Content Validity of a Mathematics Placement Test for a Gifted High School

Hannah R. Anderson, MA

Kent State University – Kent, OH

Abstract

The Content Validity of a mathematics placement test at a Science, Technology, Engineering, and Mathematics (STEM) gifted residential high school is examined. Data were collected from internal and external mathematics subject matter experts (SMEs) using a card-sorting task and were analyzed using Multidimensional Scaling (MDS) and Hierarchical Cluster Analysis (HCA). Results demonstrate some congruence between the two configurations, suggesting marginal evidence of Content Validity.

Keywords: Content Validity, Mathematics, Gifted Education, STEM Education

Author Note

Correspondence concerning this paper should be addressed to Hannah R. Anderson, MA, Doctoral Candidate of Evaluation and Measurement, School of Foundations, Leadership, and Administration, Kent State University, 300 White Hall, Kent, OH 44242. E-mail: hwilso20@kent.edu.

Content Validity of a Mathematics Placement Test for a Gifted High School

Purpose:

The purpose of this study was to demonstrate evidence of Content Validity on a mathematics placement test at a Science, Technology, Engineering, and Mathematics (STEM), gifted, residential high school. Previous research on placement exams have been conducted at the post-secondary level; however, this study extends the research to younger grade levels serving a specific, gifted population.

Theoretical Framework:

Strengthening education in the disciplines of science and mathematics has been emphasized in the United States since the early 1980s. Understanding this critical need, the National Science Foundation began a nation-wide STEM movement to cultivate a globally-recognized innovative and diverse workforce. Shortly thereafter, specialized STEM schools were formed to reinforce the importance of STEM literacy. In the current study, the selected high school enrolls academically talented Illinois students (i.e., Grades 10 through 12) in its advanced, residential college preparatory program with an emphasis in the fields of mathematics and science.

With the accountability movement ever-present since the implementation of *No Child Left Behind*, post-secondary institutions began to consider the rigor and defensibility of their placement practices (Armstrong, 2000; Clark & Watson, 1995; Morgan & Michaelides, 2005). Additionally, research has noted a void in the validation of exam scores, with no known studies examining the placement of adolescents from middle school into high school for gifted students.

Developed in 1985, the continuing purpose of this placement test is to determine a student's incoming mathematical knowledge for appropriate course placement commensurate

with ability level. Like most teacher-made tests, the items were constructed by the mathematics faculty members at the high school without any psychometric evaluation. Thus, the research question of the current study is, "What is the Content Validity of the items on a mathematics placement test for gifted, residential high school students interested in STEM?" Multidimensional Scaling (MDS) and Hierarchical Cluster Analysis (HCA) were used to address this research question.

Methods:

Participants and Procedures. The sample for this study was comprised of nine internal and eight external Subject Matter Experts (SMEs). The external SMEs were recruited from varying educational levels across the state of Illinois such as high school, community college, and 4-year universities. The SMEs were asked to complete a card-sorting task for the 107 mathematics placement exam items by arranging them into meaningful piles based on the similarity of their content. Consistent with Trochim's (1989) sorting rules, instructions included: (1) placing each item into only one pile, (2) refraining from creating as many piles as there are items, and (3) creating more than one pile. Finally, SMEs recorded the item numbers in each pile, assigning every group of items a title or name.

Data Analysis. MDS and HCA were used to compare the similarity ratings of external and internal SMEs (D'Agostino, Karpinski, & Welsh, 2011). In Content Validation studies that use MDS, similarity ratings from SMEs are compared to the original test specifications (D'Agostino et al., 2011; Li & Sireci, 2013; Sireci & Geisinger, 1992, 1995). One disadvantage in the current study is the absence of test specifications. However, prior research has demonstrated the complementary use of MDS and HCA in the development of content specifications for professional certification exams (Raymond, 1989; Schaefer, Raymond, &

Stamps White, 1992). Thus, the design of this study made use of internal SME item-similarity ratings to develop the content specifications.

The MDS analysis was conducted using Euclidean distances with the selection of dimensions based on the following three considerations: (1) the values of the fit indices (i.e., Kruskal's Stress Function and the Squared Correlation Index R^2), (2) the amount of change in fit indices from *n* to *n* – 1 dimensions, and (3) the interpretability of the dimensions (Kruskal, 1964; Whaley & Longoria, 2009). To examine the change in fit indices from *n* to *n* – 1 dimensions, a plot similar to Cattell's Scree Test (Cattell, 1966) was used.

Once the final MDS solutions were identified, the item coordinates from these solutions were analyzed using HCA with an agglomerative clustering method. A scree plot was used, along with a dendrogram, to determine the number of clusters to retain and to investigate any changes in distances between clusters. After obtaining the final cluster solutions for both the internal and external SMEs, the two configurations were compared using the Rand and adjusted Rand indices.

