International Journal of Engineering Education Vol. 29, No. 3, pp. 763–776, 2013
Printed in Great Britain

0949-149X/91 $3.00+0.00

Psychometric Properties of the Revised PSVT:R for Measuring First Year Engineering Students' Spatial Ability*

YUKIKO MAEDA**
Department of Educational Studies, Purdue University, 100 N University St., West Lafayette, IN, 47907, USA.
E-mail: ymaeda@purdue.edu

SO YOON YOON
INSPIRE, School of Engineering Education, Purdue University, West Lafayette, IN, 47907, USA. E-mail: yoon18@purdue.edu

GYENAM KIM-KANG
Measurement and Statistics Research, 3415 Boone St.,West Lafayette, IN 47906, USA. E-mail: gnkang35@yahoo.com

P. K. IMBRIE
Purdue University, 701 West Stadium Ave. West Lafayette, IN 47907, USA. E-mail: imbrie@purdue.edu

While various spatial tests are available, the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R) has been commonly used to predict students' success in the engineering field. While many studies that used the PSVT:R exist, little attention had been given to its psychometric properties in measuring spatial ability and relationships to other academic indices. The purposes of this study were (a) to characterize the item- and test-level functions of the Revised PSVT:R for the use of incoming First Year Engineering (FYE) students, and (b) to investigate its relationship to academic-related variables to provide validity evidence. Approximately 2400 FYE students enrolled in the fall of 2010 and 2011 in a large Midwestern public university completed the Revised PSVT:R. Students' academic-related variables were also retrieved from the university archive. A variety of statistical analyses, including exploratory and confirmatory factor analyses as well as item analyses, were conducted on the Revised PSVT:R scores. Pearson's product–moment correlation coefficients between the Revised PSVT:R and other academic variables were also obtained. The Revised PSVT:R measures a unidimensional subcomponent of spatial ability. Cronbach's α was 0.84. Items were relatively easy and the test provides the most precise estimate for students whose ability level is at or below average. Weak to moderate correlations were found between the Revised PSVT:R scores and the aptitude test scores. The Revised PSVT:R is a psychometrically sound instrument. However, items are relatively easy, but it is still appropriate to measure spatial visualization ability of the FYE students.

Keywords: first year engineering students; Revised PSVT:R; mental rotation; psychometric properties

1. Introduction

1.1 Importance of spatial ability for success in engineering programs

Spatial ability has received widespread attention since the early 1900s because of its link with academic and vocational success [1]. While several attempts have been made to define this construct, Lohman stated that: "It [spatial ability] is not a unitary construct. There are, in fact, several spatial abilities, each emphasizing different aspects of the process of image generation, storage, retrieval, and transformation" [2, p. 4]. For example, Michael *et al.* [3] proposed a three-factor model of spatial ability: spatial visualization (SV), spatial relations and orientation, and kinesthetic imagery. Later, McGee [4] suggested a two-factor model with SV and spatial orientation (SO). Lohman [5] agreed with the basic components of McGee's model of spatial ability, but he extended it further by distinguishing spatial relations (SR) from SO. Divergently, Carroll [6] characterized the dimensions in spatial ability broadly by including other peripheral components, such as closure speed, flexibility of closure, perceptual speed, and visual memory. Although there is no consensus of dimensions in spatial ability, researchers seems to agree with at least two core components of spatial ability: spatial visualization (SV) and spatial relation/orientation (SR/SO).

Despite the fact that the notion of spatial ability varies across studies and/or the ability is often loosely defined, researchers generally agree that spatial ability plays a crucial role in determining students' achievement in engineering courses, in particular graphic and design courses [7–10]. For example, Baartmans [7] found that students' scores on a spatial ability test were the most powerful

** Corresponding author.

* Accepted 29 January 2013.

predictors for their success in an engineering design graphics course among 11 variables, including demographic and academic aptitude variables, such as SAT and ACT scores. Field [9] reported that the use of spatial ability scores as well as mathematics course grades improved the prediction of performance in undergraduate engineering design courses compared with the use of mathematics GPA as a single predictor. Recently, the demand for spatial ability in engineering has increased as technology like Computer-Aided Design (CAD) application has been introduced in graphic design; a traditionally required skill for designing and drawing in 2-D space is now replaced by designing and manipulating objects in 3-D space using virtual models [8]. Therefore, spatial ability has received more attention than ever for its role in predicting the academic success of students.

First Year Engineering (FYE) programs provide opportunities for students seeking degrees in engineering to learn and develop fundamental knowledge by taking basic courses in STEM areas (such as calculus, physics, computer programming, and graphic design courses). Therefore, some institutions require that students successfully complete the FYE program in order to proceed in their choice of engineering major. The 17-year longitudinal study by Budny et al. [11], conducted at a large Midwestern public university, indicates one of the important roles of the FYE program. The researchers reported that 97% of freshmen who successfully completed the first year requirements in the FYE program graduated from the university, mostly with baccalaureate degrees in engineering (i.e., 89%). This indicates that successful completion of the FYE program predicts students' retention in professional schools, and their chance of graduating with a college degree. Together, these results imply that providing appropriate remedial instruction for students with lower spatial ability may help those students to be better prepared to take fundamental STEM courses in the FYE program, which may result in the successful completion of their FYE program and subsequently increase their chance of graduating from college.

In fact, Sorby and her colleagues [e.g., 12–16] reported the positive impact of interventions on spatial ability for improving spatial ability performance in engineering graphic design courses at Michigan Technological University. In their studies, the FYE students were measured for their spatial ability and then placed into the graphic courses according to their ability level. Students with low spatial ability were encouraged to take a remedial course in cultivating 3-D spatial visualization. After taking the remedial course, students increased their ability and later successfully completed graphic-related courses [12, 16]. In addition, Sorby and Baartmans [15] suggested a link between positive outcomes through the intervention and an increase in student retention in the engineering program. Other studies also supported the positive relationship [e.g., 17].

1.2 Spatial ability tests frequently used in engineering education

Given evidence of the positive correlation between spatial ability and academic success in engineering, the spatial ability of FYE students was often measured (a) to predict their performance in engineering courses [e.g., 9, 18, 19] and (b) to identify students who may benefit from participating in a remedial intervention program [13–15]. For these purposes, selecting an instrument with sound psychometric properties is a critical first step to providing useful information to support students towards their academic success.

Various tests for measuring spatial ability are currently available, partly because there is no unitary definition of spatial ability; rather, spatial ability is often defined in terms of several subcomponents [2, 20]. These include the Mental Cutting Test (MCT) [21], the Mental Rotations Test (MRT) [22], the Revised Minnesota Paper Form Board Test (RMPFBT) [23], the Differential Aptitude Tests: Spatial Relations (DAT:SR) [24], and the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R, see Fig. 1 for a sample item from each spatial test). These tests have been used frequently in research on success in engineering programs because the particular spatial ability measured by these tests seems to be tightly connected to the engineering profession [5, 25]. Each test is briefly described below.

The MCT aims to measure an individual's spatial visualization ability. For the test, an individual has to solve 25 items of 3-D objects with a cutting plane, indicating where they should be cut through. The individual's task is to imagine the trace of that cutting plane and select the correct image from the possible options [21]. The MRT is a paper-and-pencil based test that consists of 20 items to measure spatial visualization by mentally rotating 3-D objects [22]. The RMPFBT is another popular spatial visualization test originally developed in 1920s [23] to "measure aspects of mechanical ability requiring the capacity to visualize and manipulate objects in space" [26]. Respondents are asked to mentally assemble the presented pieces of a certain geometric figure and discern the correct figures from the answer choices. Although the test contains 64 multiple choice items with 2-D objects, solving a problem in this test does not involve the mental rotation of objects.

