AC 2011-1498: STUDYING THE RELIABILITY AND VALIDITY OF TEST SCORES FOR MATHEMATICAL AND SPATIAL REASONING TASKS FOR ENGINEERING STUDENTS

Laura L. Pauley, Pennsylvania State University, University Park

Laura L. Pauley, professor of mechanical engineering, joined the The Pennsylvania State University faculty in 1988. From 2000 to 2007, she served as the Professor-in-Charge of Undergraduate Programs in Mechanical and Nuclear Engineering. In 2003, Laura received the Penn State Undergraduate Program Leadership Award. Dr. Pauley teaches courses in the thermal sciences and conducts research in computational fluid mechanics and engineering education. She received degrees in mechanical engineering from University of Illinois (B.S. in 1984) and Stanford University (M.S. in 1985 and Ph.D. in 1988). She can be contacted at LPauley@psu.edu .

Jonna M. Kulikowich, The Pennsylvania State University

Jonna M. Kulikowich is Professor of Education at Penn State, University Park. Her areas of research and teaching are in test development, applied psychometrics, and statistics. Professor Kulikowich's recent program of research focuses on academic development in mathematics and science. Specifically, she is interested in the measurement of reading comprehension and problem solving in mathematics and science.

Nell Sedransk, National Institute of Statistical Sciences

Dr. Nell Sedransk, Co-Pi, is the Associate Director of the National Institute of Statistical Sciences (NISS) and Professor of Statistics at North Carolina State University. She is an Elected Fellow of the International Statistical Institute, also Elected Fellow of the American Statistical Association. She has served as Associate Editor for the Journal of the American Statistical Association, the Journal of Statistical Planning and Inference, and has been Vice-Chair of the Publication Board of the American Statistical Association. The areas of her technical expertise and current research include design of complex experiments, Bayesian inference, spatial statistics and topological foundations for statistical theory. She received her Ph.D. in Statistics in 1969 from Iowa State University. She can be contacted at sedransk@niss.org

Renata S. Engel, Pennsylvania State University, University Park

Renata S. Engel is Associate Dean for Academic Programs and Professor of Engineering Design and Engineering Science & Mechanics. A member of the Penn State faculty since 1990, she served from 2000-2006 as the Executive Director of the Schreyer Institute for Teaching Excellence. Through various collaborative efforts, she has affected changes in the engineering curriculum at Penn State, primarily to incorporate elements of design in fundamental engineering courses. Engel earned a BS in engineering science at Penn State and PhD in engineering mechanics at the University of South Florida. She can be contacted at rse1@psu.edu.

Studying the Reliability and Validity of Test Scores for Mathematical and Spatial Reasoning Tasks for Engineering Students

Background and Motivation

The purpose of this paper is to summarize the reliability and validity of scores for several mathematical and spatial-reasoning constructs that are keys to academic success in engineering. In this study, we examine the scores for two mathematical constructs and two spatial-reasoning constructs. The mathematical constructs represent students' abilities to: M1 compare and contrast mathematical operations (e.g., differentiation, integration, interpolation); and M2 represent engineering- and physics-based principles mathematically.

The two spatial reasoning constructs focus on students' skills at: S1 rotating and transforming geometric objects in three-dimensional space; and S2 translating two-dimensional images to three-dimensional images and vice versa when representing visually physics-based principles such as acceleration, equilibrium, and force.

A student's understanding of mathematical concepts and the physical interpretation of these concepts is essential in engineering courses. Learning and developing these skills, however, is often not taught in math courses and is taken as a prerequisite in engineering courses. Astatke et al.[1] investigated how a physical understanding of mathematics can be taught to students in a pre-calculus course. Cardella and Atman[2] have studied how engineering students use mathematics in an engineering capstone design course. Students in the study used mathematics as a tool, as a problem solving method, and also as a way to describe physical problems.

