

### Evidence Types Organized by Source of Evidence

Consequences of Testing	Internal Structure	Relations to Other Variables	Reliability	Response Process	Test Content
Appropriate cut score	Bayesian Network Models	Alignment with expert opinion of test user (e.g. teacher, therapist)	Alternate form	Cognitive interviews	Alignment with frameworks/standards/theory/learning trajectory
Bias as one consequence of testing	Cluster analysis	Convergent Association or Divergent Association	EAP/PV test reliability	Error related to response patterns/CTT	Construct Definition
Cost-benefit analysis	Factor Analysis- Bifactor	Correlation analysis	Generalizability theory- D-studies	Eye tracking/physiological data	Data from experts
Documentation of unintended behavior changes based on test use	Factor Analysis- Confirmatory Factor Analysis (CFA)	Discriminant validity	Generalizability theory- G-studies	fMRI (functional Magnetic Reasoning Image)	Fairness of content
Explicit intended uses and interpretations and warn against inappropriate uses	Factor Analysis- Exploratory Factor Analysis/Exploratory Structural Equation Modeling	Discrimination power	Inter-rater reliability- Kappa	Focus groups	Field Work
Impact of assessment is similar under clinical and practical implementations	Factor Analysis- Multi-trait Multi-method matrix (MTMM)	Hierarchical Linear Modeling	Inter-rater reliability- Percent agreement	Generalizability-theory (G-theory) related evidence	Literature Review
Item functioning such as DIF - unknown subgroups had to know	Factor Analysis- Not specified	Multi-trait Multi-method matrix (MTMM)	Internal consistency or alternatives- Alpha	Log data	Participant-generated content
Motivational consequences	Factor Analysis - Parallel Analysis	Statistical Testing (e.g., t-test, regression, and chi-square)	Internal consistency or alternatives- IRT or Rasch reliability	Predicted response patterns/processes based on Learning Trajectories	Revision Process
	Factor Analysis- Principal Axis Factoring (PAF)	Structural Equation Model	Internal consistency or alternatives- Omega	Rater agreement/reliability	Standard Setting
	Factor Analysis- Principal Component Analysis (PCA)	Treatment/control study	Internal consistency or alternatives- Raykov	Rater training and calibration	
	Item difficulty	Triangulation/Crystallization with qualitative data	Item remainder correlations	Sorting tasks	
	Item Scale Correlations		Kuder-Richardson formula 20	Student written work	
	Item Response Theory (IRT)		SEM measurement	Studies of respondent's speed of task completion	
	Latent Class Analysis (LCA)		Sensitivity analysis	Think alouds	
	Latent Profile Analysis (LPA)		Split-half reliability		
	Multidimensional scaling		Test - Retest		
	Rasch modeling		Yule's Y		
	TETRAD				