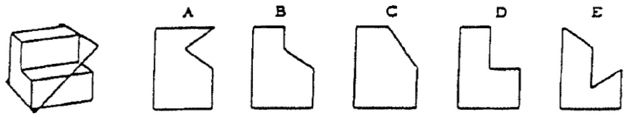
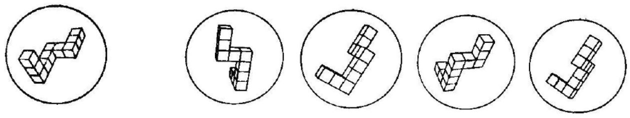
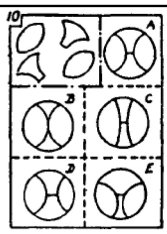
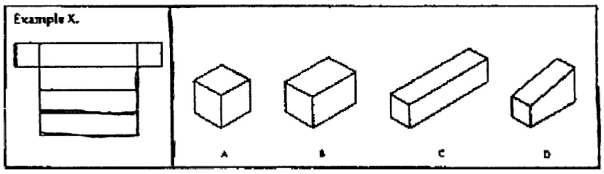
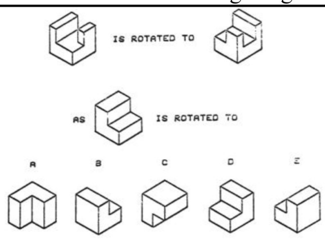
Spatial test	Sample item
Mental Cutting Test [21]	 Direction: Choose the resulting cross-section from the cut of the 3-D object.
Mental Rotations Test [22]	 Direction: Choose the two figures that are identical to the one on the far left.
Revised Minnesota Paper Form Board Test [23]	 Direction: Choose the figure that displays the pieces joined together.
Differential Aptitude Test: Spatial Relation [24]	 Direction: Choose the correct 3-D object from the four choices that would result from folding the given 2-D pattern.
Purdue Spatial Visualization Tests: Visualization of Rotations [27]	 Direction: Choose the object that has the same rotation as shown in the top line.

Fig. 1. Sample items from the popular spatial tests used in engineering education.

The DAT:SR is a subtest of a multiple aptitude battery (eight subtests) that requires examinees to indicate what an unfolded shape would look like when folded [24]. Finally, the PSVT:R is a spatial visualization test involving the mental rotation of 3-D objects [27]. The PSVT:R has been used primarily in research on educational settings in science, technology, engineering, and mathematics (STEM) disciplines for more than three decades. The test has been recognized as one of the most popular tests to measure the ability of students to spatially visualize the mental rotation in engineering education [8, 9]. This is partly because, compared with other popular spatial tests for research in engineering education, the PSVT:R is unique in that it includes a variety of 3-D objects (including objects with inclined, oblique, and/or curved surfaces), and it requires a higher level of spatial visualization ability [28]. A further description of the PSVT:R is provided in the next section.

1.3 Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:R)

Guay [27] developed the Purdue Spatial Visualization Test (PSVT), which consists of three 12-item subtests entitled *Developments*, *Rotations*, and *Views*, respectively. The PSVT:R is an extended version of the subtest, *Rotations*, to measure the 3-

D mental rotation ability of individuals aged 13 or above in 20 minutes [29]. The PSVT:R has 30 items consisting of 13 symmetrical and 17 nonsymmetrical 3-D objects that are drawn in a 2-D isometric format. In each item, the respondents' task is to mentally rotate an object in the same direction as indicated visually in the instructions, and then to select an answer from among five possible options. While there have been many studies that use the PSVT:R, little attention has been paid to the accuracy of measuring spatial ability with the test. Recently, Yue [30, 31] identified ten figural errors, such as missing lines of a rotated object, on seven of 30 items of the PSVT:R. As a result, the test was revised by Yoon [32], with Guay's permission, to eliminate all figural representation errors on items. In addition, the format of the instrument was modified to avoid possible measurement errors due to crude item presentation: the revised version has one item per page instead of two items per page to avoid any distractions to the respondent (i.e. being distracted by features of another item on the same page). In addition, if figures had different scales, they were rescaled. For interested readers, see Yoon [32] for detailed information about the revision process. The revised version of the PSVT:R is distinguished from the original one in this study by referring to it as the Revised PSVT:R.

The purpose of the study was twofold: to investigate and characterize the psychometric functions of the Revised PSVT:R for incoming FYE students and to report its relationship with academic-related variables. The specific research questions to meet the purpose were:

1. To what extent is the one-factor model of the theoretical construct measured by the Revised PSVT:R appropriate?
2. How reliable are scores on the Revised PSVT:R for indicating the spatial ability of FYE students?
3. To what extent do characteristics, such as item difficulty and item discrimination, vary across the items in the Revised PSVT:R?
4. To what extent does the Revised PSVT: R relate to academic-related variables?

2. Methods

2.1 Data source

The target population of this study includes all engineering freshmen enrolled in the First Year Engineering (FYE) Program in a large, Midwestern public university in the United States in the fall of 2010 and 2011. The acceptance criteria to enter the FYE program are the same as the university's general freshman admission criteria. In other words, students are accepted if they provide evidence of high academic performance and demonstrate that their academic aspirations are aligned to a designated program. Students who were accepted by the engineering program were invited to participate in web-based assessments by the School of Engineering Education prior to their entrance to the FYE program during the summers of 2010 and 2011. The Revised PSVT:R was administered as a part of the online assessments. For this study, we retrieved data of the Revised PSVT:R scores gathered from 2,469 FYE students, as well as relevant academic and demographic information of those students from the university archive. Of the 2,469 students, 1,888 (76.5%) were male, 580 (23.5%) were female, and one student did not report his/her gender information. Their age ranged from 15 to 38 years old, with an average age of 18.01. However, 97% of the students were in the range of 17 to 19 years old.

2.2 Data

The Revised PSVT:R responses. The Revised PSVT:R was administered online and respondents were given a maximum of 25 minutes to answer the 30 multiple choice items in the test. We considered the 25-minute time frame to be sufficient to control the impact of problem solving speed on their scores because Yoon and Mann [33] found that more than 95% of undergraduate students from various majors could complete the test within 25 minutes when no time limit was introduced. Individual's raw response on each of the 30 items was recoded as a dichotomous variable (1 for correct, 0 for incorrect) for the proceeding analysis. A raw total score was also computed by counting the number of correct responses among the 30 items.

Academic-related performance and demographics. The FYE students' academic aptitude test scores, such as SAT and/or ACT mathematics as well as composite scores, ACT science scores, and high school core and overall GPAs were retrieved from the university archive and used to evaluate the criterion-related validity of the Revised PSVT:R. The web-based assessment battery also contained several questions related to students' demographic backgrounds, including gender. The demographic information was used only for preliminary analyses.

2.3 Data analyses

As preliminary analyses, a series of descriptive analyses and an exploratory factor analysis (EFA) were first conducted on the Revised PSVT:R raw scores. Means and standard deviations of the Revised PSVT:R total scores were computed with participants as a whole and by cohort to examine if there is a significant cohort effect. Differences in mean scores and standard deviations between the

2010 and 2011 cohorts (M = 22.87, SD = 5.20 in Cohort 2010 and M = 22.24, SD = 5.19 in Cohort 2011) were negligible (Cohen's d = 0.12), but statistically significant, $t(2{,}467) = 2.71$, $p = 0.01$. We concluded that the significant differences are due to the large sample size, because there was no specific explanation to suspect a systematic difference in the Revised PSVT:R scores between two cohorts. Therefore, we merged two data sets for the proceeding analysis.