Spatial-reasoning measures have also received attention in the literature because of the importance in determining academic success in engineering. Devon, Engel, and collegues[3,4] determined that the students' ability to rotate and transform geometric objects in three-dimensional space is related to graduation and retention patterns in engineering programs. Sorby[5] has developed assessment tools and a training course to improve students' three-dimensional spatial skills. Similarly, knowing how forces are represented visually in diagrams is a skill that successful engineering students have. However, many college students have difficulty understanding how physics-based principles are represented visually. As a result, the types of problems assigned in courses like statics and thermodynamics that utilize these visual representations may be one reason these classes are perceived as difficult. Wai et al.[6] provide evidence that spatial ability is also important in other STEM (science, technology, engineering and mathematics) disciplines.

The challenge that students encounter in engineering courses potentially increases as no construct, skill, or strategy operates in isolation of one another. Research from cognitive psychology[7,8] provides ample evidence that constructs must be coordinated or integrated if students are to reach levels of competence or proficiency within their domain. Therefore, in this article, we attempt to show how to build instruments that can measure multiple mathematical and spatial reasoning constructs that are viewed as important in the domain of engineering. First, we provide definitions and example items for each mathematical and spatial reasoning variable of interest. Second, we summarize preliminary information about the reliability of scores for each scale we have developed. Third, we analyze the validity of scores by relating each outcome to well-known academic achievement criteria such as GPA and grades in coursework. Fourth, we examine a preliminary factor structure of all constructs to examine the interplay among dimensions given item responses. Finally, we also test whether means for the four constructs are similar by major area of study in engineering, and we report the overall reliability of the items for the instrument.

This paper presents the results from a survey completed by a sample of 83 students enrolled in a Thermal Science course. Most students completing the survey had declared their major area of study in civil engineering (n = 43). Other programs of study represented included: a) aerospace engineering; b) architectural engineering; c) electrical engineering, and, d) industrial engineering. For students in aerospace engineering and architectural engineering, the Thermal Science course is a required course taken in fourth semester. For students in civil engineering, this course is taken in sixth semester. The Thermal Science course is used as a technical elective in the electrical and industrial engineering programs. Most students were enrolled in a dynamics course at the time they completed the survey. Most students had enrolled previously and completed coursework in both Engineering Design as well as Calculus and Analytic Geometry. Reported grades for these courses were also available for analysis along with self-reported Grade Point Average (GPA).

Mathematical Test Items: Constructs M1 and M2

The use of mathematics in solving and communicating engineering analysis can be an obstacle for some students. In describing the use of mathematics in engineering, we have distinguished between two different constructs, listed above as:

M1 compare and contrast mathematical applications relevant to solving varied problems in engineering;

M2 understand how the engineering quantities (e.g. force, work, power, and flow rate) are described by the mathematical representations (e.g. integration, differentiation, or interpolation) presented in statics, dynamics, thermodynamics, and fluid mechanics.

Although these two constructs are similar, we have listed them separately to better define the particular usage of mathematics that a student finds challenging. The following two examples will better define these constructs.

Mathematical Test Items: Construct M1

Construct M1 refers to an understanding of the mathematical equations and solution methods without relating it to a physical quantity such as force, pressure, or power. An example of this type of problem is:

M1.3 The function $y = f(x)$ is shown on the graph. Circle all statements below that are true:

