www.activationlab.org
info@activationlab.org

Measures Technical Brief
Competency Beliefs in STEM version 1.0 March 2017

Overview

Description of the Construct

Competency beliefs are the learner's beliefs about their ability to successfully participate in diverse STEM learning situations as well as their beliefs about having the core skills of required for STEM learning. Competency beliefs are a core construct in social cognitive theory, defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986). In general, educational and psychological research has revealed that competency beliefs (or self-efficacy beliefs) are an important predictor of many types of achievement behavior (i.e., choice of task, engagement, effort, and persistence) (Schunk, Pintrich, & Meece, 2008). Educational and psychological research makes a clear distinction between people's actual competence and knowledge, and their subjective judgment and perceptions of them. Further, this body of research distinguishes between the role of actual and perceived competence in predicting their achievement tests, and achievement behaviors. For example, reasoning ability often plays a more significant role than self-efficacy in predicting their achievement in STEM learning (e.g., Lawson, Banks, & Logvin, 2007). By contrast, learners with high self-efficacy beliefs are more likely to be behaviorally and cognitively engaged in a learning process in the forms of choice, effort, persistence, and so on (Linnenbrink & Pintrich, 2003, insert math thing). We assess learner's beliefs about their competency to perform in diverse situations as well as their competency beliefs about particular skills within STEM.

Intended Uses of the Instrument

The STEM Learning Activation survey is designed for 10-14 year-old respondents. The STEM Competency Beliefs scale of the STEM Learning Activation survey is to measure an individual's STEM Competency Beliefs at the time the survey responses are collected. The construct is conceived as semi-malleable and therefore is amenable to intervention. However, changes in scale scores are not expected to occur due to single hour-long experience/intervention. In longitudinal study contexts, such as repeated measures over longer periods of time, a minimum of weeks between pre- and post-administrations is suggested. The survey is not intended for high-stakes decisions about students (e.g., pass/fail decisions, admission decision to a program) or programs (e.g., accountability).

How to Score

The STEM Learning Activation survey is designed to be constructed and calibrated under the Rasch measurement model—the Partial Credit model (PCM; Masters, 1982). Statistical or programing programs, such as Winsteps (Linacre, 2016), ConQuest (Adams, Wu, & Wilson, 2015), WinBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009), PARSCALE (Muraki & Bock, 2003), and the R packages eRm (Mair & Hatzinger, 2007) and TAM (Kiefer, Robitzsch, & Wu, 2016), are available for model estimation. In the Rasch measurement model, the probability of getting a response is specified as a logistic function of the person and item parameters. Basically, the Rasch measurement model calibration is used here mainly to inform the measurement quality and effectiveness of the STEM Learning Activation instrument. For score interpretation, a simple strategy for those who are not test developers or who are applied practitioners is to use raw scores—a summation or an average score of the STEM Competency Beliefs (i.e., an average or a sum score of all 12 items). While using a summation or an average score of the STEM Competency Beliefs for interpretation, it is noted that the contribution of each item to the latent trait (i.e., STEM Competency Beliefs) is assumed to be equal which is unlikely to be true; and that variations of an average score is limited. Also, no or few missing data is always desirable; and a prevalence of missing data issue should be handled prior to analysis.

Analytical Options

Both Rasch ability parameter estimates or raw scores can be treated as a continuous variable. Depending on researchers' and evaluators' different research interests, comparison tests (e.g., t-test, ANOVA), correlation analyses, or predication modeling (e.g., regression model, structural equation model [SEM]) can be further conducted with various scenarios.

Scale Construction

<u>Calibration Model</u>

As aforementioned, the STEM Learning Activation survey is originally constructed and calibrated under PCM (Masters, 1982). The rationale is that response option formats varies across four subscales in the STEM Learning Activation scale. The PCM is adequate for the STEM Learning Activation survey because it releases the restriction of equal threshold distance across all items which is held for Rating Scale Model (RSM). Empirically, a likelihood ratio test comparison between PCM and RSM also indicates that PCM provides better fit than RSM (i.e., LR statistics = 76.110, *df* = 22, *p* < .05).