An EFA was performed to investigate whether a single underlying factor structure exists in the Revised PSVT:R, because the test intended to measure one sub-dimension of spatial ability (the spatial visualization ability of mental rotation). The Mplus 6.0 program [34] was used for the analysis because the robust weighted least squares estimator (WLSMV) method for estimation is recommended to handle categorical data in factor analysis [35]. This analysis also served for examining the unidimensionality of the data, which is the major assumption of unidimensional IRT-based analysis [36]. There was a large difference between the estimated eigenvalue of the first factor (9.11) and the eigenvalue of the second factor (2.27), which was represented by the steep slope of the line connecting the first and the second factors. The eigenvalues for the rest of the factors were similar, and the line connecting these factors became horizontal [37]. The results suggest that the Revised PSVT:R has a single factor structure, and the unidimensionality assumption for the IRT (item response theory) analysis was also satisfied.

To address the first question, a confirmatory factor analysis (CFA) was conducted to provide evidence of appropriateness of the assumption on the structure of theoretical construct measured by the Revised PSVT:R. More specifically, the CFA was performed using the Mplus 6.0 program to test whether the one-factor model of the construct fit the sample data obtained from the FYE students. The model fit was evaluated with a Chi-square test as well as multiple CFA fit indices, including Root Mean Square Error of Approximation (RMSEA), Comparative fit index (CFI), and Tucker Lewis index (TLI). Finally, we reported descriptive statistics and frequency distributions of both raw total scores and IRT-based ability scores to summarize the score distribution of the Revised PSVT:R among the FYE students in the sample.

The item analyses applying both classical test theory (CTT) and item response theory (IRT) were conducted to address the second and the third questions. CTT and IRT are the two major frameworks frequently used in measurement research. CTT defines the observed test scores with two components: a person's true ability score and measurement error. This measurement framework has been used in a variety of testing situations because of the simplicity of its theoretical model, weak theoretical assumptions, and the small sample size requirement for applying the framework in practice [38–40]. In the CTT-based framework, item difficulty is defined as the proportion of examinees who successfully answered a particular item. Item discrimination is typically defined as the point-biserial correlation between responses (i.e., correct or incorrect response coded as a dummy variable) on a particular item and the raw total scores. One of the largest drawbacks of applying the framework is that these item statistics depend on the sample characteristics used for the analysis and the examinees' observed scores (e.g., the Revised PSVT:R raw total scores) are also determined by the selection of the items used for testing [36].

IRT is a measurement framework used to estimate the probability of obtaining a correct response on the particular item, given a respondent's ability level, that is independent of both respondents' group characteristics and the items used for testing. In the framework, a family of probability models is mathematically defined with given item parameters, including item difficulty, item discrimination, and/or guessing, depending on the selected IRT model. Although the IRT-based analysis shows some advantages over CTT-based analyses, the application of IRT is generally restricted by its strong theoretical assumptions and large sample size requirement.

Based on the CTT framework, Cronbach's alpha coefficient of internal consistency was computed using IBM-SPSS 18. Item statistics for each item, such as item difficulty and item discrimination were also computed using BILOG-MG 3.0 [41]. Following the CTT-based item analyses, a three- parameter logistic (3-PL) IRT model was used to generate the item parameter estimates, and an individual's ability parameter estimates based on the examinees' response patterns on the Revised PSVT:R. The 3-PL model was a reasonable choice, since it was more realistic to assume the effect of guessing on responses (because the Revised PSVT:R consists of multiple choice items with five alternatives). Moreover, our preliminary analyses with a 1-PL IRT model, as well as the analysis using the CTT framework, have shown that items vary in their difficulties and discrimination powers [42]. In addition, the sample size is large enough to use the 3-PL model. Chi-square item-fit indices were used to evaluate the adequacy of the 3-PL IRT model to the sample data [38].

To address the last research question, we computed Pearson's product–moment correlation coefficients between the Revised PSVT:R total scores and other academic variables (i.e., SAT composite or ACT composite scores, and high school core and overall GPAs). These academic variables were selected because the correlations between these variables and mental rotation ability scores measured by other tests and the original PSVT:R were often reported in empirical studies [e.g., 7], and it relates to our investigation of the similarities of the magnitude of the relationship among these with the Revised PSVT:R.

3. Results

3.1 The structure of the theoretical construct measured by the Revised PSVT:R

A CFA was conducted to test the fit of the data obtained from the Revised PSVT:R to the hypothesized factor structure, a single factor model. As shown in Table 1, factor loadings were reasonably high, ranging from 0.377 to 0.663 across items with $M = 0.528$ and $SD = 0.081$. As expected from the fact that the Chi-square test is well known for its sensitivity to a large sample size [35, 38, 39], the result of the Chi-square test ($\chi^2 = 1623.06$, $df = 405$, $p < 0.001$) was significant. Other fit indices revealed evidence that the single factor model was a good fit to the sample data [35, 43]: the Root Mean Square Error of Approximation (RMSEA) = 0.035, the comparative fit index (CFI) = 0.928, and Tucker Lewis index (TLI) = 0.923. Results from both the EFA and CFA supported the fact that the Revised PSVT:R measures a single factor.

3.2 Item and test characteristics of the Revised PSVT:R

Score reliability. The Cronbach's alpha coefficient of internal consistency for the Revised PSVT:R with the current sample was 0.839, which is considered to be reasonably high, meaning that at least 83.9% of the total score variance is due to true score variance. All 30 items of the test appeared to be worthy of inclusion because the removal of any items did not increase the score reliability.

Table 1. Factor loadings by Confirmatory Factor Analysis

Item	Factor loading	Item	Factor loading	Item	Factor loading
1	0. 398	11	0. 610	21	0. 546
2	0. 414	12	0. 624	22	0. 382
3	0. 377	13	0. 426	23	0. 523
4	0. 452	14	0. 593	24	0. 526
5	0. 502	15	0. 426	25	0. 497
6	0. 605	16	0. 511	26	0. 607
7	0. 553	17	0. 454	27	0. 563
8	0. 564	18	0. 663	28	0. 541
9	0. 630	19	0. 556	29	0. 542
10	0. 632	20	0. 616	30	0. 516

Item characteristics. The CTT-based item difficulty and item discrimination statistics are reported in Table 2. Item difficulties represented by percent correct values ranged from 32.7 to 93.6 with a mean of 74.7. Most of the items were answered correctly by more than half of the students, except for Item 22 (45.6 percent correct responses) and Item 30 (32.7 percent correct responses). These two items are considered to be difficult items for the given sample. On the other hand, Item 4 (93.6 percent correct responses) is considered as the easiest among these 30 items. While there was wide variation in item difficulties across items, the item difficulty overall was considered as appropriate because Lord [44] indicated that, with five-option multiple choice items, the ideal average item difficulty is 70 to maximize the discrimination among respondents.

As also shown in Table 2, CTT-based item discrimination estimates ranged from 0.198 (Item 3) to 0.441(Item 18) with a mean of 0.351, indicating that the discrimination power of the items varies, although all items appeared to contribute to correctly discriminate students for their ability. For example, the correlation was low for the five easiest items (Item 1 through Item 5), meaning that these items are less useful in differentiating students by their levels of spatial visualization ability. This is because almost all students could identify the correct response on these items regardless of their level of spatial visualization ability. For the rest of the items, they discriminated moderately between students with low and high total scores.