a. $\left.\frac{dy}{dx}\right|_1 > \left.\frac{dy}{dx}\right|_2$

b. $\left.\frac{dy}{dx}\right|_1 < \left.\frac{dy}{dx}\right|_2$

c. location 1 is an inflection point

d. $\left.\frac{d^2y}{dx^2}\right|_1 > 0$

e. $\left.\frac{d^2y}{dx^2}\right|_1 < 0$

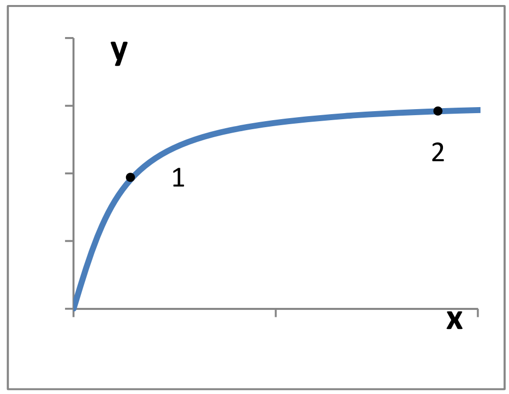

Figure 1. Example of Construct M1.

To answer this question, a student must have an understanding of derivatives but there is no relation to physical quantities. Problems of this type can also be presented using different variables, say (y,T) instead of (x,y). Although the problem still uses variables with no physical interpretation, some students will find the second problem to be much more difficult because textbooks and instructors in calculus classes use (x,y) in most if not all problems. This finding might lead us to change the variable names throughout a calculus course and not always use (x,y).

Reliability Results for Construct M1

In total, 11 items were designed to measure this construct. Scores were assigned dichotomously as correct =1 or incorrect = 0 for each item. To estimate the reliability of scores, we used the the Kuder-Richardson 20 (KR20) procedure, which is an index of internal consistency[9]. The KR-20 estimate for the scores was .61 after one item was deleted because the item-total correlation coefficient was negative. For a 10-item scale, this reliability estimate seems acceptable.

Validity Results for Construct M1

To examine the validity of scores, we studied patterns of association with the three academic achievement external criteria: a) reported grades in Engineering Design; b) reported grades in Analytic Geometry and Calculus; and, c) reported GPA. We hypothesized that scores for Construct M1 should be significantly related to both overall GPA and grades in the mathematics course. There was some support for this hypothesis. Since the distributions for M1 scores and GPA were approximately normal, we used the Pearson correlation coefficient to study the association between scores. The Pearson correlation coefficient between M1 scores and GPA was +.33. Grades in the Engineering Design and Analytic Geometry and Calculus courses were reported as letter grades, thus these course grades were analyzed as ordinal scales of measurement. Therefore, we studied the relations between M1 scores and these ordinal grade scales using the Spearman Rank correlation coefficient. The Spearman Rank correlation coefficient between M1 scores and mathematics grades was +.31. The Spearman Rank correlation coefficient between M1 scores and Engineering Design grades was .21.

Mathematical Test Items: Construct M2

The second use of mathematics tested is Construct M2 that applies a physical meaning to the variables in the equation.

M2.5 If *h* represents the height of water in a tank and t represents time, what does the following equation tell you about the height of the water in the tank?

$$\frac{dh}{dt} = -5$$

a. The height of the water is negative.
b. The height of the water does not change with time.
c. The height of the water is increasing with time.
d. The height of the water is decreasing with time.
e. Insufficient information given to answer the question.

Figure 2. Example of Construct M2.

Figure 2 is an example of Construct M2. This question has added a physical meaning to each variable and asks for a physical interpretation of the differential equation. To answer this question requires several skills: understanding the definition of a derivative, using variables (t,h) instead of (x,y), and relating this equation to a physical process. For an unsteady problem such as this, the physical process cannot be easily communicated using a figure drawn on paper. Students must be able to mentally visualize a "movie" to understand the problem. Similar complications occur for three-dimensional problems that are shown as a two-dimensional representation. Another example of Construct M2 is shown later in this paper as Figure 5.

Reliability Results for Construct M2

We used KR20[9] to estimate the reliability of scores for Construct M2. Six items comprised this scale, and the reliability coefficient was .55. No items had to be deleted given inspection of the item-total correlation coefficients. While only six items, the item difficulty estimates ranged from .13 to .94. We find these initial reliability statistics very good given the short scale.

