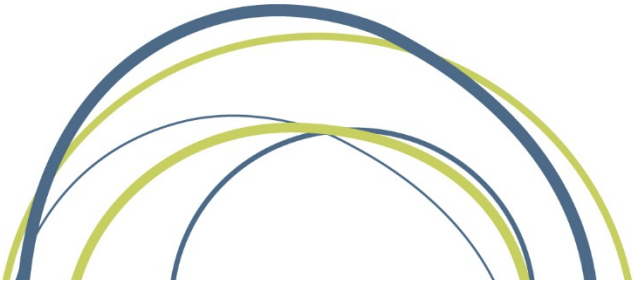
ISTE ESSENTIAL CONDITIONS INVENTORY REPORT ON VALIDITY AND RELIABILITY

11/12/14

Cathleen A. Kennedy, Ph.D.



This report presents evidence of the validity and reliability of the 2014 Essential Conditions Inventory (inventory) relative to its intended purpose. The measure is intended to assess the degree to which respondents meet expectations in the 14 ISTE Essential Conditions to effectively leverage technology for learning established by the International Society for Technology in Education.

The inventory comprises 144 Likert-type questions across 14 subscales (conditions), and the number of response categories ranges from two to six. ISTE staff members developed the items based on their expertise in technology management in education and extensive experience in defining the Essential Condition standards. An initial inventory comprising 137 questions was field-tested with 73 respondents in March 2014. Psychometric analysis of those data revealed that, for some conditions, more items were necessary to achieve adequate reliability, and several items required improvements for clarity. Ninety-eight respondents completed the updated inventory in July and August 2014.

METHOD

A one-parameter partial credit model (Masters, 1982), a polytomous extension of the Rasch model (Rasch, 1960/1980), was utilized to evaluate the data and provide a relatively straightforward interpretation of person and item parameter estimates. Among the 14 conditions, a total of 12 item types were used, from dichotomous "yes-no" questions to customized Likert-type questions with response categories such as *never, sporadically, periodically*, and *regularly*. Most conditions required at least three item types, and the categorical response options were not readily equated across items. Because so many different item types were used, and to more precisely monitor the distance between categories on different items, the partial credit model was chosen for the analyses rather than a rating scale model. Each scale (one per condition) was analyzed individually in a unidimensional model.

Infit and outfit mean square values are presented as quantitative evidence of construct validity. These values can be used to detect the relative fit of items to a unidimensional construct and identify potential construct-irrelevant variance. To evaluate the fit of items to the model, both weighted and unweighted mean squares (MNSQs) and associated *t*-statistics were evaluated, with MNSQs indicating magnitude and *t*-statistics indicating significance of fit to the model. Smith, Schumacker, and Bush (1998) suggested that misfit is evident when MNSQ values are larger than 1.3 for samples less than 500; Adams and Khoo (1996) indicated that items with adequate fit will have weighted MNSQs between 0.75 and 1.33; and Wright and Linacre (1994) stated that MNSQ values of less than 0.6 or greater than 1.4 indicate a practically significant amount of misfit for partial credit items. Items exhibiting large mean squares are considered the greater threat to validity because they may be measuring unmodelled noise.

Test reliability is evaluated using an item response theory (IRT) approach. In IRT, the standard errors of measurement differ across the scale because they are conditional on the observed proficiencies, unlike in classical test theory, where the standard error is assumed to be the same at all levels of ability. The IRT marginal reliability coefficient is a measure of the overall reliability of the test based on the average conditional standard errors across different points on the scale. We also include Cronbach's alpha coefficient (α) as a traditional measure of internal consistency.

FINDINGS

The analyses indicate that the questions produce coherent unidimensional scales for each condition and assess most respondents' locations with good to very good reliability. Table 1 summarizes the findings for each condition. The reliability is quite high for scales with so few items, indicating that the questions are targeting relevant information about technology planning, implementation, and management. The Equitable Access condition could be examined and updated with additional questions, or better questions, to improve reliability. One or two conditions (e.g., Implementation Planning and Skilled Personnel) could be modified to achieve better unidimensional alignment, but they are currently producing reliable measures.

Table 1. Relevant IRT statistics for each Essential Condition of the Summer 2014 administration of the ISTE Essential Conditions Inventory.

Essential Condition	No. Items	IRT Rel.	a Rel.	Infit MNSQ Range	Outfit MNSQ Range
Shared Vision	10	.88	.88	0.75 – 1.32	0.84 – 1.34
Empowered Leaders	12	.90	.88	0.81 – 1.55	0.80 – 1.51
Skilled Personnel	7	.89	.90	0.63 – 1.90	0.59 – 2.34
Engaged Communities	10	.89	.89	0.76 – 1.23	0.75 – 1.43
Technical Support	15	.87	.87	0.69 – 1.71	0.65 – 1.94
Curriculum Framework	13	.91	.92	0.80 – 1.19	0.84 – 1.24
Consistent & Adequate Funding	5	.84	.88	0.76 – 1.21	0.67 – 1.25
Assessment & Evaluation	9	.87	.87	0.81 – 1.20	0.85 – 1.16
Professional Development	7	.84	.82	0.87 – 1.24	0.84 – 1.48
Implementation Planning	11	.95	.95	0.75 – 2.33*	0.72 – 2.58*
Equitable Access	11	.81	.76	0.83 – 1.29	0.80 – 1.38
Support Policies	17	.90	.93	0.81 – 1.23	0.69 – 1.43
Supportive External Context	11	.86	.86	0.87 – 1.29	0.85 – 1.38
Student-Centered Learning	6	.85	.88	0.77 – 1.22	0.75 – 1.23

*IP is the least coherent scale. This appears to be due primarily to one item, which could be modified or excluded in the future (IP1). The scale is highly reliable, however, with the item included.

The findings from the psychometric analyses indicate that the vast majority of questions worked very well to inform a unidimensional condition scale and produced good to very good reliability.