<u>Survey Questions</u>

The STEM Competency Beliefs scale has 12 items, with response options from 1 to 4.

Item ID	Prompt	Response Options and Coding
CB1	I can do the math problems I get in class	4=All of the time; 3=Most of the time; 2=Half the time; 1=Rarely
CB2	I can understand scientific information on websites for kids my age	4=All websites; 3=Most websites; 2=A few websites; 1=None of them
CB3	If I did my own project in an after school science club, the project would be…	4=Excellent; 3=Good; 2=Okay; 1=Poor
CB4	I am the technology expert in my house	4=All of the time; 3=Most of the time; 2=Half the time; 1=Rarely
CB5	I can understand the science in books for adults	4=All of the time; 3=Most of the time; 2=Some of the time; 1=A little of the time
CB6	I think I am very good at: Figuring out how to fix things that don't work.	4=YES!; 3=yes; 2=no; 1=NO!
CB7	I think I am very good at: Giving evidence when I tell my opinion.	4=YES!; 3=yes; 2=no; 1=NO!
CB8	I think I am very good at: Explaining my solutions to math problems.	4=YES!; 3=yes; 2=no; 1=NO!
CB9	I think I am very good at: Solving problems.	4=YES!; 3=yes; 2=no; 1=NO!
CB10	I think I am very good at: Coming up with my own science investigations.	4=YES!; 3=yes; 2=no; 1=NO!
CB11	I think I am very good at: Coming up with new ways to solve technical problems.	4=YES!; 3=yes; 2=no; 1=NO!
CB12	I think I am very good at: Coming up with new ideas when working on projects.	4=YES!; 3=yes; 2=no; 1=NO!

<u>Psychometric Properties</u>

To examine the measurement quality of the STEM Competency Beliefs Scale, the fit assessment includes checking unidimensionality assumption, item characteristics, as well as PCM model-level and item-level fit statistics. Analyses are based on a total sample of 205 middle school youth.

Unidimensionality (*Construct Validity*)

The STEM Competency Beliefs items are presumed to measure essentially only one latent trait— STEM Competency Beliefs. A confirmatory factor analysis (CFA) with WLSMV estimator (i.e., an estimator for modeling categorical or ordinal data) was conducted to assess the unidimensionality assumption. The purpose is to know whether the set of 12 items measure what it purports to measure.

Results of all CFA model fit indicators meet the criteria (see Hu & Bentler, 1999), indicating that data well fit the one-factor model (i.e., χ^2 = 84.059, *df = 54, p* = 0.005*;* Comparative Fit Index [CFI]=0.974; Tucker-Lewis Index [TLI] = 0.969; Root Mean Square of Approximation [RMSEA] = 0.052, Standardized Root Mean Square Residual [SRMR] = 0.070). All of 12 items generate significantly positive factor loadings to the factor, ranging from 1.699 to 2.417 ($p < .05$).

Descriptive statistics

Table 1 is item-level descriptive statistics, including the number of observed samples; and mean, standard deviation (SD), and median of each item. There are 204 to 205 cases for each of the items (i.e., only one missing value for items 1-5), with mean values ranging from 2.34 to 3.41. Table 2 gives the frequency for response options for each of the items. Overall, there are more observations for response options 3, response 4, and then followed by response option 2 and response option 1. Items 4 and 5, when compared to the rest of the items, have more observations for response option 1 and fewer observations for response option 4; this means that these 2 items are more difficult for students to choose higher item categories (e.g., YES!).

Item characteristics

Figure 1 is item characteristic curve (ICC) plots for each of the items, and Figure 2 is a person-item map. In the Rasch measurement model, item and ability parameters are on the same scaling continuum. The PCM allows for different category distances across items (see Figure 1 and Figure 2). These 12 items have varying difficulty levels and are displayed in order of difficulty in Figure 2 (i.e., items are ordered from the easiest [item9] to the most difficult [item 5] for students to endorse higher item categories). The higher item categories are observed to be chosen by students with stronger abilities to agree the statements.