Following the item analyses based on the CTT, the item analyses with a 3 parameter logistic (3-PL) IRT model were conducted. The estimates of item discrimination parameter (a), item difficulty (b), and guessing (c) for each item are also reported in Table 2. Item discrimination parameter (a) estimates ranged from 0.74 (Item 1) to 1.842 (Item 26). As consistent with the CTT results, relatively easy items showed low item discrimination power. Item difficulty parameters ranged from –3.128 (Item 3, the easiest item) to 0.966 (Item 30, the most difficult item) with a mean of –0.021, which indicates that the Revised PSVT:R consists of relatively easy items for the population. Note that the easiest item (Item 3) identified by the IRT analysis is different from the one (Item 4) identified by the CTT analysis, while Item 3 was the second easiest according to the CTT analysis. The guessing parameter estimates ranged from 0.064 to 0.260 with an average of 0.186. The estimates are in the expected range because the Revised PSVT:R items are five-option multiple choice items. Interestingly, the hardest item showed the lowest guessing parameter estimate (0.064), which indicates almost no probability of answering the item (Item 30) correctly for students

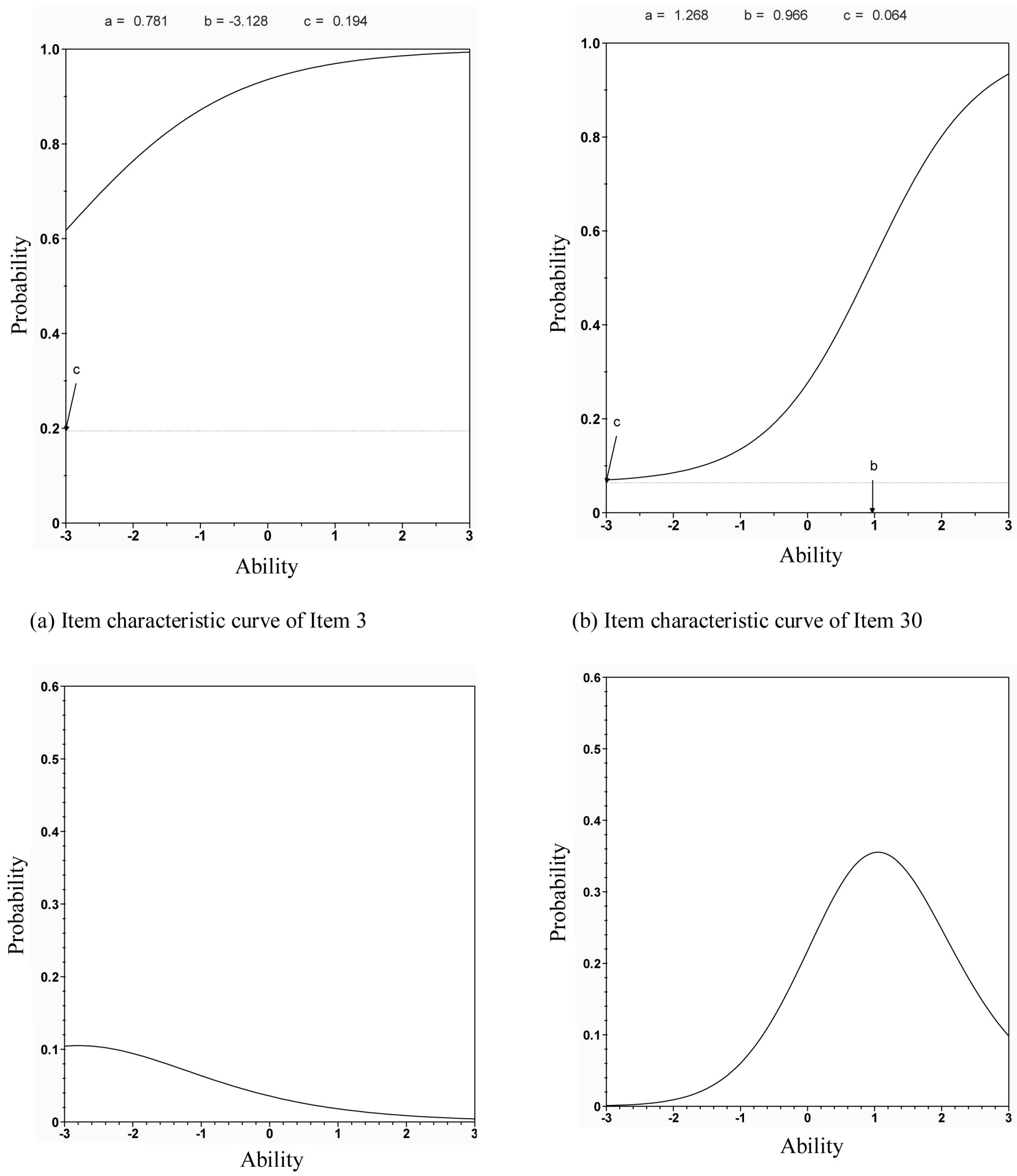
Table 2. Item characteristics based on Classical Test Theory (CTT) and the 3-parameter logistic Item Response Theory Model (IRT)

CTT			**3-PL IRT**			
Item	**Percent correct**	**Item total correlation**	**Item discrimination (a)**	**Item difficulty (b)**	**Guessing (c)**	**Chi-square item-fit (p)**
1	90.8	0.208	0.740	–3.068	0.182	18.0 (0.081)
2	91.0	0.218	0.821	–2.806	0.206	4.5 (0.953)
3	92.0	0.198	0.781	–3.128	0.194	9.6 (0.475)
4	93.6	0.226	0.967	–2.940	0.182	17.3 (0.100)
5	88.4	0.295	1.054	–2.067	0.175	9.5 (0.579)
6	82.2	0.403	1.519	–1.121	0.219	6.9 (0.736)
7	87.2	0.335	1.264	–1.701	0.190	19.5 (0.053)
8	80.6	0.372	1.305	–1.178	0.181	13.3 (0.277)
9	89.6	0.373	1.557	–1.700	0.196	4.8 (0.903)
10	81.0	0.422	1.487	–1.173	0.142	11.7 (0.385)
11	75.3	0.427	1.559	–0.734	0.198	16.4 (0.128)
12	70.6	0.443	1.469	–0.651	0.121	11.2 (0.427)
13	65.4	0.300	0.931	–0.376	0.187	10.1 (0.606)
14	78.3	0.404	1.515	–0.864	0.228	14.8 (0.139)
15	74.0	0.294	0.884	–0.926	0.214	9.8 (0.630)
16	79.1	0.345	1.069	–1.276	0.152	18.3 (0.075)
17	68.9	0.320	0.866	–0.805	0.122	21.8 (0.040)
18	83.3	0.441	1.710	–1.313	0.119	11.9 (0.371)
19	78.0	0.381	1.408	–0.855	0.246	14.1 (0.228)
20	75.9	0.430	1.491	–0.848	0.160	8.3 (0.595)
21	71.7	0.385	1.404	–0.516	0.232	11.1 (0.436)
22	45.6	0.271	1.470	0.893	0.242	16.4 (0.128)
23	70.6	0.367	1.148	–0.695	0.152	11.4 (0.408)
24	77.2	0.357	1.151	–1.069	0.164	7.6 (0.752)
25	67.0	0.354	1.337	–0.209	0.260	13.4 (0.265)
26	64.6	0.439	1.842	–0.159	0.215	15.6 (0.155)
27	63.4	0.404	1.766	–0.029	0.250	29.9 (0.002)
28	68.9	0.385	1.372	–0.408	0.215	9.8 (0.547)
29	55.4	0.390	1.410	0.118	0.157	16.4 (0.127)
30	32.7	0.337	1.268	0.966	0.064	20.0 (0.045)

with low ability levels. Finally, Chi-square item fit indices were also included in Table 2. All but Items 17, 27, and 30 fit the 3-PL IRT model relatively well. Although the significant Chi-square statistics were expected, given the large sample size, the interpretation of the results requires caution.