Validity Results for Construct M2

No correlation coefficient was greater than .11 when studying Construct M2 with the external criteria. This pattern of correlation coefficients suggests that Construct M2 is representing a latent trait that may not be reflected in engineering coursework. Further, the pattern of results is different than that observed for Construct M1. We did compute the Pearson correlation coefficient for scores representing both M1 and M2. The degree of association was .44. While this value does support a significant relationship, descriptively it is only moderate. As such, it appears that there is evidence to support the hypothesis that Construct M1 is a different type of mathematical trait than Construct M2.

Spatial Reasoning Items: Constructs S1 and S2

Two spatial reasoning constructs are important in engineering education:

Construct S1 involves the ability to rotate and transform geometric objects in three-dimensional space. Similar to the Construct M1 in mathematics, this spatial reasoning can be perceived as a general one that does not include reference to specific engineering- or physics-based principles. Yet, the literature documents clearly[4] that students who solve problems well in engineering have strong general spatial reasoning strategies.

Construct S2 requires translation of two-dimensional images to three-dimensional and vice versa when solving engineering problems. This construct includes the interpretation of figures, diagrams, and word descriptions that represent engineering- or physics-based principles. There are two different skills that are included in this construct:

1. Three-view two-dimensional projection drawing to a three-dimensional perspective drawing.
2. Relating different visual and mathematical representations of unseen quantities such as velocity, force, pressure, or temperature.

Spatial Reasoning Items: Construct S1

An example of Construct S1 is shown below in Figure 3. This figure was used with permission from a Mental Rotation Test developed by Devon et al.[3,4] Three items from that test were included in the current investigation.

S1.2 Which figure below is a rotation of the first?

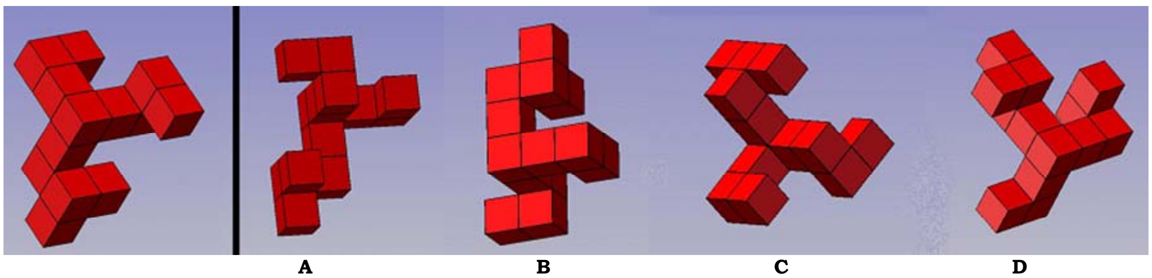

Figure 3. Example of 3D rotation, Construct S1.

Reliability Results for Construct S1

The reliability estimate for Construct (S1) was only .33; however, the scale included only three items. The item difficulty estimates were: .48, .66, and .94. Based on these results, we will include more items on the survey in future studies.

Validity Results for Construct S1

No correlation coefficient was greater than .17 when studying the associations among Construct (S1) scores and the external criteria. This pattern of results suggests that the type of spatial reasoning measured by the scale is not related to variables that lead to assignment of grades in engineering coursework.

Spatial Reasoning Items: Construct S2

Engineering also includes the analysis and interpretation of unseen quantities such as velocity, force, pressure, and temperature. Engineers often describe unseen quantities visually in graphs and figures. Students sometimes have difficulty in interpreting these graphs and figures, sometimes considering both coordinates as spatial coordinates and the plotted curve as a physical line or boundary. When the quantity is plotted using a cross-section of the geometry, the spatial visualization also presents a challenge. Figure 4 presents an example of the laminar velocity profile in a pipe presented in three different ways: using a velocity profile, surface contour, and uniform velocity contours. Each representation includes two different answers. In each row of answers the student needs to decide if the first, the second, or neither of the figures describes the given velocity profile.   These types of representations are used in many engineering courses.