Model level fit

Infit is the weighted fit index, which is sensitive to irregular response patterns to items that are close to a person's ability level; and outfit is the unweighted fit index, which is sensitive to unexpected response patterns to items with difficulty away from a person's ability level (Linacre, 2002). Mean-square statistics (MNSQ) are chi-square statistics divided by degree of freedom (Linacre, 2002).

Overall, data fit the Partial Credit Model well (i.e., INFIT MNSQ = 0.93; OUTFIT MNSQ = 0.92), indicating a productive measurement (i.e., MNSQ values are between 0.5 and 1.5; see Linacre, 2002). The reliability of the STEM Competency Beliefs scale is good (i.e., Cronbach's alpha = 0.83; polychoric alpha = 0.87), indicating that this scale is creating a well-defined latent trait—STEM Competency Beliefs.

Item level fit

Infit and outfit mean-square statistics for 12 items are acceptable (see Table 3), ranging between 0.73 and 1.22, which indicates productive of measurement (see Linacre, 2002).

Table 1 Item Level Descriptive Statistics

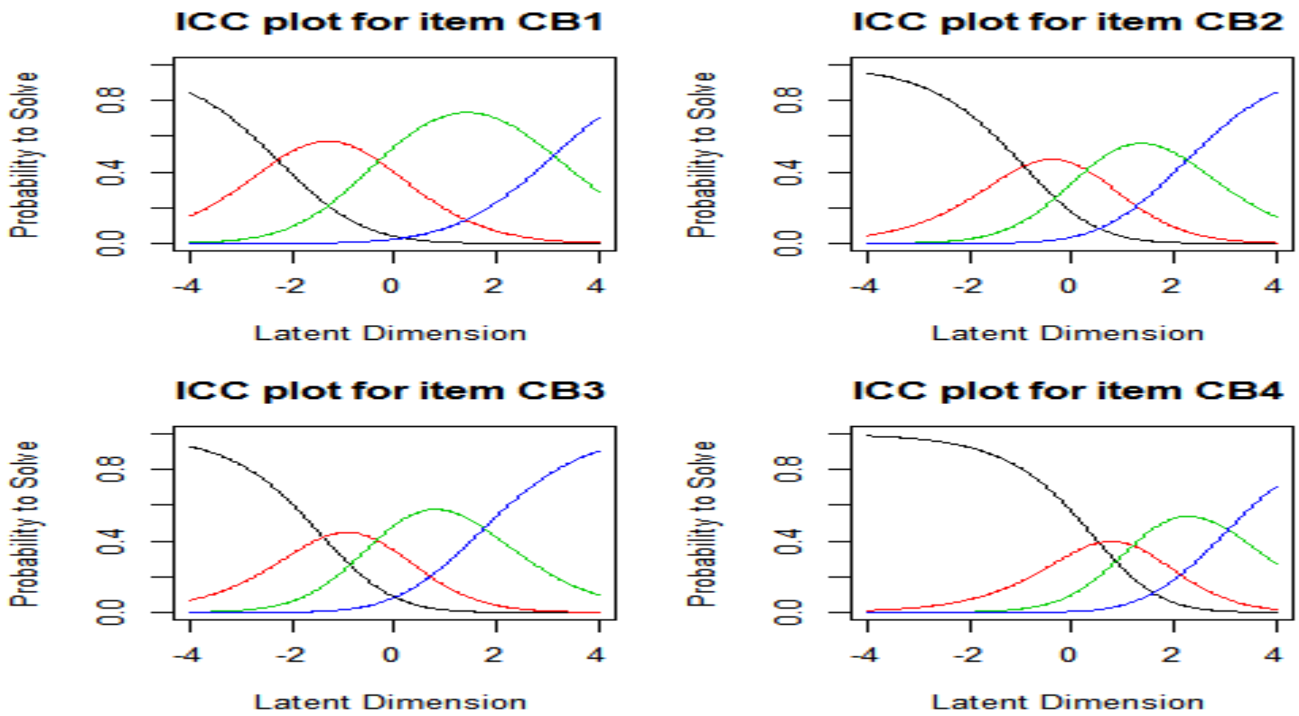
	N	Mean	SD	Median	SE
CB1	204	3.13	0.61	3	0.04
CB2	204	3.19	0.77	3	0.05
CB3	204	3.40	0.68	3.5	0.05
CB4	204	2.72	0.96	3	0.07
CB5	204	2.34	0.81	2	0.06
CB6	205	3.13	0.79	3	0.06
CB7	205	3.27	0.71	3	0.05
CB8	205	3.03	0.82	3	0.06
CB9	205	3.40	0.59	3	0.04
CB10	205	2.97	0.80	3	0.06
CB11	204	3.02	0.80	3	0.06
CB12	205	3.41	0.70	4	0.05

