Classical test theory and Rasch analysis of test of understanding of vectors (TUV)

C Y Ly1, S Rakkapao1, 2, K Nualtong3 and W Sumathakulawat1*

1Department of Physics, Division of Physical Science, Faculty of Science, Prince of Songkla University, Songkhla, 90110, Thailand
2Thailand Center of Excellence in Physics, Commission on Higher Education, 328 Si Ayuthaya Road, Bangkok 10400, Thailand
3Department of Mathematics and Statistics, Division of Computational Science, Faculty of Science, Prince of Songkla University, Songkhla, 90110, Thailand

*Corresponding author’s E-mail: wanida.s@psu.ac.th

Abstract. This study aimed to apply classical test theory (CTT) and Rasch analysis of item response theory (IRT) to investigate Thai participants’ understanding of vectors and the characteristics of the test of understanding of vectors (TUV). A Thai version of the 20-items TUV was administered to 1,021 first-year science and engineering students at Prince of Songkla University, before and after common instructions. Although the engineering students (pre-score = 62%) had a significantly higher background of vectors than the science students (pre-score = 24%), proved by the effect size of Cohen’s $d$, both groups have improved their learning into a medium level of the normalized gain ($0.3 < g < 0.7$) after the instructions. Both groups improved the most with a vector representation concept. The participants were found to have the most difficulties with the unit vector and the cross product. However, the TUV is fairly easy for the assessed Thai first-year students. Overall, the item difficulty index derived from CTT was in moderate range ($0.33-0.69$). The difficulty index of the Thai-TUV obtained from CTT and Rasch analysis has a strong correlation indicating a reliable assessment tool.

1. Introduction
A vector is a basic concept in physics courses at a university level. Students, who completely understand vectors, can easily apply their knowledge on other related physics concepts, as found for forces and motions [1-2]. Several assessment instruments for monitoring and diagnosis students’ understanding of vectors in physics education research (PER) were published [2-4]. A study of Knight (1995) developed open-ended questions of vectors covering a vector component, magnitude, direction, and two graphical vectors addition [2]. Excluding the component of vectors, Nguyen and Meltzer (2003) proposed a similar research [3]. Barniol and Zavala (2014) developed the test of understanding of vectors (TUV) and deployed it to Mexican students [4]. To explore the understanding of entire sub-topics of vectors required for an introductory physics course at a university level, on a single exam, the current study applied the TUV to collect data.

The TUV is a 20-item test with five choices per item covering ten vector concepts and popularly used in PER community [5-8]. The TUV was applied to 2,392 Thai university students and investigated by the three-parameter logistic model of item response theory and the item response curves technique. It
revealed the effectiveness of the test and its distractors [5]. Based on the results of TUV, tutorial worksheets were designed for instructions in high school and university levels [6]. Moreover, the TUV was applied to 889 Croatian university students and analyzed by the Rasch analysis. Reportedly, the test was easy for the target group and more difficult items were required [7]. The model analysis technique was also introduced to analyze the TUV items [8].

This study aimed to apply the TUV 1) to diagnose Thai student understanding of vectors and, 2) to investigate the relation of item difficulty indices calculated by the classical test theory (CTT) and Rasch analysis. The students’ understanding found in this research will be useful for educators in designing lesson plans, teaching materials and assessment tools for vectors and other related courses. Moreover, the result of CTT and Rasch analysis will demonstrate the reliability of the TUV in applying to Thai participants.

2. Data collection
The 20-item TUV was translated into Thai language and validated by a group of Thai physics professors. It was administered to first-year science (SC; N = 340) and engineering (EN; N = 681) students of Prince of Songkla University, Thailand via the online learning management system of the university, before and after the common instruction within two weeks as pre-test and post-test (N total = 1,021). The common instruction refers mostly to lecture-based methods, some in-class problem-solving activities, and homework assignments.

3. Difficulty index derived from classical test theory (CTT) and Rasch analysis

3.1. Difficulty index derived from CTT ( \( P \) )
In the CTT framework, the item difficulty index ( \( P \) ) characterizes the proportion of students who correctly answer the item to the total number of students. The difficulty index ranges from 0 to 1. An item with a difficulty index close to 1 means an easy item and vice versa. The standard criterion for the acceptable range of the item difficulty index is 0.4-0.6 [9].

