

Supplemental Material S3. Ratings of measurement properties and risk of bias for individual studies of theory of mind assessment measurement properties.

Assessment	Citation	Methodological Quality/Rating of Sufficiency of the Measurement Property						
		Structural Validity	Internal Consistency	Measurement Invariance	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing
IRI	Davis, 1980	V ?	V +	NR	A ?	NR	NR	NR
	Davis, 1983	NR	NR	NR	NR	NR	NR	V + V - ¹
	Melchers et al., 2015	NR	V + V - ¹	NR	NR	NR	NR	Convergent V+ V- ¹ Convergent
MASC	Fossati et al., 2017	A + (adolescent) A - (adult)	V +	NR	NR	NR	NR	V + Convergent
	Lahera et al., 2014	NR	V +	NR	I ?	NR	V +	V + Convergent
	Martinez et al, 2017	NR	V +	NR	NR	NR	V +	NR
	Muller et al., 2016	I ?	V +	NR	NR	NR	V +	V + Convergent
NEPSY II Affect Recognition	Korkman et al., 2007	NR	D ?	NR	A ?	I ?	NR	V + Group Differences
NEPSY II Theory of Mind	Korkman et al., 2007 Internal Consistency	NR	D ?	NR	A ?	I ?	V - V + ²	V + Group Differences

Assessment	Citation	Methodological Quality/Rating of Sufficiency of the Measurement Property						
		Structural Validity	Internal Consistency	Measurement Invariance	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing
	McKown et al., 2013	NR	D ?	NR	D ?	NR	NR	V + Convergent
6-Step Theory of Mind Scale	Peterson et al., 2012	NR	NR	NR	NR	NR	NR	V + Convergent
TASIT	Mathersul et al., 2013	NR	NR	NR	NR	NR	NR	V + Group Differences
	McDonald, Flanagan & Rollins, 2017 ³	NR	D ?	NR	A ?	NR	NR	NR
	Morrison et al., 2019	NR	D ?	NR	NR	NR	NR	V + Group Difference & Convergent
TASIT-S	McDonald et al., 2018	NR	NR	I + (Part 1,2) I – (Part 3)	NR	NR	NR	NR
	McDonald, Flanagan & Honan, 2017; Honan et al., 2016 (same studies)	V +	I ? (Part 1) V+ (Part 2,3)	NR	NR	NR	NR	V + Convergent

Assessment	Citation	Methodological Quality/Rating of Sufficiency of the Measurement Property						
		Structural Validity	Internal Consistency	Measurement Invariance	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing
ToMI-2	Greenslade & Coggins, 2016	NR	I ? ⁴	NR	NR	NR	NR	V + Group Difference & Convergent V + Group Differences
	Hutchins et al., 2016 (First 42 Items)	I ?	I ? ⁴	NR	A ?	V ?	V +	
	Hutchins et al., 2016 (All Items)	NR	I ? ⁴	NR	NR	V ?	NR	NR
	Jiang et al., 2016	NR	I ? ⁴	NR	D +	NR	NR	NR
	Lee et al., 2021	I +	NR	NR	NR	NR	NR	NR
	Pujals et al., 2016	NR	I ? ⁴	NR	NR	NR	NR	V + Group Differences
ToMI-SR-A	Crehan et al., 2021	NR	NR	NR	NR	NR	NR	V + Group Difference & Convergent V + Group Differences
	Hutchins et al., 2019 ⁵	I ?	V +	NR	NR	NR	V +	V + Group Differences
	Hutchins et al., 2021 ⁵	I ?	V +	NR	NR	NR	V +	V + Group Differences
ToMTB	Hutchins et al., 2008	NR	D ?	NR	V + V – (six items)	NR	NR	NR

Assessment	Citation	Methodological Quality/Rating of Sufficiency of the Measurement Property						
		Structural Validity	Internal Consistency	Measurement Invariance	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing
	Hutchins et al., 2010	NR	D ?	NR	NR	NR	NR	V + Convergent

Notes: Methodological quality rated V = very good, A = adequate, D = doubtful, or I = inadequate based on COSMIN risk of bias process. Ratings of measurement property sufficiency are + for sufficient, - for insufficient, or ? for indeterminate. NR = not rated in the study. ¹ IRI internal consistency and convergent validity rated - for personal distress subscale only. ² NEPSY II Theory of mind scale reported to have sufficient specificity but insufficient sensitivity to identify children with ASD. ³ Measurement properties based on sample including people with schizophrenia. ⁴ Internal consistency calculated for all items rather than for each set of unidimensional items. ⁵ Hutchins et al., (2019; manual) and Hutchins et al., (2021) appear to report the same studies.