Table 2 Response Options Frequency

	Response 1	Response 2	Response 3	Response 4
CB1	0%	11%	63%	25%
CB2	2%	15%	44%	39%
CB3	1%	8%	41%	50%
CB4	13%	24%	41%	22%
CB5	13%	48%	31%	8%
CB6	3%	15%	47%	35%
CB7	1%	13%	45%	41%
CB8	4%	20%	44%	31%
CB9	0%	4%	50%	45%
CB10	1%	29%	40%	29%
CB11	2%	25%	43%	31%
CB12	1%	9%	38%	52%

Table 3 Item level fit statistics

	Outfit (Unweighted)		Infit (Weighted)	
	MNSQ	t	MNSQ	t
CB1	1.07	0.67	1.08	0.79
CB2	1.03	0.33	1.04	0.44
CB3	0.77	-1.98	0.83	-1.71
CB4	1.22	2.14	1.21	2.17
CB5	1.13	1.32	1.10	1.04
CB6	0.82	-1.81	0.82	-1.90
CB7	0.93	-0.59	0.97	-0.29
CB8	0.83	-1.72	0.86	-1.53
CB9	0.73	-2.64	0.77	-2.22
CB10	0.86	-1.47	0.89	-1.29
CB11	0.91	-0.92	0.88	-1.37
CB12	0.80	-1.65	0.78	-2.23


ICC plot for item CB1
ICC plot for item CB2
ICC plot for item CB3
ICC plot for item CB4
Probability to Solve
0.0
0.4
0.8
-4
-2
0
2
4
Latent Dimension