3.2. Difficulty index derived from Rasch analysis ( \( b \) )
The Rasch analysis is the one-parameter logistic model of item response theory (IRT) developed by George Rasch in 1960 [10]. IRT defines that the performance of an examinee on a test item can be predicted from the item’s typical features (called item parameters) and the examinee’s ability (called person parameter), which their relationship can be plotted using the logistic function. In Rasch analysis for dichotomous data, item \( i \) difficulty parameter ( \( b_i \) ) associates with the person ability ( \( \theta \) ) as written by

\[
P_i(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}
\]

with the ability ( \( \theta \) ). The \( \theta \) is usually standardized to zero mean and one standard deviation. In practice, its values are usually in the interval [-3,3]. The \( b \) parameter for Rasch analysis is the \( \theta \) at the inflection point of the curve, in which the probability is 0.5. The greater of a difficulty parameter, the harder of an item for the students to require a 50% chance of their ability to get the correct answer [11]. In this study, a free download Rasch analysis program named JMetrik [12] was applied to resolve the \( b \) parameter of the TUV items.

4. Results and discussion

4.1. Students’ understanding of vectors
Thai students’ understanding of vectors was analyzed from their responses to the 20-item TUV. The average pre-scores and post-scores in percentage for ten sub-concepts were shown in figure 1. The average normalized gain \( \langle g \rangle = (%<\text{post}>-%<\text{pre}>)/(100-%<\text{pre}>), \) and its errors \( \Delta <g> \) were calculated [13]. Overall, we found that the engineering (EN) students (pre-score = 62%) had a higher background of vectors than the science (SC) students (pre-score = 24%) as proved by the effect size of Cohen’s \( d \) [14].Remarkably, both groups gained their learning into a medium level \( <g> \pm \Delta <g>=0.55 \)
± 0.03 for SC; = 0.65 ± 0.03 for EN) after the instruction. The unit vector and the cross product of vectors were the most difficult concepts of the sample groups at the beginning of the courses. 58% of SC and 30% of EN students believed that the unit vector has both x and y components of one unit. Moreover, 43% of SC and 16% of EN students were confused about the cross product by applying \( \hat{i} \times \hat{i} = \hat{i} \). This observation was similarly observed in previous studies [4-5, 7]. After the instruction, both concepts remained as most difficult for SC students to grasp and these students were likely to least improve learning on the unit vector (\(|g| ± Δ < g > = 0.49 ± 0.03\)). EN students least improved their learning on the direction and addition of vectors, and a dot product of vectors. For example, 30% of students at pre-instruction drew the tip-to-tip method to sum two vectors, which had the same magnitude and set an obtuse angle. The percentage of EN students, who held this misunderstanding increased to 46% at post-instruction. Moreover, 34% of EN students believed that a dot product of two vectors was the magnitude of a vector between the two vectors pointing up to the right.

![Figure 1. Percentage of average pre-score and post-score, and the average normalized gain with its errors (<g> ± Δ < g >) for ten concepts of TUV responded by Thai engineering (EN) (N = 340), and science (SC) (N = 681) participants. An error bar displays one standard deviation. The ten concepts of TUV were represented by a number, where 1 = direction, 2 = magnitude, 3 = component, 4 = unit vector, 5 = vector representation, 6 = addition, 7 = subtraction, 8 = scalar multiplication, 9 = dot product, and 10 = cross product.](image)

**Figure 1.** Percentage of average pre-score and post-score, and the average normalized gain with its errors (\(<g> ± Δ < g >\)) for ten concepts of TUV responded by Thai engineering (EN) (N = 340), and science (SC) (N = 681) participants. An error bar displays one standard deviation. The ten concepts of TUV were represented by a number, where 1 = direction, 2 = magnitude, 3 = component, 4 = unit vector, 5 = vector representation, 6 = addition, 7 = subtraction, 8 = scalar multiplication, 9 = dot product, and 10 = cross product.