In Fig. 2, we report the item characteristic curve (ICC) and its corresponding item information function (IIF) for the easiest (Item 3) and the most difficult (Item 30) items to highlight the differences in their item characteristics. The ICC represents the probability of obtaining the correct response by a respondent given his or her ability level. As shown in the ICCs in Fig. 2, the probability of obtaining the correct response has a curvilinear relationship with respondents' ability. The guessing parameter estimate of the easiest item (Item 3) was 0.194, which is the second largest among 30 items. It means that with about 19% of chance, students can get a correct answer on this item with guessing. The item difficulty estimate (b_3) for Item 3 was –3.128, which is the outside of range of the ability continuum in Figure 2(a). The probability of answering the item correctly at the ability of –3.128 was 0.597, which is substantially shifted up from 0.50 because of the influence of guessing. In addition, the item discrimination estimate for the item was 0.781, which is low (Fig. 2(a)). Therefore, the ICC for Item 3 depicts a shallow curve across ability levels, which suggests that the probability of answering Item 3 correct is above 0.6, regardless of student's level of spatial ability, however, the probability increases as the ability level increases and becomes almost 1.0 when the ability level is at 1.0.

In contrast, the ICC of Item 30 shows that the probability of obtaining the correct response on this item was extremely low for students with low ability and that the inflection point of the curve was shifted to the right of the middle point along the ability continuum where the item difficulty is located (b_{30} = 0.966) (Fig. 2(b)). Also the ICC indicates that the probability of getting a correct response does not increase much until the ability level of –1.0, but the probability shows a steep increase as the ability level increases. This is because the item discrimination estimate of Item 30 was high (a_{30} = 1.268). The item discrimination of the hardest item was much larger than that of the easiest item. Therefore, the ICC for the Item 30 was far steeper than that for the Item 3. This indicates that the probability of answering Item 30 correctly was highly sensitive to the variation of ability around the point at the item difficulty, b_{30} = 0.966. The guessing parameter estimate for Item 30 was close to zero (i.e., 0.064). This indicates that guessing does not have much impact on obtaining a correct response.



(a) Item characteristic curve of Item 3

(b) Item characteristic curve of Item 30

(c) Item information curve of Item 3

(d) Item information curve of Item 30

Fig. 2. Item characteristic curves and item information functions for the easiest item (Item 3) and the hardest Item (Item 30) from the 3-PL IRT model.

The item information function (IIF) indicates how precisely a particular item of the test measures the spatial visualization ability at each level of the ability. Item information is maximized at the point where the item difficulty is located. The amount of information depends on the item discrimination of a given item. In comparing the IIF for the two items in Fig. 2(c) and (d), Item 30 provided greater amount of information (0.355) than Item 3 (0.102) because Item 30 had the larger discrimination parameter. The IIF of Item 30 was quite peaked and its maximum was shifted to the right (b_{30} = 0.966) of the middle on the ability continuum. Item 30 functioned best at the ability level in terms of the precision of ability estimation and estimated spatial ability more precisely at the identified item difficulty level than Item 3. On the other hand, the IIF of Item 3 was relatively flat and its maximum information at the ability of –3.128 was substantially low (0.102). This means that Item 3 produces the most precise estimate of spatial ability at a substantially low ability level with an insufficient amount of information. The precision of the ability estimate slightly decreases as the ability level increases. In other words, the IIF of Item 3 implies that the item is not so helpful in measuring spatial ability because it is too easy for the population.

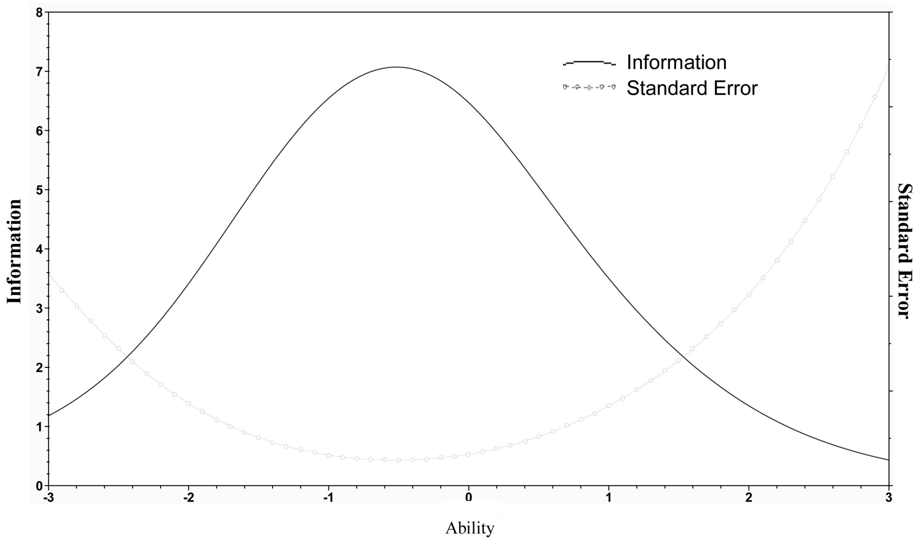


Fig. 3. Test information function of the three-parameter logistic (3-PL) IRT model.

Figure 3 shows the test information function (TIF), which was produced by summing up the IIFs of all 30 items in the Revised PSVT:R. The TIF is symmetric and peaked, where the maximum information is available, between the ability levels of –1.0 and 0.0. The information level dropped off rapidly in both directions from the peak, indicating that the Revised PSVT:R will provide the most precise estimates when the test is used to measure students whose ability levels are somewhere between –1.0 and 0.0. It also indicates that the test will provide less precise estimates of spatial ability when it is used for students with high or low spatial ability. This result is consistent with the fact that the 3–PL IRT item difficulty parameters were mostly located below zero (i.e., an average ability level), except three items (Items 22, 29, and 30).

3.3 Correlations with academic variables

Pearson product–moment correlation coefficients among academic variables and the Revised PSVT:R scores were reported in Table 3. All correlations are statistically significant at the alpha of 0.05. We found weak to moderate correlations between the standardized aptitude tests (SAT or ACT) and the Revised PSVT:R scores (r = 0.271 with ACT composite scores; r = 0.251 with SAT composite scores). Note that the ACT/SAT mathematics scores showed higher correlations with the Revised PSVT:R than the composite scores (r = 0.319 with ACT math and r = 0.321 with SAT math). Interestingly, these correlations were higher for female than male students; the correlation between ACT science scores and the Revised PSVT:R with all examinees is 0.249, while male students showed a lower correlation (0.173) than female (0.289). On the other hand, the correlation between the Revised PSVT:R and high school overall GPA was negligible (r = 0.071).

4. Discussion

Engineers are required to visualize and represent their ideas involving abstract objects on paper or computer screens and to communicate with others about the ideas graphically [45, 46]. Many research findings support a positive relationship between spatial ability, especially spatial visualization and/or mental rotation ability, and success in engineering courses [e.g., 4, 15, 18, 47]. One of the frequently used spatial ability tests in the field of engineering is the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). However, the literature review indicated a lack of empirical studies in which item and test properties of the PSVT:R were evaluated under a measurement framework. In addition, the original PSVT:R contained errors in certain items [30, 31]. Thus, after the revision of these errors in the test, it is important to provide the psychometric evidence of the Revised PSVT:R [32] for use with FYE students. We first summarize and then discuss our findings for each research question and, finally, if applicable, we present potential future directions of research related to the questions.