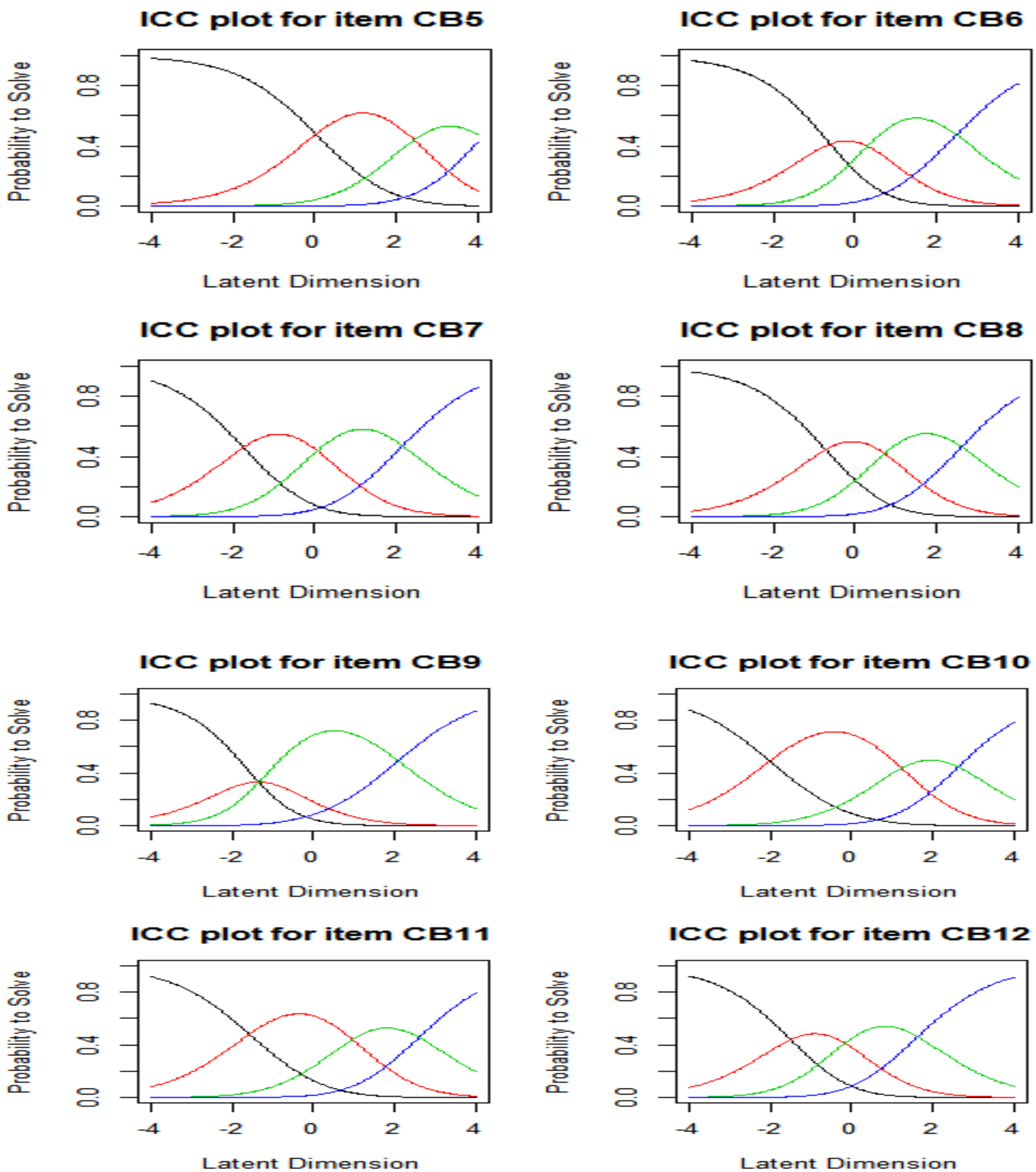


Figure 1 Item Characteristic Curve

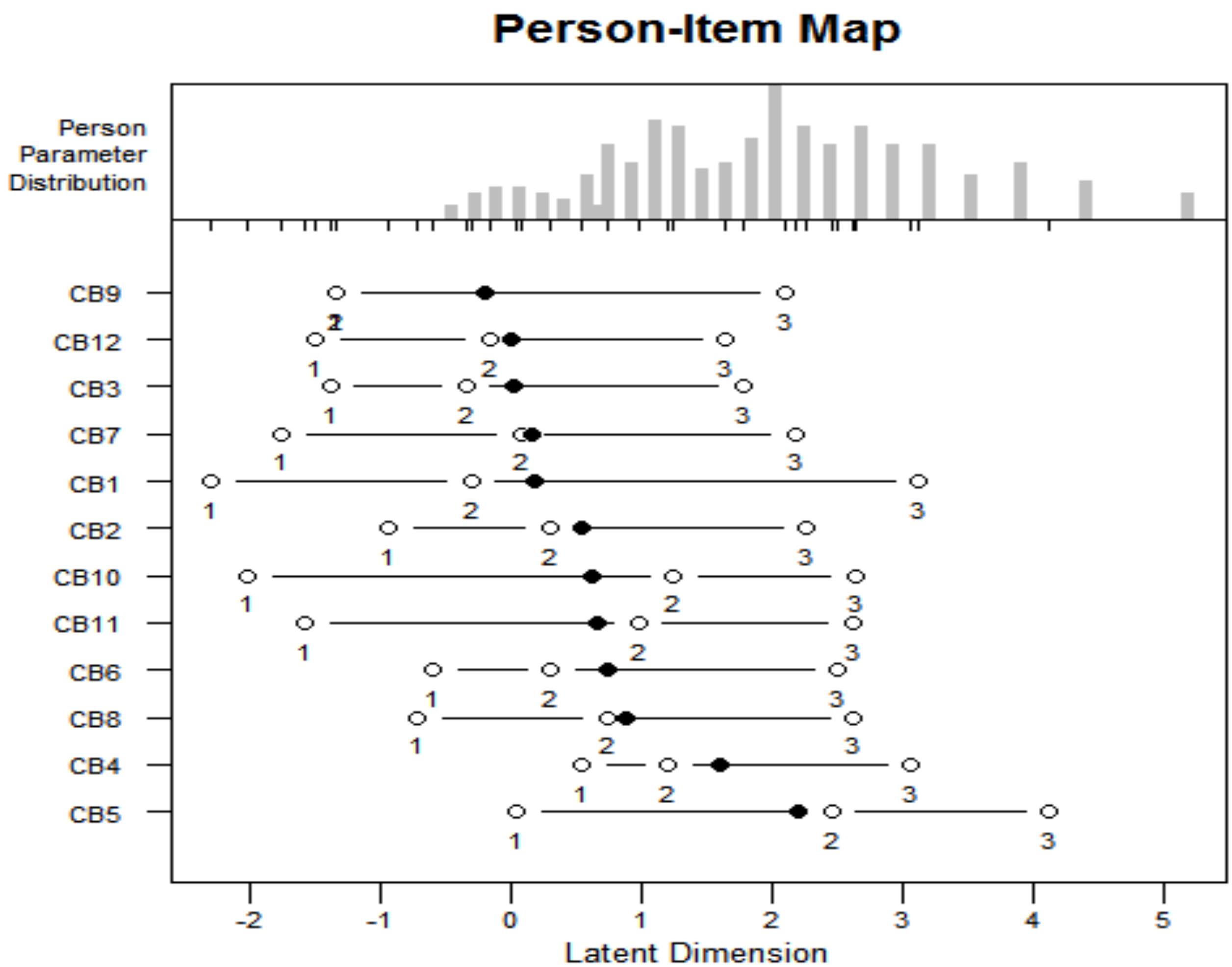


Figure 2 Person-Item Map

This research report was supported by the National Science Foundation under Award Number 1337186. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation.

Suggested Citation:
Chen, Y.-F., Cannady, M. A., Schunn, C., & Dorph, R. (2017) Measures Technical Brief: Competency Beliefs in STEM. Retrieved from http:// www.activationlab.org/wp-content/uploads/2017/06/CompetencyBeliefs_STEM-Report_20170403.pdf

References

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, 47(1), 105-113.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Byrne, B. M. (2010). *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming* (2nd ed.). New York: Routledge.

Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO 2.1 for Windows* [computer software]. Chicago, IL: Scientific Software International.

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7).

Gadermann, A. M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3).

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*, 6(1), 1–55.

Kiefer, T., Robitzsch, A., Wu, M., & Robitzsch, M. A. (2013). *TAM: Test analysis modules* (Version 1.1) [computer software]. <http://cran.r-project.org/package=TA>.

Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching,* 44(5), 706-724.

Linacre, J. M. (2008) *Winsteps* (Version 3.67.0) [computer software]. Winsteps.com, Chicago.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly,* 19, 119-137.

Mair, P., & Hatzinger, R. (2007) Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9):1–20.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in Education: Theory, Research, and Applications* (Third ed.): Pearson Education Inc.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8:3, 370.

Wright, B. D., & Stone, M. H. (1979). Best test design. *Rasch Measurement*. Chicago: Mesa Press.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2.0: Generalised Item Response Modelling Software* [computer software]. Melbourne: Australian Council for Educational Research.