

Appendix 2: risk of bias assessment

Citation(s)	Measurement error			Criterion validity			Hypotheses testing			Responsiveness		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
PBDE (1994)				30	Very good	Open road $r=0.74$; $p<0.001$ Closed road $r=0.44$; $p<0.05$ (+), result closed road ignored, because low correlation between results open road and closed road	30	Very good	7 hypotheses confirmed (7+)			
Pooled or summary result (overall rating)				30		0.44-0.74 (+)	30		7+ (+)			
WURT (1997)				123	Very good	τ -b=0.60 τ -b/category: 0.26-0.69 (-)	123	Very good	τ -b=-0.27, hypothese confirmed (1+)			
Pooled or summary result (overall rating)				123		τ-b<0.70 (-)	123		1+ (+)			
New Haven (2003)							35	Adequate	3 hypotheses confirmed (3+)			
Pooled or summary result (overall rating)							35		3+ (+)			
TRIP (2003)												
TRIP (2005)				38	Very good	$r=0.8$ (+)						
Pooled or summary result (overall rating)				38		$r=0.8$ (+)						

Note: n: number; r: Pearson correlation coefficient; p: p-value; τ -b: Kendall rank correlation coefficient

Citation(s)	Measurement error			Criterion validity			Hypotheses testing			Responsiveness		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
RIRT (2005)												
RIRT (2012)												
Pooled or summary result (overall rating)	80		A major dimension and a minor dimension (?)	80		0.80 and 0.93 (+)				20		0.87 (+)
SMS (2006)				95	Very good	r=0.94 (+)						
SMS (2010)				127	Very good	ROC analysis: AUC=0.906 (sensi.=0.91 and speci.=0.87) (+)						
Pooled or summary result (overall rating)				222		r=0.94; AUC=0.906 (+)						
P-Drive (2010)												
P-Drive (2011)				85	Adequate	ROC analysis: optimal cut-off at 85 (?)						
P-Drive (2015)				24	Very good	R ² =0.445 (-)						
P-Drive (2015)				99	Very good	ROC analysis: area under the curve=0.98 (sensi.=0.93 and speci.=0.92) (+)						
Pooled or summary result (overall rating)				208		(±)						

Note: n: number; α : Cronbach alpha; r: Pearson correlation coefficient; IR: inter-rater; ROC: receiver operating characteristic; AUC: area under the curve; R²: coefficient of determination (square of Pearson's coefficient)

Citation(s)	Measurement error			Criterion validity			Hypotheses testing			Responsiveness		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
CDAS (2012)												
Pooled or summary result (overall rating)												
NNDA (2012)												
Pooled or summary result (overall rating)												
DOS (2013)	33	Adequate	SEM=3% (?)									
Pooled or summary result (overall rating)	33		SEM=3% (?)									
RODE (2015)												
Pooled or summary result (overall rating)												
UWO (2016a)							35	Very good	2 hypotheses confirmed (2+)			
UWO (2016b)												
Pooled or summary result (overall rating)							35		2+ (+)			

Note: n: number; SEM: standard error of the measurement

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
PBDE (1994)				30	Doubtful	Closed road $\alpha=0.78$ Open road $\alpha=0.89$ (?)				30	Doubtful	IR: closed road $r=0.84$; open road $r=0.74$ (+)
Pooled or summary result (overall rating)				30		0.78-0.89 (?)				30		0.74-0.84 (+)
WURT (1997)										TR: 63	Doubtful	TR: $\tau-b=0.76$ (?)
Pooled or summary result (overall rating)										63		$\tau-b=0.76$ (?)
New Haven (2003)				35	Doubtful	$\alpha=0.88$ (?)				357	Adequate	IR: ICC (scale)=0.99 ICC/items: for 26, ICC: 0.911-0.998 and for 10, ICC>0.80 (+)
Pooled or summary result (overall rating)				35		0.88 (?)				357		ICC>0.80 (+)

Note: n: number; α : Cronbach alpha; r: Pearson correlation coefficient; $\tau-b$: Kendall rank correlation coefficient ; ICC: intraclass correlation coefficient; IR: inter-rater; TR: test-retest

