Development of a representational conceptual evaluation in the first law of thermodynamics

S P Sriyansyah[1] and A Suhandi[2]

[1]Department of Physics Education, School of Postgraduate Studies, Indonesia University of Education, Jl. Dr. Setiabudhi No. 229, Bandung 40154
[2]Department of Physics Education, FPMIPA, Indonesia University of Education, Jl. Dr. Setiabudhi No. 229, Bandung 40154

Email: syakti.perdana.sriyansyah@student.upi.edu

Abstract. As part of an ongoing research to investigate student consistency in understanding the first law of thermodynamics, a representational conceptual evaluation (RCET) has been developed to assess student conceptual understanding, representational consistency, and scientific consistency in the introductory physics course. Previous physics education research findings were used to develop the test. RCET items were 30 items which designed as an isomorphic multiple-choice test with three different representations concerning the concept of work, heat, first law of thermodynamics, and its application in the thermodynamic processes. Here, we present preliminary measures of the validity and reliability of the instrument, including the classical test statistics. This instrument can be used to measure the intended concept in the first law of thermodynamics and it will give the consistent results with the ability to differentiate well between high-achieving students and low-achieving students and also students at different level. As well as measuring the effectiveness of the learning process in the concept of the first law of thermodynamics.

1. Introduction

Students' conceptual understanding has become the focus of Physics Education Research (PER) for many years. Much previous research has reported learning difficulties encountered by students [1]. PER has shown that many students who usually do well on quantitative problem solving have serious conceptual difficulties and one of the topics which many students have difficulty therein is thermodynamics. Previous research on students at university level indicates that the student has serious conceptual difficulties, particularly in the first and second law of thermodynamics [2-5]. These difficulties seem to be perceived by students around the world [6-7]. Therefore, it is important for every instructor to detect students' difficulties in the beginning of instruction in order to design an effective learning methods to be used. For that purpose, research-based conceptual assessment plays an important role in developing a comprehensive instrument that needed.

The instrument most widely developed taking the concept of mechanics, electricity and magnetism, but still limited on thermodynamics [see Ref.1]. Even it has been developed representational variant instrument of FCI [8] which not only assess students' conceptual understanding, but also the students' ability to interpret multiple representations consistently in the context of the force concept. As for the concept of thermodynamics, it has been developed Thermodynamics Concept Survey (TCS) [7] that

used to assess students' understanding of ideal gas law and the first law of thermodynamics. The TCS addresses several representations in a variety of context, but it does not provide a systematic evaluation of students' ability to use multiple representations when the context is fixed (representational and scientific consistency). The existing tests like the TCS are limited in that they do not permit a comprehensive evaluation of students' skills in using multiple representations. Whereas, the representational ability need to be taken into account in the learning of thermodynamics [2]. This deficiency led us to develop an instrument called Representational Conceptual Evaluation in the first law of Thermodynamics (RCET) that measures students' ability in using multiple representations and assess conceptual areas that has not been included in the TCS. Our focus on students' conceptual understanding, representational consistency and scientific consistency of the concept of the first law of thermodynamics.

2. Overview

The representational conceptual evaluation aims to evaluate students' understanding, representational consistency, scientific consistency of the concepts including work, heat, the first law of thermodynamics, and its application in the thermodynamic processes. Most RCET items were adapted from previous physics education research findings [2,3,6] and designed as an isomorphic multiple-choice test with three different representations. Some items of RCET were taken from some textbooks that has been reviewed with highly careful considerations. This test adopts the form of RFCI tests and the instrument used by Meltzer [9]. The current version of RCET consisted of 30 items with three different representations (verbal, mathematics, diagram/pictorial) and concerning ten themes. The term 'theme' used for three isomorphic items that keep the physical concept and context of the items as similar as possible. A brief description of themes, categories of conceptual areas, along with the item number and its representations, is given in table 1.

Table 1. Themes, categories of conceptual areas, items and representations of RCET including verbal (V), mathematics (M), diagram (D) and pictorial (P).