4.1 To what extent is the one-factor model of the theoretical construct measured by the Revised PSVT:R appropriate?

The results of EFA and CFA supported the unidimensionality of the construct measured by the Revised PSVT:R. While several studies reported correlations between the original PSVT:R and

Table 3. Pearson product–moment correlation coefficients among the spatial ability and academic variables

	2	3	4	5	6	7	8
1. Revised PSVT:R	0.271 0.320 0.239	0.251 0.298 0.241	0.071 0.154 0.101	0.055 0.101 0.094	0.319 0.349 0.246	0.321 0.379 0.262	0.249 0.289 0.173
2. ACT Composite	1.000	0.783 0.826 0.763	0.306 0.377 0.307	0.301 0.388 0.296	0.675 0.734 0.651	0.599 0.704 0.552	0.806 0.832 0.800
3. SAT Composite		1.000	0.352 0.326 0.371	0.269 0.344 0.252	0.617 0.704 0.580	0.617 0.661 0.613	0.594 0.607 0.564
4. HS GPA Overall			1.000	0.767 0.775 0.761	0.241 0.323 0.262	0.237 0.229 0.275	0.181 0.221 0.211
5. HS GPA Core				1.000	0.223 0.317 0.247	0.171 0.187 0.203	0.177 0.249 0.204
6. ACT Math					1.000	0.732 0.781 0.691	0.548 0.618 0.496
7. SAT Math						1.000	0.516 0.576 0.459
8. ACT Science							1.000

Note. First row is the correlation of total sample. Second row is the correlation of female FYE students. Third row is the correlation of male FYE students. Missing cases were excluded pairwise. All correlations were statistically significant at $p < 0.05$.

other mental rotation ability as convergent validity evidence [48–50], the result of this study suggested additional evidence of constructed-related validity regarding the adequacy of the theoretical assumption of a single latent factor structure measured by the Revised PSVT:R. However, these results only support the fact that all items in the Revised PSVT:R contribute to measuring a single factor. The analysis itself does not indicate whether or not the factor is the subcomponent of spatial ability, which is spatial visualization ability in mental rotation as Guay intended [27]. Thus, further studies investigating construct-related validity of the Revised PSVT:R with other spatial tests will help to define the spatial factor that the test measures.

4.2 How reliable are scores on the Revised PSVT:R for the first year engineering students?

With the sample of 2,469 FYE students, we found a Cronbach's internal consistency reliability of 0.839, which indicates high score reliability when used with FYE students. Note that the magnitude of reliability depends on the characteristics of the sample to which the test was administered. The obtained reliability for the Revised PVST:R with FYE students was comparable to those with other engineering student cohorts reported by Sorby and Baartmans (r = 0.82) [15] and with general undergraduate cohorts (e.g., $r = 0.81$ [51]; $r = 0.86$ [49]; r = 0.86 [32]).

4.3 To what extent do characteristics, such as item difficulty and item discrimination, vary across the items in the Revised PSVT:R?

Guay attempted to order items by difficulty in the test, which would be the order of complexity in rotation. However, the analyses of item difficulty conducted by applying both CTT and IRT frameworks revealed that the items were not ordered by item difficulty level [27]. We found that the fourth item of the test was the easiest according to the CTT analysis, and the third item was the easiest according to the IRT-based analysis. The last item was the most difficult according to both frameworks, as Guay intended. Fatigue might affect respondents' motivation and selection of the correct responses at the end of the assessment session, which resulted in an increase in item difficulty for the items located at the end of the instrument. However, we gave a closer look to the relatively easy and difficult items to provide insight into the nature of the items and its link to item difficulty for future research.

In Fig. 4, we summarized the visual differences of item characteristics among relatively easy and hard items (i.e., Items 3, 4, 22, and 30), which might affect the item difficulty levels. Guay ordered the items based on the degree of rotation and the number of rotations required to solve the spatial tasks. The first six items, including two of the easiest items (Items 3 and 4), can be solved by a single 90° rotation around one of the three axes of a 3-D Cartesian coordinate

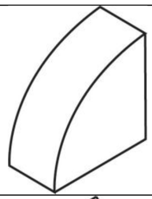
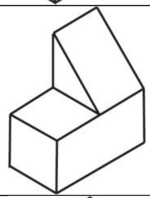
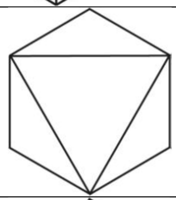
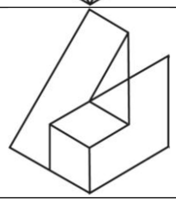
Item	Figure	Required rotation
3		One 90° rotation around one axis
4		One 90° rotation around one axis
22		One 90° rotation around one axis and another 90° rotation around a different axis
30		One 90° rotation around one axis and another 180° rotation around a different axis

Fig. 4. Rotations required solving problems in Item 3, 4, 22, and 30.

system. The next eight items (Items 7 to 14) in the PVST:R can be solved by a single 180° rotation around one of the three axes, and the next eight items (Items 15 to 22) can be solved by two 90° rotations around two different axes. The last eight items (Item 23 to 30), including the most difficult item (Item 30), can be solved by a combination of a single 90° rotation and a single 180° rotation around two different axes [28, 32]. The shape of the objects used in the item for rotation might also affect the item difficulty, because some shapes may be more difficult than others to mentally rotate. While the objects used in the two easiest items (Item 3 and 4) are simple 3-D shapes, the objects used in the two most difficult items are relatively complex. According to Hegarty and Waller:

> Performance on tests of spatial abilities depends on execution of basic cognitive processes such as encoding a visual stimulus, constructing a visual image, retaining an image in working memory, transforming an image, and comparing a visual stimulus to an image in working memory. [52, p. 136]

Mumaw *et al.* [53] also mentioned that individual differences in spatial ability tend to occur at any step of cognitive processing. Although we could identify the level of item difficulty and probability of guessing responses, further investigation of required cognitive processing for each items would be beneficial to understand how specific components of a figure (e.g., the degree of complexity of the shape defined with inclined, oblique, and/or curved surfaces, etc.) and the required tasks to solve the item (the degrees and/or direction of rotations) contribute to the determination of item difficulty and response style (e.g., random guessing), given an individual's ability level. Mumaw and others further suggested that the speed of cognitive processing to solve spatial tasks would produce differences in individuals' test scores. It may be also interesting to scrutinize the features of 3-D objects in the Revised PSVT:R and to investigate their relationship to the level of cognitive processing and processing speed.

Another important finding regarding the item characteristics was that, in general, items were relatively easy to solve for the population, although our results showed that the test functions well to discriminate the level of spatial visualization ability among FYE students. Our IRT-based analysis indicates that the test provides the most precise estimate for students whose ability level are somewhere between –1.0 and 0.0 (i.e., raw total scores between about 16 and 23). As an examinee's ability level becomes progressively either lower or higher than the range, the amount of measurement error increases. This may have an important implication for setting up a cut-off score for selecting individuals for a remedial program or advising students to take a certain course. Moreover, additional items with different item difficulty levels may be need to adapt the test to students with a wide variety of ability levels.

Finally, we found a significant gender difference in both raw total scores and ability scores. On average, male students outperformed female students by about three points. The magnitude of the difference was relatively large (Cohen's $d = 0.678$), which is consistent with the size reported in the meta-analytic study [54]. It is worth investigating whether items functions differently by gender, which may result in differential performances by gender as a future study.

4.4 To what extent is the Revised PSVT: R related to academic variables?

Pearson's correlation coefficients between the Revised PSVT:R scores and the standard aptitude scores (used as a proxy of academic outcomes) showed positive, but weak, correlations. As we speculated, slightly higher correlations were observed for mathematics and science sub-scores. Interestingly, the sub-score correlations with the Revised PSVT:R were higher for female than male students. We found almost no correlation with high school overall and core GPAs. Thus, weak to moderate correlations with the aptitude test scores provided evidence that the test scores on the Revised PSVT:R may probably provide different information when used to predict FYE academic outcomes along with other standardized test scores as criterion variables.