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
TRIP (2003)										27	Adequate	IR: total score ICC: 0.62-0.84 (+), video VS video ICC: 0.80-0.84 and real-life VS video ICC: 0.62-0.64 with large CI considered in the risk of bias
TRIP (2005)										38	Adequate	IR: total score ICC=0.83 (+)
Pooled or summary result (overall rating)										65		0.62-0.84 (+)
RIRT (2005)										20	Doubtful	IR: r=0.87 (+)
RIRT (2012)	80	Inadequate	2 subscales (one explains 31% and the other 8% of the variance) (?)	80	Very good	α /subscales=0.93 and 0.80 (+)						
Pooled or summary result (overall rating)	80		A major dimension and a minor dimension (?)	80		0.80 and 0.93 (+)				20		0.87 (+)

Note: n: number; α : Cronbach alpha; r: Pearson correlation coefficient; ICC: intraclass correlation coefficient; CI : confidence interval; IR: inter-rater

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
SMS (2006)				95	Doubtful	$\alpha=0.94$ (?)				IR: 33 TR: 10	Adequate	IR: ICC: 0.88 (dich. scores) 0.94 (scale with 4 pts/item) (+) TR: ICC: 0.91 (dich. scores) 0.95 (scale with 4 pts/item) (+)
SMS (2010)												
Pooled or summary result (overall rating)				95		0.94 (?)				43		0.88-0.95 (+)
P-Drive (2010)	205	Very good	3 items outside the criteria for goodness-of-fit Unidim. scale (PCA: 59.1% for the principal component and 4.9% for the second one, with a rule of thumb set at resp. 50% and 5%) (-)			205	Inadequate		Differences in item functioning for 4 items between CVA and MCI (-)			
P-Drive (2011)												

Note: n: number; α : Cronbach alpha; PCA: principal component analysis; ICC: intraclass correlation coefficient; IR : inter-rater; TR: test-retest; CVA : cerebrovascular accident; MCI: mild cognitive impairment

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability			
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	
P-Drive (2015)										24	Very good	IE: ICC random effect=0.950 (CI 95%: 0.889-0.978); ICC/category: 0.875-0.963 (+)	
P-Drive (2015)	99	Doubtful	3 items outside the criteria of goodness-of-fit Unidim. scale PCA: 80.3% for the principal component and 2.4% for the second one, with a rule of thumb set at resp. 60% and 5%) (-)										
Pooled or summary result (overall rating)	304		3 items outside the criteria of goodness-of-fit but PCA suggests unidim. (-)				205					24	0.875-0.978 (+)
CDAS (2012)	47	Inadequate	2 subscales (one explains 14% and the other 12% of the variance) (?)	47	Very good	$\alpha=0.89$ $\alpha=0.73$ (+)							
Pooled or summary result (overall rating)	47		2 dimensions of equal importance (?)	47		0.73 and 0.89 (+)							

Note: n: number; α : Cronbach alpha; PCA: principal component analysis; ICC: intraclass correlation coefficient; CI: confidence interval; IR: inter-rater; CVA: cerebrovascular accident; MCI: mild cognitive impairment

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
NNDA (2012)										6	Inadequate	IR: Level of agreement: 100% for 7/25 items and discrepancies between minor errors and no error 13/25 and discrepancies between minor errors/no error and major errors 5/25 and significant discrepancies for 6/150 observations (4%) (-)
Pooled or summary result (overall rating)										6		Use of level of agreement's, doesn't take into account chance(-)
DOS (2013)										33	Adequate	IR: ICC=0.91 (+)
Pooled or summary result (overall rating)										33		0.91 (+)
RODE (2015)										24	Adequate	IR: ICC: 0.84-0.97 (+)
Pooled or summary result (overall rating)										24		0.84-0.97 (+)

Note: n: number; IR: inter-rater; ICC: intraclass correlation coefficient

Citation(s)	Structural validity			Internal consistency			Cross-cultural validity/Measurement invariance			Reliability		
	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)	n	Meth. Qual.	Results (rating)
UWO (2016a)												
UWO (2016b)										34	Doubtful	IR: GRS (dich.), $\kappa=0.892$; GSR (4 levels), $\kappa=0.952$ PERS, error 1 : $\kappa=0.888$; error 2 : $\kappa=0.847$; error 3 : $\kappa=0.902 (+)$
Pooled or summary result (overall rating)										34		0.847-0.952 (+)

Note: n: number; κ : Cohen's kappa; IR: inter-rater; GRS: global rating scale; PERS: priority error rating score