

ANHANG

CITO-Sprachstanderhebungstest

Die wesentlichen messtheoretischen Standards, denen gute Softwareprodukte genügen müssen, werden als Hauptgütekriterien bezeichnet. Von entscheidender Bedeutung ist die Validität (Gültigkeit), das heißt, dass die Schlussfolgerungen aus den Testwerten empirisch abgesichert werden.

Ein weiteres Merkmal ist die Reliabilität (Zuverlässigkeit), das Ausmaß, wo Testwerte ohne Messfehler sind. Diesen Bereich kann man am empirisch ermittelten Reliabilitätskoeffizienten ablesen.

Bei einer an 448 Vorschulkindern von Duisburg durchgeführten Vorstudie, wo auch zusätzlich noch zwei nicht-digitale Testkomponenten (Satznachahmung und aktiver Wortschatz) erfasst wurden, konnten bezüglich der Reliabilität, Validität und Konsistenz des CITO-Sprachtests folgende Schlussfolgerungen aus der Datenanalyse gezogen werden:

- Alle Testaufgaben weisen ein hohes Maß an interner Konsistenz der einzelnen Subtests in deutscher (d) und türkischer (t) Sprache auf, wie die Berechnungen des Koeffizienten Alpha und die durchschnittliche Item-Rest-Korrelation (RIR) in folgender Tabelle zeigen.

Subtest	Gruppe	Kinder	Items	P-Wert	Alpha	Durchschn. RIR
		(N)	(N)			
PW-d	D	154	67	82.70	.84	.29
PW-d	T	163	67	63.20	.83	.26
PW-t	T	148	61	68.20	.69	.19
KB-d	D	154	65	83.80	.84	.28
KB-d	T	162	65	70.00	.80	.23
KB-t	T	145	65	68.80	.82	.25
PB-d	D	151	30	80.90	.90	.48
PB-d	T	160	30	64.80	.87	.40
PB-t	T	143	30	67.80	.87	.40
TV-d	D	149	20	77.40	.81	.38
TV-d	T	154	20	57.70	.66	.25
TV-t	T	140	20	64.10	.71	.28

Abbildung 1: Interne Konsistenz

- Bezüglich Validität konnte gezeigt werden, dass die ausgewählten Itemzusammensetzungen eindimensional und populationsunabhängig sind und sich sowohl für deutsche als auch für zweisprachige Kinder eignen.

Bei der weiterführenden Untersuchung des Sprachtests an 5330 Duisburger Vorschulkindern, wo beide nicht-digitalen Testkomponenten ausschieden, konnte das Ergebnis zeigen, dass „CITO“ als digitales Diagnoseinstrument über ausreichende Zuverlässigkeit und Validität verfügt. Die Reliabilitätskoeffizienten in folgender Tabelle zeigen, dass die Messgenauigkeit der deutschen und türkischen Version gut bis sehr gut ausfällt. Deshalb wird der Test in allen Grundschulen Duisburgs als Schuleingangsüberprüfung durchgeführt und auch in Bremerhaven und Düsseldorf flächendeckend eingesetzt.

Aufgabe	Aufgaben insg.	N	Durchschnitt	P-Wert	alpha	Sd
PW	(60)	4660	42.47	70.80	.95	12.60
KB	(65)	4527	46.70	71.80	.91	10.28
PB	(30)	4428	19.83	66.10	.90	7.03
TV	(20)	4355	12.80	64.00	.86	4.49

Abbildung 2: Zuverlässigkeit der deutschen Testversion

Aufgabe	Aufgaben insg.	N	Durchschnitt	P-Wert	alpha	Sd
PW	(60)	967	39.66	66.10	.85	7.52
KB	(65)	958	40.81	62.80	.84	8.55
PB	(30)	946	17.38	57.90	.84	6.13
TV	(20)	936	10.94	54.7	.69	3.62

Abbildung 3: Zuverlässigkeit der türkischen Testversion

Ein weiteres Kriterium umfasst die Aufgabenkonstruktion, -analyse, und die Normierung. Normierungen sind nach Konak et al. (2005) das Produkt von Prozessen der Entscheidungsfindung und wurden bei „CITO“ auf Grund von Literatur zu den Testbereichen, Vorerfahrungen aus dem seit 15 Jahren in den Niederlanden verwendeten Test „Zweisprachigkeit“, Analyse von Ergebnissen aus dem Test an 5330 Kindern und Expertenmeinungen gebildet. Dabei wurde von der Frage ausgegangen, wie sprachfähig ein ca. sechsjähriges Kind sein sollte.

Normierungsdaten geben Auskunft über die Anzahl von richtigen Antworten, um in einer bestimmten Kategorie eingeordnet werden zu können.

In den folgenden Abbildungen (siehe sind die Normierungen zur deutschen und türkischen Testversion aller Testbereiche ersichtlich.

	Förderbedürftig	Befriedigend	Gut
Passiver Wortschatz 60	0-33	34-40	41-60
Kognitive Begriffe 65	0-39	40-45	46-65
Phonologisches Bewusstsein 30	0-13	14-19	20-30
Textverständnis 20	0-8	9-12	13-20

Abbildung 4: Normierungstabelle der deutschen Testversion

	Förderbedürftig	Befriedigend	Gut
Passiver Wortschatz 60	0-36	37-40	41-60
Kognitive Begriffe 65	0-35	36-42	43-65
Phonologisches Bewusstsein 30	0-12	13-19	20-30
Textverständnis 20	0-9	10-12	13-20

Abbildung 5: Normierungstabelle der türkischen Testversion

Hinsichtlich der Nutzung der Auswertungstabellen muss festgestellt werden, dass immer auch Beobachtung von Pädagogen und Pädagoginnen notwendig ist, um allenfalls Änderungen der individuellen Förderung durchführen zu können.

Multivariate Tests ^b						
Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz
Zeit	Pillai-Spur	,941	3,166E2	1,000	20,000	,000
	Wilks-Lambda	,059	3,166E2	1,000	20,000	,000
	Hotelling-Spur	15,828	3,166E2	1,000	20,000	,000
	Größte charakteristische Wurzel nach Roy	15,828	3,166E2	1,000	20,000	,000
Zeit * Gruppe	Pillai-Spur	,240	6,331 ^a	1,000	20,000	,021
	Wilks-Lambda	,760	6,331 ^a	1,000	20,000	,021
	Hotelling-Spur	,317	6,331 ^a	1,000	20,000	,021
	Größte charakteristische Wurzel nach Roy	,317	6,331 ^a	1,000	20,000	,021
a. Exakte Statistik						
b. Design: Konstanter Term + Gruppe						
Innersubjektdesign: Zeit						

Abbildung 6: Auswertungsergebnisse der Varianzanalyse