S2.4. Mark all figures that are a visual representation of $u(r) = 1 - r^2$ where r is the radial coordinate and u is the velocity.

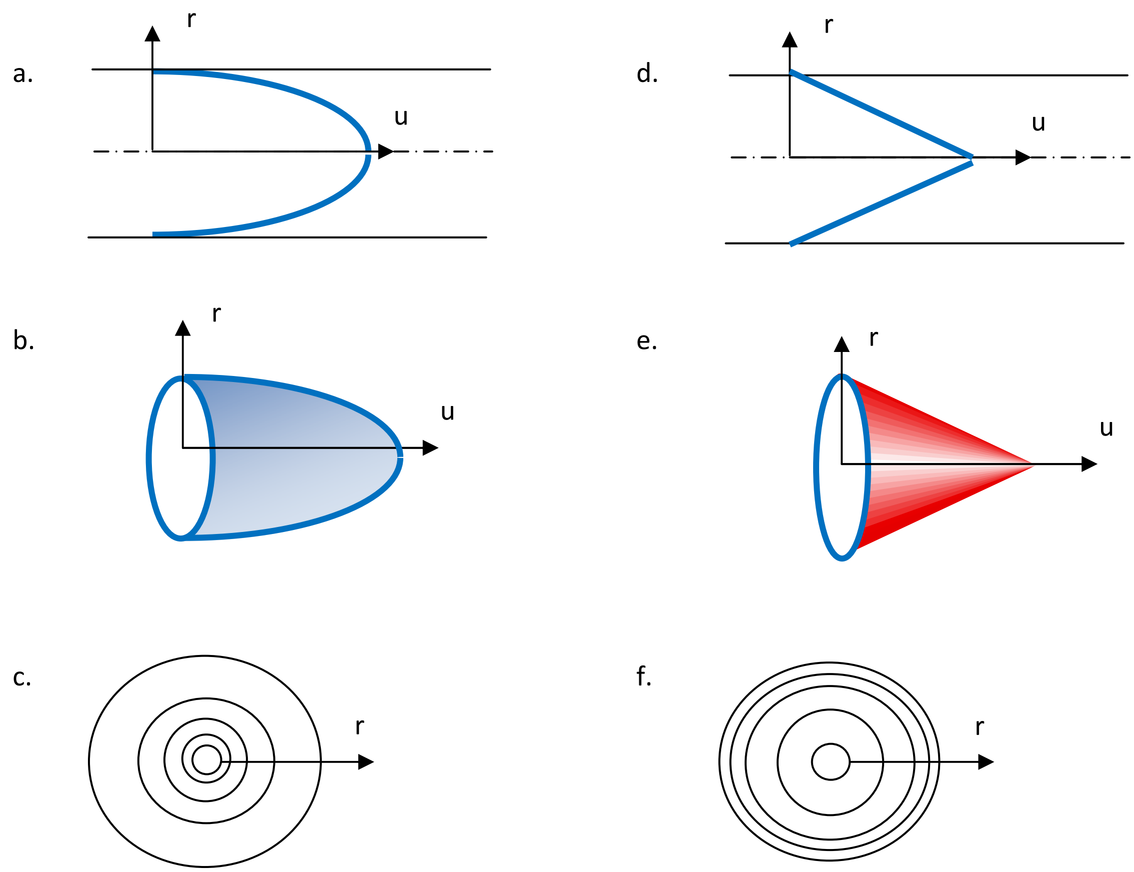

Figure 4. Example of various representations for Construct S2.

Reliability Results for Construct S2

Four items comprised the scale for Construct S2. The KR20 estimate was .50, which we interpreted as a very good reliability coefficient given the short scale. Item difficulty values were in the range of .27 to .81.