Related to this point, one of the most important analyses, which we could not investigate in the current study, is to investigate the predictive power of the test scores on the Revised PSVT:R. Because one of the main purposes of using the Revised PSVT:R scores in a FYE program is to predict future academic performance, it is worthwhile, for example, to investigate how performance on the Revised PSVT:R is related to the first semester GPA or retention in the FYE program. Furthermore, it may be interesting to establish evidence to support the use of the spatial ability scores to predict student retention in the engineering program. The evidence of predictive validity for the inferences will provide further implication for the use of the Revised PSVT:R scores in instruction and curriculum design to ascertain the role that ability plays in students' academic success.

5. Conclusions

Cognitive assessments are used in a variety of instructional settings in the current educational system. On some occasions we make serious educational decisions based on assessment results. For example, providing appropriate educational guidance for selecting courses based on assessment results is one of the main reasons that assessments are used in educational settings. The PSVT:R is one of the cognitive assessments used to measure examinees' mental rotation ability, and has been used to make inferences about examinees' future academic success in STEM fields. The PSVT:R has also been used as a placement test in some programs. However, without a full understanding of the nature of the test, fair use of the test in making educational decisions cannot be guaranteed. In particular, the use of a psychometrically sound instrument is critical for ensuring appropriate interpretation and valid application of the results in order to make decisions and/or judgments.

This study was conducted to investigate the psychometric properties of the Revised PSVT:R for use with FYE students. We focused on FYE students because the PVST:R has been used frequently with this population. Thus, identifying item functions specific to this population can help future users of the test enhance their understanding of test scores. We found that the Revised PSVT:R measures an unidimensional concept with high reliability. Scores on the Revised PSVT:R represent a different student attribute from those that are typically represented by academic variables, such as GPA and standardized test scores (e.g., SAT and ACT scores). Although the instrument seems slightly easy for the FYE population, items vary in terms of difficulty and discrimination power to measure different levels of student ability with an acceptable amount of measurement error. However, with further analyses of the relationship between item functions (i.e., item difficulty and discrimination) and item structures (e.g., the complexity of the required mental rotation of an object to solve an item), the utilities (e.g., appropriateness of cut-off scores) and interpretation of the test scores could be enhanced.

In summary, the study provided sound and detailed psychometric evidence of the Revised PSVT:R for the use of the FYE population. We encourage current and future users of the test to scrutinize the information provided by the study to ensure appropriate use and interpretation of the scores for their educational purposes.

References

1. M. Hegarty and D. Waller, Individual differences in spatial abilities. In P. Shah and A. Miyake (eds.), *The Cambridge Handbook of Visuospatial Thinking*, Cambridge University Press, New York, 2006.
2. D. F. Lohman, *Spatial ability and G.* Paper presented at the

first Spearman seminar, University of Plymouth. 1993. Retrieved on January, 13, 2011 from http://faculty.education.uiowa.edu/dlohman/pdf/spatial_ability_and_G.pdf
3. W. B. Michael, J. P. Guilford, B. Fruchter and W. S. Zimmerman, The description of spatial-visualization abilities, *Educational and Psychological Measurement,* **17**, 1957, pp. 185–199.
4. M. G. McGee, Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influence, *Psychological Bulletin*, **86**, 1979, pp. 889–918.
5. D. F. Lohman, Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (ed.), *Advances in the Psychology of Human Intelligence*, Erlbaum, Hillsdale, NJ, 1988, pp. 181– 248.
6. J. Carroll, *Human Cognitive Abilities: A Survey of Factor-Analytical Studies*, Cambridge University Press, New York, 1993.
7. B. J. Baartmans, Gender differences in spatial visualization and predictors of success in an engineering design course, *Proceedings of the National Conference on Women in Mathematics and the Sciences, St. Cloud*, MN.1990.
8. M. Contero, F. Naya, P. Company, J. L. Saorin and J. Conesa, Improving visualization skills in engineering education, *Computer Graphics and Applications*, IEEE, **25**, 2005, pp. 24–31.
9. B. W. Field, Visualization, intuition, and mathematics metrics as predictors of undergraduate engineering design performance, *Journal of Mechanical Design*, **129**, 2007, pp. 735–743.
10. K. L. Norman, Spatial visualization–A gateway to computer-based technology. *Journal of Special Educational Technology*, **12**, 1994, pp. 195–206.
11. D. Budny, G Bjedov and W. LeBold, Assessment of the impact of the freshman engineering courses, *Proceedings of the 27th Annual Frontiers in Education Conference,* 2, 1997, pp. 1100–1106.
12. H. Gerson, S, Sorby, A.Wysocki and B. Baartmans, The development and assessment of multimedia software for improving 3-D spatial visualization skills, *Computer Applications in Engineering Education*, **9**, 2001, pp. 105–113.
13. S. A. Sorby, Improving the spatial skills of engineering students: Impact on graphics performance and retention, *Engineering Design Graphics Journal*, **65**(3), 2001, pp. 31–36.
14. S. A. Sorby, A course in spatial visualization and its impact on the retention of female engineering students, *Journal of Women and Minorities in Science and Engineering*, **7**, 2001, pp. 153–171.
15. S. A. Sorby and B. J. Baartmans, The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students, *Journal of Engineering Education*, **89**, 2000, pp. 301–307.
16. K. M. Ullman and S. A. Sorby, Enhancing the visualization skills of engineering students through computer modeling, *Computer Application in Engineering Education*, **3**, 1995, pp. 251–257.
17. B. L. Kinsey, E. Towle, E. O'Brien and C. F. Bauer, Analysis of self-efficacy and ability related to spatial tasks and the effect on retention of engineering students, *International Journal of Engineering Education*, **24**, 2008, pp. 488–494.
18. R. Devon, R. Engel and G. Turner, The effects of spatial visualization skill training on gender and retention in engineering, *Journal of Women and Minorities in Science and Engineering,* **4**, 1998. pp. 371–380.
19. M. Peters, P. Chisholm and B. Laeng, Spatial ability, student gender, and academic performance, *Journal of Engineering Education*, **84**,1995, pp. 69–73.
20. D. Voyer, S.Voyer and M. P. Bryden, Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables, *Psychological Bulletin,* **117**, 1995, pp. 250–270.
21. College Entrance Examination Board, *CEEB Special Aptitude Test in Spatial Relations,* College Entrance Examination Board, New York, 1939.
22. S. G. Vandenberg and A. R. Kuse, Mental rotations, a group test of three-dimensional spatial visualization, *Perceptual and Motor Skills*, **47**, 1978, pp. 599–604.
23. W. H. Quasha and R. Likert,The revised Minnesota paper form board test, *Journal of Educational Psychology*, **28**, 1937, pp. 197–204.
24. G. K. Bennett, H. G. Seashore and A. G. Wesman, *Differential Aptitude Tests*, The Psychological Corporation, New York, NY, 1973.
25. C. L. Miller, A historical review of applied and theoretical spatial visualization publications in engineering graphics, *Engineering Design Graphics Journal*, **60**(3), 1996, pp. 12–33.
26. Pearson Education, Inc., Minnesota Paper Form Board Test—Revised. Retrieved from January 18, 2011, http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-4183-30X&Mode=summary.
27. R. B. Guay, *Purdue Spatial Visualization Test,* Purdue Research Foundation, West Lafayette, IN, 1976.
28. J.Yue, Spatial visualization by orthogonal rotations. Paper session presented at the American Society for Engineering Education (ASEE) Annual Conference and Expositions, Salt Lake City, Utah, June, 2004.
29. R. B. Guay, Spatial ability measurement: A critique and an alternative. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED189166), 1980.
30. J. Yue, Spatial visualization by isometric drawing, *Proceedings of the 2006 IJME-INTERTECH Conference,* Union, New Jersey, 2006.
31. J. Yue, Spatial visualization by realistic 3D views, Paper session presented at the American Society for Engineering Education (ASEE) Annual Conference and Expositions, Honolulu, Hawaii, June, 2007.
32. S. Y. Yoon, Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT:R), Doctoral Dissertation, Retrieved from ProQuest Dissertations and Theses. (Order Number: 3480934), 2011.
33. S. Y. Yoon and E. L. Mann, Exploring spatial ability of undergraduate students: Association with gender, STEM majors, and gifted program experiences. Manuscript submitted for publication, 2011.
34. L. K. Muthén and B. O. Muthén, *Mplus User's Guide*, 6th edn, Muthén & Muthén, Los Angeles, CA, 2010.
35. T. A. Brown, Confirmatory Factor Analysis for Applied Research, Guilford Press, New York, NY, 2006.
36. S. E. Embretson and P. Reise, Item Response Theory, Lawrence Erlbaum Associates, Mahwah, NJ, 2000.
37. H. Tinsley and D. Tinsley, Uses of factor analysis in counseling psychology research, *Journal of Counseling Psychology,* **34**, 1987, pp. 414–424.
38. R. K. Hambleton and H. Swaminathan, *Item Response Theory, Principles and Applications*, Kluwer Nijhoff Publishing, Boston, MA, 1985.
39. R. K. Hambleton, H. Swaminathan and H. J. Rogers, *Fundamentals of Item Response Theory*, Sage Publications, Inc., Newbury Park, CA, 1991.
40. S. Reise and J. Henson, A discussion of modern traditional psychometric as applied to personality assessment scales, *Journal of Personality Assessment*, **81**, 2003, pp. 93–103.
41. R. Zimowski, E. Muraki, R. Mislevy and D. Bock, *BILOG-MG*, SSI Scientific Software, Lincolnwood, IL, 1996.
42. Y. Maeda and S. Y. Yoon, Measuring spatial ability of first year engineering students with the Revised PSVT:R. Paper presented at the 118th American Society for Engineering Education (ASEE) Annual Conference and Exposition, Vancouver, Canada, June, 2011.
43. M. W. Brown and R. Cudeck, Alternative ways of assessing model fit, In K. A. Bollen and J. S. Long (eds.), *Testing Structural Equation Models*, Sage, Newbury Park, CA, 1993, pp. 136–162.
44. F. M. Lord, The relation of the reliability of multiple-choice tests to the distribution of item difficulties, *Psychometrika*, **17**, 1952, pp. 181–194.
45. S. Hsi, M. C. Linn and J. E. Bell, The role of spatial reasoning in engineering and the design of spatial instruction, *Journal of Engineering Education*, **86**, 1997, pp. 151–158.
46. W. E. Mack, The effect of training in computer-aided design