Theme	Conceptual Areas in a Given Theme	Representations
T1	System's internal energy	1V, 15M, 23D
T2	Work is a process-dependent function	2D, 11V, 25M
T3	Internal energy is a state-function	3D, 13V, 21M
T4	Heat is a process-dependent function	4M, 14D, 22V
T5	Adiabatic compression	5V, 18M, 26D
T6	Adiabatic free-expansion	6D, 19V, 27M
T7	Isochoric process	7M, 17D, 28V
T8	Isothermal process	8V, 12D, 29M
T9	Cyclic process	9M, 20P, 30V
T10	Specific heats of Gases	10D, 16V, 24M

3. Validity

RCET was administered to 95 second-year students who have completed a thermodynamics course at the end of second semester (year 2014/2015). Test response of the students was used to calculate an item analysis and the quality of the test, including validity and reliability. Validity is defined as the extent to which the test accurately measures the intended concept. Here we presented two types of validity, criterion validity and expert validity.

3.1. Criterion validity

Criterion validity is the extent to which the test gives results similar to other independent measures, such as an examination test valued by faculty [10]. This validity was measured by correlating students'

score of the examination test and the test response of RCET. Pearson correlation coefficient between RCET scores and students' final exams shows highly correlation of 0.49 (N=51), significant at the 0.01 level. This correlation is considered "medium" [11] indicating that the construct measured on RCET are related to other students' performance on the independent measures. In other word, the RCET items have been precisely measured the desired concept.

3.2. Expert validity

For further strengthen the validity of RCET, then we do the content validity by the expert judgements. The RCET items were reviewed by five physics faculty member at Indonesia University of Education (three are experienced thermodynamics lecturer and two are the expert of evaluation and learning on physics education). We provide them the RCET and corresponding learning objectives that matched to each RCET item. We would like to know whether the experts agree with the questions in RCET measure the intended concepts and matched learning objectives.

All of the experts agreed that the questions in RCET are valuable and useful. But there are small notes provided by the experts. The expert suggested to fix some *P-V* diagram and descriptions of thermodynamics process. Overall, experts agreed that RCET items were appropriate and accordance with the learning objectives and the desired concept. Here we provide a sample of the alternatives for theme 2 with two representations that used in figure 1.

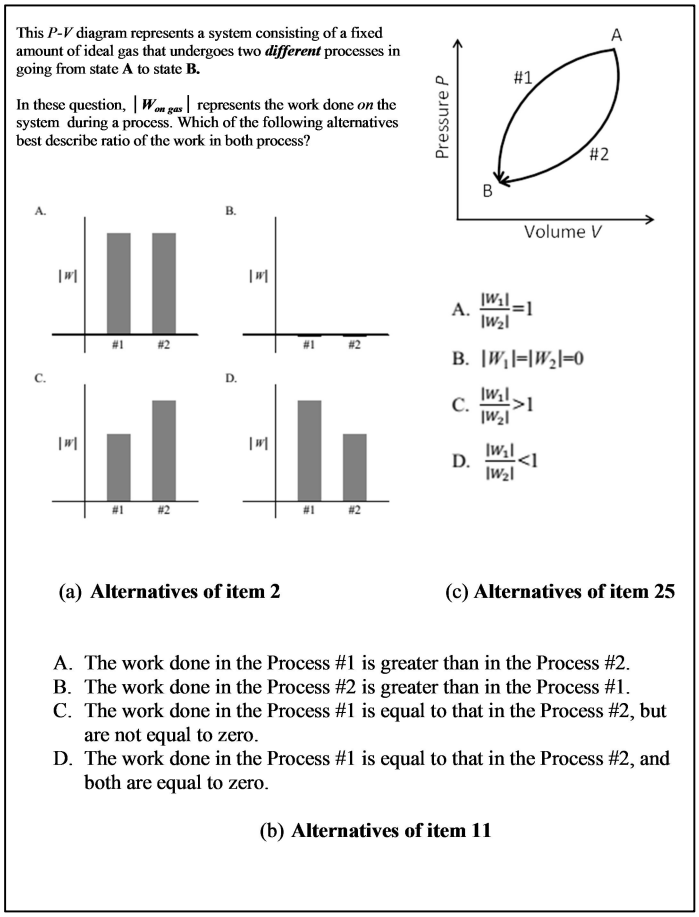

Figure 1. Sample of the question in theme 2 and the alternatives of three items in theme with two representations, (a) diagram representation, (b) verbal representation, and (c) mathematics representation.

4. Statistical indices

Based on the number of statistical analyses in the classical test theory, we evaluate the item analysis and test analysis. Here we present the evaluation of RCET that measure difficulty index, discrimination index, and point biserial coefficient for item analysis, while the reliability index and Ferguson's delta for test analysis. Those five measures have been often used to evaluate the multiple-choice test item in Physics Education Research [12]. The purpose of these measures is to examine whether the test is reliable and discriminating. Table 2 shows the evaluation results of the test.