Validity Results for Construct S2

To a small degree, scores for Construct S2 were associated with reported grades in Analytic Geometry and Calculus, +.25. Also, the correlation coefficients between Construct S2 scores and mathematical construct scores (M1), +.51, and (M2), +.40, were not only significant, but also these indices of association were greater than that computed between Construct S1 and Construct S2 scores, +.24. From these results, we conclude that there is likely an important variable interplay between this type of spatial reasoning and mathematical performance required to be successful in engineering courses.

Factor Analysis of Four Constructs

Therefore to explore further how the four constructs, M1, M2, S1, and S2 were related, we ran a factor analysis using the 23 item scores (i.e., M1 = 10 items; M2 = 6 items; S1 = 3 items, S2 = 4 items, Total Items = 23). While the sample size (n = 83) is considered small for a factor analysis given the frequency of items[10,11], we thought that study of the factor loadings and cross loadings might help us describe some of the patterns in the relations between constructs for this particular sample of participants. In particular, M1 and S2, were the most strongly correlated pair given the four sets of scores for this sample.

Orthogonal and oblique solutions were compared based on a principal axis factor analysis. The orthogonal solution results in a factor structure where the four constructs are statistically independent of one another while the oblique solution summarizes the factor correlations between constructs[10,11], Therefore, in the oblique solution, the correlation coefficients between factors may be sufficiently large to suggest that these factors can be combined.

A comparison of the orthogonal and oblique factor solutions demonstrated that both factor loadings and cross-loadings for a four-factor solution were comparable. Since only one cross-loading was greater than .40 for either solution, the orthogonal solution was interpreted as it is much easier to describe than a oblique solution[11]. The four-factor solution explained 29.44 percent of the variance in scores. Eigenvalues were: a) 3.34 for Factor 1; b) 1.42 for Factor 2; c) 1.20 for Factor 3; and d) .82 for Factor 4.

Table 1 presents the orthogonal solution for four factors. Only the factor loadings are presented in the table given the orthogonal solution. As such, the cross loadings are not included. The items are listed for each factor given the size of the loadings. Eleven items loaded onto Factor 1. The first six of these items represented 2 M1 items, 1 M2 item, and 3 of the 4 S2 items. All of these loadings were greater than .40, which implies that those item scores contribute to defining the construct or the latent trait. Inspection of the set of items demonstrates that many of these items focused on concepts related to acceleration, motion, and rates of change. The loadings for this factor also show that this construct or latent trait represents a combination of both M1 and S2 skills. Thus, the pattern of loadings also helps to describe why the M1 and S2 total scores were moderately correlated.

Construct	Factor1	Factor2	Factor3	Factor4
S2 (item4)	.69			
M1 (item3)	.56			
M1 (item10)	.55			
M2 (item4)	.48			
S2 (item3)	.47			
S2 (item1)	.42			
M1 (item8)	.33			
M1 (item9)	.30			
M1 (item1)	.30			
S1 (item2)	.27			
M1 (item2)	.19			
M1 (item7)		.53		
M2 (item3)		.44		
M1 (item4)		.42		
M1 (item6)		.36		
M2 (item1)			.82	
M2 (item2)			.63	
M2 (item3)				.61
M1 (item11)				.55
M2 (item6)				.40
S2 (item2)				.39
S1 (item1)				.32
S1 (item3)				.25

Table 1. Orthogonal factor loadings for M1, M2, S1, and S2 items.

Factor 2 includes 3 M1 items and 1 M2 item. All of these items appear to test knowledge of functions or algebraic expressions.

Factor 3 includes two of the M2 items, and these items also focus on algebraic expressions; however, the stem of both problems introduces a visual representation that illustrates the distribution of pressure on surfaces. One of these problems is shown in Figure 5. As such, the stem of each item requires understanding of physics principles, while the option set includes possible algebraic equations that reflect the distribution of pressure on surfaces given arrows presented in each stimulus diagram.