on the spatial visualization ability in selected gifted adolescents (unpublished doctoral dissertation), Virginia Tech, 1992, Retrieved January 18, 2011 from http://scholar.lib.vt.edu/theses/available/etd-02032004-161628.
47. W. H. Gillespie, Using solid modeling tutorials to enhance visualization skills (unpublished doctoral dissertation), University of Idaho, Moscow, ID, 1995.
48. T. J. Branoff, Spatial visualization measurement: A modification of the Purdue Spatial Visualization Test—visualization of rotations, *Engineering Design Graphics Journal*, **64**(2), 2000, pp. 14–22.
49. R. B. Guay and E. McDaniel (eds.), *Correlates of Performance on Spatial Aptitude Tests,* U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 1978.
50. S. A. Sorby, Spatial abilities and their relationship to effective learning of 3-D solid modeling software, *Engineering Design Graphics Journal*, **64**(3), 2000, pp. 30–35.
51. H. M. Alkhateeb, Spatial visualization of undergraduate education majors classified by thinking styles, *Perceptual and Motor Skills*, **98**, 2004, pp. 865–868.
52. M. Hegarty and D. Waller, A dissociation between mental rotation and perspective-taking spatial abilities, *Intelligence,* **32**, 2004, pp. 175–191.
53. R. J. Mumaw, J. W. Pellegrino, R. V. Kail and P. Carter, Different slopes for different folks: Process analysis of spatial aptitude, *Memory & Cognition*, **12**, 1984, pp. 515–521.
54. Y. Maeda and S. Y. Yoon, A meta-Analysis on gender differences in mental rotation ability measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educational Psychology Review*, **25**(1), 2013, pp. 69–94, doi: 10.1007/s10648-012-9215-x.

Yukiko Maeda, is an assistant professor of Educational Studies in the area of Applied Measurement and Research Methodology at Purdue University. She received her Ph.D. in Quantitative Methods in Education from University of Minnesota. She has expertise in educational measurement and applied statistics including survey design, investigating effects of assumption violations of statistical tests, particularly used in multilevel analysis and meta-analysis. She has been involved with various STEM education research projects sponsored by NSF and IES. She is currently involved in two NSF-funded STEM education research projects, *Preparing To Teach Algebra: A Study Of Teacher Education,* and *Implementation of a Multidimensional Assessment Tool to Explore the Impacts of Pedagogy On Undergraduate Student learning* as a Co-PI and methodologist.

So Yoon Yoon, is an INSPIRE post-doctoral research associate in the School of Engineering Education at Purdue University. She received her Ph.D. and M.S.Ed. in Educational Psychology with the specialties in Gifted Education and Research Methods & Measurement, respectively, from Purdue University. She also holds an M.S. in Astronomy & Astrophysics and a B.S. in Astronomy and Meteorology from Kyungpook National University in South Korea. Her work centers on the development and validation of instruments, particularly useful for P-16 STEM education settings, such as the Revised PSVT:R and the Teaching Engineering Self-efficacy Scale (TESS), the evaluation of engineering teacher professional development programs, and the investigation of P-16 students' spatial ability to understand its association with their academic performance and talents in STEM fields.

Gyenam Kim-Kang, is a Senior Researcher at the Measurement and Statistics Research located in West Lafayette, Indiana. She obtained her Ph.D. from the Department of Educational Psychology with emphasis on Measurement and Statistics in Education at the University of Minnesota, Twin Cities. She received her Master's Degree focusing on Educational Technology in the Department of Curriculum and Instructional Systems at the University of Minnesota. She obtained her Bachelor's degree in Education from Yonsei Unversity in Seoul, Korea. She was an associate professor in the Department of General Education at the Korea Nazarene University and visiting scholar in the Department of Psychology at the University of Minnesota and in the Department of Educational Studies at Purdue University. Her primary research interests include the development of optimal conditions of various options in implementing adaptive self-referenced testing in measuring individual change in the computerized adaptive testing using item response theory, such as item selection criterion, termination criteria and test statistics for detecting individual change with adaptive measurement of change. Her previous researches include analyses of empirical power and experiment-wise type I error rates in the area of multiple comparison procedures in repeated measures design and of pooling procedures of appropriate error terms in a mixed analysis of variance using Monte Carlo Simulation. She was also interested in investigating validity analyses on religious oriented scales such as the QUEST scale, Religious Maturity Scale, Faith Maturity Scale and Spiritual Well-being Scale.

P. K. Imbrie is the Director of the College of Engineering Honors Program and an Associate Professor, by courtesy, in the School of Engineering Education and School of Aeronautics and Astronautics at Purdue University. He holds B.S., M.S. and Ph.D. degrees in Aerospace Engineering from Texas A&M University. His research interests include educational research, solid mechanics, experimental mechanics, microstructural evaluation of materials, and experiment and instrument design. He has been involved with various research projects sponsored by NSF, NASA and AFOSR, ranging from education-related issues to traditional research topics in the areas of elevated temperature constitutive modeling of monolithic super alloys and environmental effects on titanium based metal matrix composites. His current research interests include: epistemologies, assessment, and modeling of student learning, student success, student team effectiveness and global competencies; experimental mechanics; and piezospectroscopic techniques.