt auf ein anderes Konstrukt (AV) hat
- Probanden mit unterschiedlichen manifesten Testwerten sollten innerhalb eines experimentellen Untersuchungsdesigns unterschiedliche Werte auf der AV haben
- Sind die Effekte hypothesenkonform spricht das für die Konstruktvalidität des Tests

2. Konstruktvalidität: Testebene

Experimentelle Ansätze: Konstrukt als UV



2. Konstruktvalidität: Testebene

Korrelative Ansätze

- Theoretisch begründete Annahme, dass das interessierende Konstrukt mit einem anderen Konstrukt, einem Verhaltensmaß oder einer anderen Personenvariable (z. B. Alter, Geschlecht) zusammenhängt
- Die Theorie muss eine Hypothese zur Richtung und Höhe der Korrelation erlauben
- Stimmen die empirisch ermittelten Korrelationen mit den postulierten Zusammenhängen überein, spricht das für die Konstruktvalidität des Tests

2. Konstruktvalidität: Testebene

Korrelative Ansätze

- **Konvergente Validität:** Korrelation mit konstruktnahen Variablen → Erwartung hoher Zusammenhänge
- **Diskriminante Validität:** Korrelation mit konstruktfernden Variablen → Erwartung niedriger oder nicht vorhandener Zusammenhänge
- Der **Multitrait-Multimethod-Ansatz** erlaubt die gemeinsame Prüfung der konvergenten und diskriminanten Validität
 - Multitrait: Korrelationen zwischen unterschiedlichen Traits
 - Multimethod: Korrelationen zwischen unterschiedlichen Methoden

2. Konstruktvalidität: Testebene

Korrelative Ansätze: Konvergente und diskriminante Validität

Table 5

Relations to the Big Five and Self-Esteem (Study 4)

Trait correlate	NARQ	
	r/β	
	ADM	RIV
Neuroticism	−.16/−.25	.19/.28
Extraversion	.31/.39	−.11/−.24
Openness	.25/.31	−.08/−.18
Agreeableness	−.04/.11	−.42/−.46
Conscientiousness	.08/.16	−.19/−.25
Self-esteem	.33/.49	−.23/−.42

2. Konstruktvalidität: Testebene

Trait correlate	NARQ	
	<i>r/β</i>	
	ADM	RIV
Pathological narcissism		
Overall	.39/.19	.60/.52
Grandiosity	.59/.46	.51/.33
Vulnerability	.25/.03	.57/.55
Entitlement	.59/.43	.57/.40
Grandiosity	.73/.72	.31/.03
Impulsivity	.04/−.08	.26/.29
Anger	.16/−.09	.58/.62
Machiavellianism	.17/−.10	.64/.67
Psychopathy	.33/.21	.39/.31
Enhancement		
General	.46/.56	−.03/−.25
Agentic	.32/.43	−.11/−.27
Communal	.05/.27	−.46/−.57

2. Konstruktvalidität: Testebene

Beispiel

Multitrait-Multimethod Matrix zur Überprüfung der Konstruktvalidität des deutschen Beck Depression Inventory

		S (Selbstbeurteilung)			F (Fremdbeurteilung)		
		D	A	N	D	A	N
S	D	.86					
	A	.46	.81				
	N	.47	.42	.81			
F	D	.34	.49	.17	.75		
	A	.47	.55	.29	.46	.92	
	N	.39	.34	.27	.64	.43	.73

Reliabilität konvergente Validität diskriminante Validität

Anmerkungen:

D (Depression): BDI (S) bzw. Hamilton-Depressionsskala (F)

A (Angst): ZUNG-Self Rating (S) bzw. COVI-Angstskala (F)

N (Neurotizismus): EPI-N (S) bzw. Ratingskala von Amelang (F)

2. Konstruktvalidität: Testebene

- Multitrait-Multimethod Analyse des NEO-PI-R und der Bipolaren Adjektiv-Rating-Skalen (BARS), jeweils erfasst mit Selbstbeurteilungen (S) und Fremdbeurteilungen (F)
- Fett: Monotrait-heteromethod Korrelationen