Table 2. Evaluation of the RCET

Evaluation Measure	Values of the RCET	Desired Values
Difficulty index (*P*)	Average of 0.53	0.30-0.90
Discrimination index (*D*)	Average of 0.45	≥0.30
Point biserial coefficient (*r-pbi*)	Average of 0.35	≥0.20
Reliability index (*r-KR 20*)	0.77	≥0.70
Ferguson's delta (δ)	0.97	≥0.90

4.1. Item difficulty index

Item difficulty is a measure of the easiness of an item and is defined as the proportion of correct responses [12]. Practically, the acceptable values of the item difficulty index, ranging from 0.3-0.9. As shown in figure 2, the average value of the item difficulty index is 0.53. It means that the question on the RCET are neither extremely difficult or extremely easy. The items with difficulty index around 0.5 ideally have the highest reliability. Nevertheless, there are some items (question 4 "Q4" and 7 "Q7") that have difficulty index below 0.3. After both of the questions are re-examined, the context and their description is very clear. Thus, it can be caused by the students who are not familiar with one of the representation that is presented in the isomorphic items. This reason is supported by two isomorphic items with similar context and concepts which have a satisfactory difficulty index.

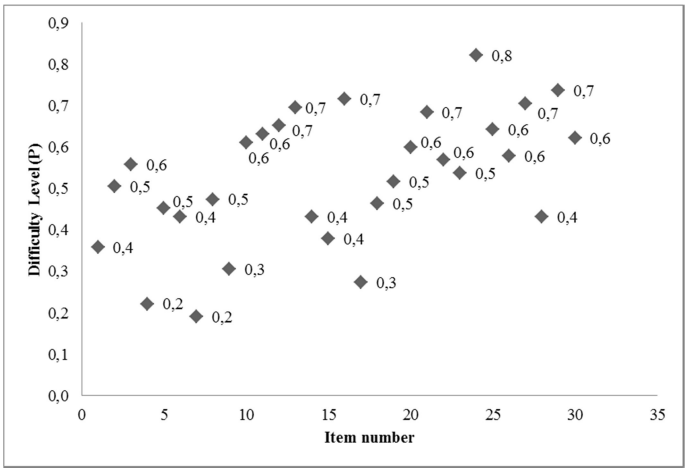

Figure 2. Difficulty index of each item on the RCET

4.2. Item discrimination index

The item discrimination index is a measure of the discriminatory power of an item [8]. It indicates how well the test discriminates students with different abilities. The acceptable value of the discrimination index should be equal or greater than 0.3. If the item has low discriminatory index, it may be caused by the description of the question that is unclear or related to its difficulty index,

whether too difficult or too easy. As shown in figure 3, the average value of the item discriminatory index shows highly 0.45. It means that overall the question on the RCET are able to distinguish low-achieving students and high-achieving students. Some of the questions with discrimination index below 0.3, related to the high or low difficulty index. It means that the question may be too difficult or too easy, so it is not able to discriminate students.

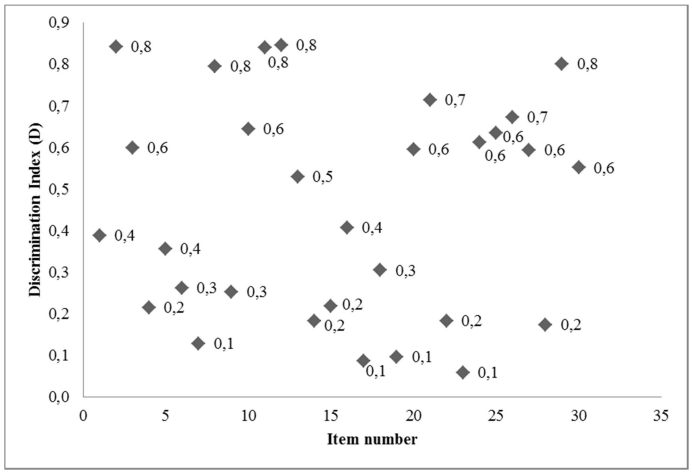

Figure 3. Discrimination index of each RCET item

4.3. Point biserial coefficient

The point biserial coefficient is a measure of individual item reliability, indicates how consistently an item measures students' performance in relation to the whole test [8, 12]. As shown in figure 4, almost all the test items have a value above 0.2 except eight items. The average point biserial coefficient for the RCET is 0.35, shows that the RCET items are reliable and consistent. The desired value is equal or higher than 0.20 [12]. Although, a low coefficient indicates that the item may not measure the same material at the same level, but the item in RCET is the isomorphic items which different only in the representations used. So, if there are some questions with low point biserial value, it was probably caused by students who are not familiar with those representations.