M2.2. Pressure *P(x,y)* varies on the surface of a rectangular plate as shown, where (*x,y*) are coordinate distances from the lower left hand corner of the plate. *a* and *b* are positive constants with appropriate units.

Which function describes the distribution of pressure on the plate?

a. $P(x, y) = ax + b$

b. $P(x, y) = -ax + b$

c. $P(x, y) = ay + bx$

d. $P(x, y) = ay + b$

e. None of the above

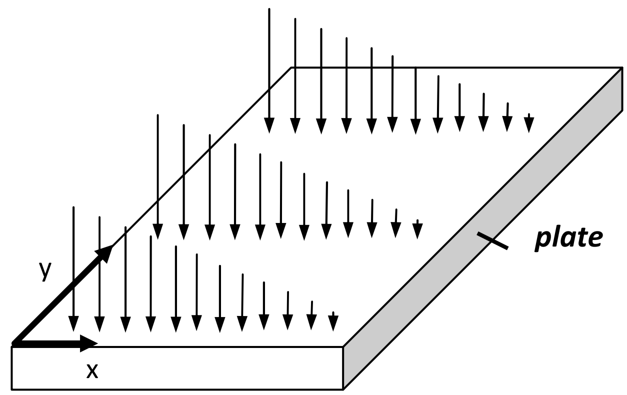

Figure 5. Example of pressure visualization question for Construct M2.

Factor 4 also includes two of the M2 items along with 2 of 3 of the S1 items. This set of items appears to require the skill to understand derivatives given real-world problem-solving scenarios in physics. As such, the representations included in both the stem and options of the items for Factor 4 are different when compared to those representations of the stems and items of Factor 3. However, selecting correct responses to both sets of M2 items does require mapping understanding of physics principles to mathematical expressions of these principles. It is also important to note that 2 of the 3 S1 items are the last to be summarized in the factor analysis with loadings on Factor 4 of .32 and .25, respectively. Again, because of the small sample size, results should be described cautiously. However, the results do imply two things: a) S1 is a very different type of construct than M1, M2, and S2; and, b) more S1 items should be included on the test as the sample size is also increased to represent more sets of responses provided by students.

Differences Among Major Areas of Study and Overall Reliability

Finally, we compared the means for the M1, M2, S1, and S2 scores by major area of study. Specifically, we analyzed mean differences for students whose major was in civil engineering (n = 43) compared to all other major areas of specialization (n = 40). Results of independent t-tests for comparison of two groups showed no significant differences, t's < 1.65, p's > .10. These results indicate that scores for M1, M2, S1, and S2 were similar for all major areas of study represented in this investigation. Additionally, and as one final estimate of the reliability of scores, we analyzed the KR20 for the complete set of 23 items summarized in the factor analysis. The KR20 was .76, which suggests there was high internal consistency given the complete scale. Given that the factor structure demonstrated that, with the exception of Factor 3 (i.e., M2: pressure on surfaces of a plate), that each factor represented combinations of M1, M2, S1, and S2 items, future research should explore whether there is possibly one second-order mathematical-spatial reasoning trait that is related to various first-order constructs that represent different kinds of mathematical and spatial reasoning skill given the types of representation (e.g., equations, graphs) and physics concepts (e.g., acceleration and motion vs. forces and pressure) in both the stems and options of the items. A significantly greater sample size would be needed to conduct this analysis than the sample size for which we could summarize results in this manuscript; however, descriptive patterns do suggest that results may be related to how various external representations are mapped. Research in cognitive and educational psychology by Ainsworth and her colleagues[12,13] would also indicate that the line of inquiry on how students use multiple external representations is one to consider as we plan additional investigations in our program of research.