4.2. Difficulty index of Thai-TUV
The item difficulty indices of Thai-TUV (pre-test) derived from CTT (\(P\)) and Rasch analysis (\(b\)) were shown in table 1. \(P\) values were [0.33-0.69] indicating moderate ranges. It demonstrated that 33-69% of the respondents can correctly answer a question. \(b\) parameters were [-1.36, 2.01]. Moreover, a strong correlation \((r = 0.99)\) between \(P\) and \(b\) of the 20-item was observed. The strength of invariance of the item parameter across samples was found to be similar for the two frameworks, which usually considered as the theoretical superiority IRT framework. Although the theoretical difference between CTT and Rasch analysis, the item difficulty indices derived in this study are comparable, which is consistent with a work of Fan (1998) [15]. It suggests that the obtained difficulty indices are independent of Thai participants’ abilities indicating a reliable TUV in Thai version.
Table 1. Item difficulty indices of 20-item Thai-TUV derived from CTT ($P$) and Rasch analysis ($b$).

| Item | q1 | q2 | q3 | q4 | q5 | q6 | q7 | q8 | q9 | q10 |
|------|----|----|----|----|----|----|----|----|----|-----|
| $P$  | 0.56 | 0.49 | 0.33 | 0.69 | 0.62 | 0.66 | 0.68 | 0.50 | 0.64 | 0.63 |
| $b$  | -0.14 | 0.67 | 2.00 | -1.26 | 0.11 | -1.36 | -0.91 | 0.48 | -0.66 | -0.89 |

| Item | q11 | q12 | q13 | q14 | q15 | q16 | q17 | q18 | q19 | q20 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $P$  | 0.57 | 0.49 | 0.56 | 0.69 | 0.50 | 0.34 | 0.54 | 0.52 | 0.52 | 0.62 |
| $b$  | -0.08 | 0.63 | -0.25 | -1.04 | 0.54 | 2.01 | 0.03 | 0.36 | 0.42 | -0.63 |

5. Conclusion

The 20-item TUV was applied to 1,021 Thai students to observe their understanding of vectors and to investigate a correlation of item difficulty indices obtained between CTT and Rasch analysis. It revealed that Thai participants had difficulty with a unit vector, cross and dot product, subtraction and direction, as well as the unit vector was the least improvement concept. A revised instruction to promote students’ learning on vectors is required. The students’ misunderstanding extracted by a reliable TUV test can be used as significant resources to design instructional management in a further step.

Acknowledgments

The project is funded by Royal Scholarship under Her Royal Highness Princess Maha Chakri Sirindhorn of Faculty of Science, Prince of Songkla University, Thailand.

References

[1] Sirait J, Hamdani and Oktavianty E 2017 J. Turk. Sci. Educ. 14 82
[2] Knight R D 1995 Phys. Teach. 33 74
[3] Nguyen N L and Meltzer D E 2003 Am. J. Phys. 71 630
[4] Barniol P and Zavala G 2014 Phys. Rev. ST. Phys. Educ. Res. 10 010121
[5] Rakkapao S, Prasitpong S and Arayathanitkul K 2016 Phys. Rev. Phys. Educ. Res. 12 020135
[6] Barniol P 2016 EURASIA J. Math. Sci. Technol. Educ. 12 2387
[7] Susac A, Planinic M, Klemencic D and Milin Sipus Z 2018 Phys. Rev. Phys. Educ. Res. 14 023101
[8] Kusindrastuti B, Reyes M G, Rakkapao S and Prasitpong S 2019 J. Phys.: Conf. Ser. 1380 012014
[9] Doran R L 1976 Sci. Educ. 60 199
[10] Boone W J and Scantlebury K 2006 Sci. Educ. 90 253
[11] Hambleton R K, Swaminathan H, and Jane Rogers H 1991 Fundamentals of Item Response Theory (Newbury Park London New Delhi: Sage Publications)
[12] Meyer J P 2014 Applied Measurement with jMetrik (New York and London:Routledge)
[13] Hake R R 1998 Am. J. Phys. 66 64
[14] Cohen J 1988 Statistical Power Analysis for the Behavioral Sciences (2nd ed.) (New Jersey: Lawrence Erlbaum Associates)
[15] Fan X 1998 Educ. Psychol. Meas. 58 357