Data:

The Mathematics Placement Test was created in 1985 by current faculty members. The two-part exam was designed to measure mathematical knowledge needed prior to entering into a Calculus sequence. Students are given a total of 130 minutes to complete 107 items, without using a calculator. Part I of the assessment contains 50 short-answer questions covering content such as simplifying expressions, functions, and exponents. Part II of the assessment has 57 multiple-choice questions related to topics such as graphing, Trigonometry, and Geometry. The multiple-choice questions provide students with five response options which include the correct response, three distractor responses, and a final response option of "I don't know." All responses

of the placement test are graded by mathematics faculty members using an answer key for dichotomous scoring (i.e., “Correct” or “Incorrect”). Therefore, final scores on the assessment can range from a minimum of 0 to a maximum of 107.

Results/Conclusions:

Multidimensional Scaling. The item-dissimilarity matrix for both internal and external SMEs was subjected to multidimensional scaling (MDS). Based on the analysis, stress indices and proportions of variance explained were compared for each configuration between six and nine dimensions as shown in Table 1 below.

Table 1

Fit Indices for Multidimensional Scaling Analysis

Number of Dimensions	Internal SMEs		External SMEs	
6	S = .12542	T = .99210	S = .13417	T = .99096
7	S = .11681	T = .99315	S = .11328	T = .99356
8	S = .09656	T = .99533	S = .09704	T = .99528
9	S = .08645	T = .99626	S = .08649	T = .99625

Note. S = Kruskal's Stress (Stress-I), T = Tucker's Coefficient of Congruence

Taking into account the interpretability of the dimensions, in addition to the information above, the final MDS solution contained eight dimensions for both internal and external SMEs. Next, HCA was conducted on the final item coordinates of the eight dimensional solutions.

Hierarchical Cluster Analysis. The goal of Hierarchical Cluster Analysis (HCA) in the current study was to compare the domain structure of the internal SMEs versus that of the external SMEs. To determine how many clusters to retain, a scree plot, a dendrogram, and the agglomeration schedule were examined for large distances between two cluster groupings, suggesting a possible final solution. Taking into consideration the cluster structure and interpretability, it was determined that a six cluster solution would be retained for both internal and external SMEs. Thus, the final six clusters were: (1) Algebraic Operations, (2) Solving Equations, (3) Graphing Functions, (4) Evaluating Functions, (5) Trigonometry, and (6) Geometry.

Finally, in order to quantify the degree of consensus between the two configurations, the Rand and adjusted Rand indices were calculated. The Rand index was 0.630, with an adjusted Rand index of 0.127. Therefore, the results suggest a slight congruence between the two domain definitions, providing marginal evidence of Content Validity.

Significance of Study:

This study provides a first step in analyzing the psychometric properties of the mathematics placement test. Results are intended to act as a reference for other schools with a STEM and/or gifted education focus to begin the validation process and extend and improve upon the educational testing practices at other levels of schooling. Moreover, validity is context- and population-specific (Schmidt & Hunter, 1977). Thus, mathematics placement exams designed for the general student population can produce biased results without further psychometric scrutiny and documentation. Evidencing the necessary psychometric support for the sample used and the context of the study through rigorous Content Validation procedures is needed to ultimately produce reliable and valid scores resulting in unbiased study results.

References

Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research and Practice, 24*(8), 681-695.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological assessment, 7*(3), 309-319. doi:10.1037/1040-3590.7.3.309

D'Agostino, J., Karpinski, A., & Welsh, M. (2011). A method to examine content domain structures. *International Journal of Testing, 11*(4), 295-307.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1-27.

Li, X., & Sireci, S. G. (2013). A new method for analyzing content validity data using multidimensional scaling. *Educational and Psychological Measurement, 73*(3), 365-385.

Morgan, D. L., & Michaelides, M. P. (2005). *Setting cut scores for college placement* (2005-9). Retrieved from New York:

Raymond, M. R. (1989). Applications of multidimensional scaling to research in the health professions. *Evaluation & the Health Professions, 12*(4), 379-408.

Schaefer, L., Raymond, M., & Stamps White, A. (1992). A comparison of two methods for structuring performance domains. *Applied measurement in education, 5*(4), 321-335. doi:10.1207/s15324818ame0504_3

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of applied psychology, 62*(5), 529.

Sireci, S. G., & Geisinger, K. F. (1992). Analyzing test content using cluster analysis and multidimensional scaling. *Applied psychological measurement, 16*(1), 17-31.

Sireci, S. G., & Geisinger, K. F. (1995). Using subject-matter experts to assess content representation: An MDS analysis. *Applied psychological measurement, 19*(3), 241-255.

Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and program planning, 12*(1), 1-16.

Whaley, A. L., & Longoria, R. A. (2009). Preparing card sort data for multidimensional scaling analysis in social psychological research: A methodological approach. *The Journal of social psychology, 149*(1), 105-115.