4.4. Kuder-Richardson reliability index

The quality of a test, one of which can be seen from the test reliability value. Reliability is defined as the overall consistency and stability of a test measure.

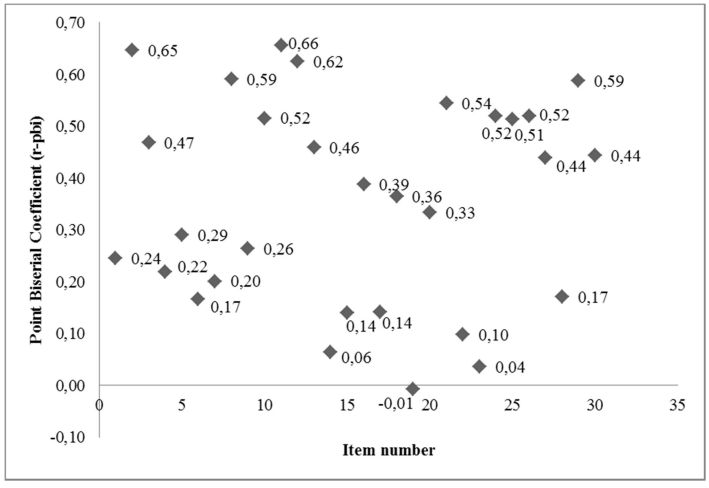

Figure 4. Point biserial coefficient of each RCET item

A widely used criterion for a reliable instrument is the coefficient of KR-20 index should be equal or higher than 0,7. The reliability coefficient of RCET is 0.77, it is slightly higher than the accepted standard value. It indicates that RCET has good internal consistency, each item on RCET measure the same characteristic and has high correlations between individual test items [8].

4.5. Ferguson's delta

The last evaluation of RCET is Ferguson's delta. It is defined as a measure of the discriminatory power of a test. The calculation of Fergussion delta for RCET gives value of 0.97, that means this test is good at discriminating among students at different levels. Generally, the desired value of Ferguson's delta is higher than 0.9.

5. Conclusion

The item analysis and the test analysis have been evaluated for RCET. The validity and reliability of RCET show that this instrument can be used to measure the intended concept in the first law of thermodynamics and it will give the consistent results. This instrument differentiates well between high-achieving students and low-achieving students and also students at different level. Overall evaluation of RCET gives positive results. Thereby, this instrument can be used to measure student understanding and a student's ability in using multiple-representations (such as representational and scientific consistency). As well as measuring the effectiveness of the learning process on the concept of the first law of thermodynamics. We are still in the process of improving the quality of the test by collecting data from various populations of university students.

Acknowledgments

We greatly thank the students and the faculty members, including A. Rusli, S. Karim, D. Saepuzaman, J. Rokhmat, S. Ayyubi, who have given many contributions in developing this instrument.

References

[1] McDermott L C and Redish E F 1999 *Am. J. Phys.* **67**(9) 755-767
[2] Meltzer D E 2004 *Am. J. Phys.* **72** (11) 1432-1446
[3] Loverude M *et al* 2002 *Am. J. Phys.* **70** (2) 137-148
[4] Leinonen R *et al* 2013 *Phys. Rev. ST Phys. Educ. Res.* **9** 0201122013
[5] Cochran M J and Heron P R L 2006 *Am. J. Phys.* **74**(8) 734-741
[6] Rozier S and Viennot L 1991 *Int. J. Sci. Educ.* **13**(2) 159-170
[7] Wattanakasiwich *et al* 2013 *Int. J. Innov. Sci. Math. Educ*. **21**(1) 29-53
[8] Nieminen P *et al* 2010 *Phys. Rev. ST. Phys. Educ. Res.* **6**(2) 020109(12)
[9] Meltzer D E 2005 *Am. J. Phys.* **73**(5) 463-478
[10] Ryan Q X *et al* 2013 *PERC Conf. 2014* (Minneapolis, MN)
[11] Cohen J 1988 *Statistical Power Analysis for the Behavioral Science 2nd ed* (Routledge, NY)
[12] Ding L and Beichner R 2009 *Phys. Rev. ST. Phys. Educ. Res.***5** 0201032009