	NEO-PI-R-Form S				
	N	E	O	A	C
NEO-PI-R-S					
N	(.93)				
E	-.24**	(.88)			
O	.10*	.43**	(.89)		
A	-.05	-.04	.09*	(.82)	
C	-.41**	-.02	-.22**	.08*	(.90)
NEO-PI-R-F					
N	.62**	-.18*	.13	.03	-.27**
E	-.10	.63**	.31**	-.08	.00
O	.26**	.21**	.56**	-.05	-.11
A	.09	-.04	.07	.55**	-.04
C	-.29**	-.07	-.22**	.09	.64**
BARS179-F					
N	.57**	-.14	.11	.05	-.03
E	.06	.47**	.15	-.11	-.08
O	.15	.13	.38**	-.20	-.15
A	.05	-.04	.01	.41**	-.11
C	-.07	-.20	-.36**	.19	.52**
BARS179-S					
N	.81**	-.11*	.14**	.06	-.19**
E	-.07	.76**	.20**	.03	-.05
O	-.05	.22**	.60**	-.13**	-.02
A	-.09	-.13**	.02	.69**	-.01
C	-.12**	-.08	-.31**	.10*	.82**

2. Konstruktvalidität: Testebene

Korrelative Ansätze: Konvergenz von Selbst- und Fremdbeurteilungen

NARQ measures	<i>M</i>	<i>SD</i>	d_{sex}	α	r_{tt}	r_{so}
1. Narcissism	2.46	0.73	.42	.88	.79	.44
2. Admiration	2.77	0.94	.28	.87	.79	.51
3. Grandiosity	2.43	1.04	.45	.73	.76	.43
4. Uniqueness	3.02	1.14	.38	.73	.72	.31
5. Charmingness	2.87	1.10	.11	.76	.69	.45
6. Rivalry	2.14	0.78	.24	.83	.76	.27
7. Devaluation	1.63	0.82	.60	.75	.62	.31
8. Supremacy	2.47	1.25	.39	.83	.79	.30
9. Aggressiveness	2.32	0.88	.10	.66	.64	.11

3. Kriteriumsvalidität

- Kriteriumsvalidität ist gegeben, wenn vom Testwert auf ein für diagnostische Entscheidungen praktisch relevantes Kriterium geschlossen werden kann
- Auswahl der Kriterien:
 - Hängt vom Anwendungszweck des Tests ab: Kriterien müssen unmittelbar relevant für die zu treffende diagnostische Entscheidung sein
 - Kriterien müssen hinreichend reliabel sein
- Vorgehen: Untersuchung des Zusammenhangs zwischen Testwert und externen Kriterien

3. Kriteriumsvalidität

- Arten:
 1. Vorhersagevalidität
 2. Übereinstimmungsvalidität
 3. Retrospektive Validität
 4. Inkrementelle Validität
- Kriteriumsvalidität ist u.a. abhängig von:
 - Inhaltsvalidität des Prädiktors und Kriteriums
 - Reliabilität des Prädiktors und Kriteriums
 - Kriteriumskontamination und Kriteriumsdefizienz

3. Kriteriumsvalidität

Vorhersagevalidität & inkrementelle Validität

Table 1

Predictive Validity for Overall Job Performance of General Mental Ability (GMA) Scores Combined With a Second Predictor Using (Standardized) Multiple Regression

Personnel measures	Validity (r)	Multiple R	Gain in validity from adding supplement	% increase in validity	Standardized regression weights	
					GMA	Supplement
GMA tests ^a	.51					
Work sample tests ^b	.54	.63	.12	24%	.36	.41
Integrity tests ^c	.41	.65	.14	27%	.51	.41
Conscientiousness tests ^d	.31	.60	.09	18%	.51	.31
Employment interviews (structured) ^e	.51	.63	.12	24%	.39	.39
Employment interviews (unstructured) ^f	.38	.55	.04	8%	.43	.22
Job knowledge tests ^g	.48	.58	.07	14%	.36	.31
Job tryout procedure ^h	.44	.58	.07	14%	.40	.20
Peer ratings ⁱ	.49	.58	.07	14%	.35	.31
T & E behavioral consistency method ^j	.45	.58	.07	14%	.39	.31
Reference checks ^k	.26	.57	.06	12%	.51	.26
Job experience (years) ^l	.18	.54	.03	6%	.51	.18
Biographical data measures ^m	.35	.52	.01	2%	.45	.13
Assessment centers ⁿ	.37	.53	.02	4%	.43	.15
T & E point method ^o	.11	.52	.01	2%	.39	.29
Years of education ^p	.10	.52	.01	2%	.51	.10
Interests ^q	.10	.52	.01	2%	.51	.10
Graphology ^r	.02	.51	.00	0%	.51	.02
Age ^s	-.01	.51	.00	0%	.51	-.01

Literatur zu dieser Sitzung

Moosbrugger & Kelava (2012). Kapitel 7.