Summary

While preliminary, we interpret our results as encouraging as we attempt to measure four different constructs that we believe are keys to academic success in engineering. Two constructs represent mathematical skills or traits. Two constructs represent spatial reasoning strategies or traits. Correlation coefficients indicated there were some moderate associations between pairs of constructs (e.g., M1 and S2), and a preliminary factor analysis showed that part of the relation between constructs may be related to the types of external representations (e.g., diagrams, equations, graphs) presented in the test items or the primary physics concepts (e.g. acceleration, force, motion) for which the items have been designed to assess achievement in engineering. In the future, we will plan studies so that more items are included on each scale. Further, we will attempt to replicate the initial findings reported in this manuscript. If correlation coefficients between scores are at best moderate, then this pattern of results has important implications for assessment, teaching, and research in engineering education. Simply, it implies that the skill set required to succeed in engineering may be multidimensional. As such, a set of various tasks or tests are needed to help students understand their profiles of strengths and weaknesses as they attempt to meet degree requirements in this, sometimes, challenging field.

In future research, it might be important to study how students think aloud as they solve problems not only within one construct set, but also, how they compare and contrast the problems they attempt to solve across sets. A mixed-methodological approach where we study both test scores with verbal reports might help us understand more about the complex processing required to complete exercises successfully. Further, verbal protocol summaries might reveal the role of one or more external representations given how students interpret problems in engineering and the processes they use to select or construct solutions for these problems.

References

[1]Astatke, Y., Leigh-Mack, P., Alao, S., Craig J. Scott, C.J., Davis, A., Davy, A.B., Njage, E., Salimian, M., (2005) “Enhancing Conceptual Transfer of Mathematical Concepts and Motivating Students using Engineering Performance Tasks”, 2005 ASEE Annual Conference & Exposition, June 12-15, Portland, OR.

[2]Cardella, M., Atman, C. (2007), “Engineering Students’ Mathematical Thinking: In the Wild and with a Lab-Based Task,” 2007 ASEE Annual Conference & Exposition, June 24-27, Honolulu, HW, Paper 2007-2853.

[3]Devon, R., R.S. Engel, R.J. Foster, D. Sathianathan, G.F.W. Turner, (1994), “The Effect of Solid Modeling Software on 3-D Visualization Skills,” *Engineering Design Graphics Journal* **58(2):**4-11.

[4]Devon, R., R. Engel, G. Turner, (1998), “The Effects of Spatial Visualization Skill Training on Gender and Retention in Engineering,” *Journal of Women and Minorities in Science and Engineering* **4:** 371-380.

[5]Sorby, S., (2009), "Educational Research in Developing 3-D Spatial Skills for Engineering Students," *International Journal of Science Education*, **31(3):** 459 – 480.

[6]Wai, J., Lubinski, D., and Benbow, C. P., (2009), Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. Journal of Educational Psychology, **101:** 817-835.

[7]Alexander, P.A. (2003), “The development of expertise: The journey from acclimation to proficiency,” *Educational Researcher* **32**: 10-14.

[8]Kulikowich, J. M. (2007), “Toward developmental trajectories: A commentary on assessing measures of mathematical knowledge for teaching,” *Measurement: Interdisciplinary Research and Perspectives* **5(2-3)**: 187-194.

[9]Kaplan, R. M., D.P. Sacuzzo, (2009, 7th ed.). *Psychological testing: Principles, applications, and issues*. Belmont, CA: Wadsworth.

[10]Gorsuch, R. L., (1983, 2nd ed.). *Factor Analysis*. Hillsdale, NJ: Erlbaum.

[11] Costello, A. B., J. Osborne (2005). “Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis,” *Practical Assessment Research & Evaluation*, 10(7): Available online: http://pareonline.net/getvn.asp?v=10&n=7

[12] Ainsworth, S. (2008), “The educational value of multiple representations when learning complex scientific concepts,”. In J. K. Gilbert & M. Reiner & M. Nakhlel (Eds.), Visualization: Theory and Practice in Science Education (pp. 191-208)New York: Springer.

[13] Ainsworth, S.E (2006), “DeFT: A conceptual framework for learning with multiple representations Learning and Instruction,” 16(3): 183-198.