What are the Impacts of Dynamic Representations on Students' Answering Behavior in the Two-tier Test in Electricity?

Jing-Wen Lin*, Yu-Lun Wu, Yen-Ching Lin
Department of Natural Science, Taipei Municipal University of Education, Taiwan
jwlin@tmue.edu.tw*

Abstract

This study adopted two formats of a two-tier test to diagnose sixth-graders' alternative conceptions in electricity. One was a static representation, and the other a dynamic representation. The major purposes of this study are to: (1) explore students' alternative conceptions or knowledge deficiencies; (2) investigate the impacts of dynamic representation to their answering behavior. Sixty-two students were matched by the score on an achievement test in electricity, and divided into two groups for the representation format. Each participant was individually interviewed using the two-tier test. A confidence rating was also used to indicate the strength of a student's conceptual understanding. This study identified six typical responses, four being robust alternative conceptions in high confidence. In addition, answering behavior might be changed by the dynamic representation. The confidence of the lower score students is encouraged, and one of the higher score students is decreased. Interview data showed the evidence that the lower score students might figure out the context of items with the assistance of dynamic representation. As well as the higher score students could obtain more hints from the inconsistency between their internal and external representations while answering. Finally, implications for the design of a two-tier test were discussed.

Key words: Answering behavior, Confidence, Dynamic representation, Electricity, Two-tier test

Background, Framework, and Purpose

A two-tier test (Treagust, 1988) has emerged as one of the most popular types of paper-and-pencil instruments for efficiently investigating large-scale students' conceptions. However, it cannot tell us students' mistakes are due to the lack of knowledge or their alternative conception (Hasan, Bagayoko, & Kelley, 1999). Therefore, more and more researchers suggested a confidence-based assessment, in which a student's rating of his confidence in a response, could be taken into consideration as a solution for this problem (e.g. Caleon & Subramaniam, 2010; Hasan et al., 1999). On the other hand, with the rapid technological advance, the employment of dynamic representations has become one of the essential current trends in computer-based assessment (e.g. PISA 2012). It is worth noticing that representation formats might influence students' answering behavior. Skopeliti and Vosniadou (2007) indicated that providing different representations would make pupils' internal representations inconsistent with the external ones. Furthermore, Lin and Wu (2011) used animation-based assessment to investigate 5th and 7th graders' alternative conceptions in electricity. They also found the employment of the dynamic representations changed students' answering behavior. For example, it increased students' scientific responses, and there was a significant high in 7th grade. They claimed the dynamic representations would be helpful for the younger students understanding the context of test items, while the older ones obtaining the more clues to guess the correct answers. Accordingly, this study based on Lin & Wu's study, and used "answering confidence rating" in advance to reflect the strength of students' conceptual understanding and internal, estimated belief in their own accuracy (Caleon & Subramaniam, 2010). The major purposes of this study are to: (1) explore 6th graders' alternative conceptions or knowledge deficiencies in electricity; (2) investigate the impact of dynamic representations on students' answering behavior, which include judging an answer, confidence in answering, understanding the context of test items, and the consistency between internal and external representations. It was predicted that the pattern of answering behavior on the higher score students would be similar to the one of the 7th graders, while the lower score ones would be like as the 5th graders in Lin & Wu's study.

Methodology

Participants and Procedures. The participants of the study were 62 6th graders in Taiwan. The detailed research procedure is illustrated in Figure 1.

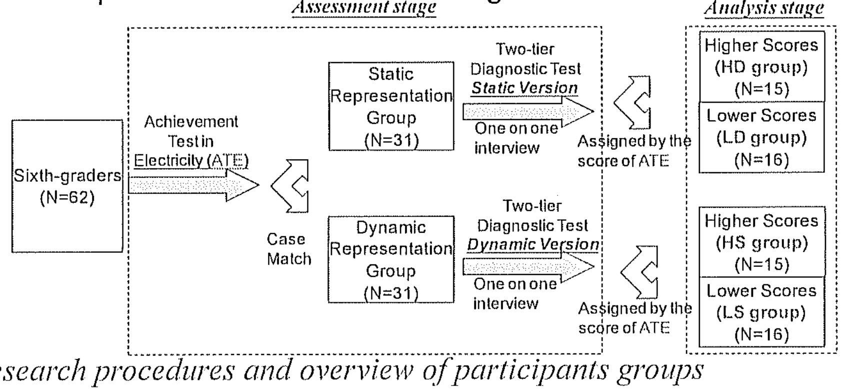

Figure 1 The research procedures and overview of participants groups

Instruments. Two instruments were adopted in this study: the achievement test in electricity (ATE), and the two-tier test. The ATE was developed by Liu (2010) according to the Grade 1-9 Curriculum Guidelines in Taiwan (Cronbach's alpha was .81). It was used for matching participants in the assessment stage, and dividing them into HD, LD, HS and LS groups in the analysis stage (Figure1). The two-tier test was revised according to Lin and Wu's study (2011). The first-tier of each item was related to the *science content*, while the second-tier presented some *reasons* for the given answer for the first tier. Two formats were designed, and the greatest difference was showed in the second tier. One was static representations (with pictures and texts, Cronbach's alpha was .66), and the other was dynamic representations (with voices and animations, Cronbach's alpha was .71). The content of two formats of the assessment was identical. Moreover, the 5-points Likert scale for each item was designed for understanding students' confidence level in answering.

Data analysis. In the quantitative data, the descriptive statistics were used to analyze all of 62 students' data at first. However, equal confidence level that in different difficulties items is questionable, 10 students' confidence data were deleted. That is, only 52 students' confidence data were analyzed. In addition, both score and confidence were then found an abnormal distribution by the Shapiro- Wilk test. Thus, U test was adopted to compare the differences of the scores and answering confidence between H and L groups in two assessment formats. As regarding to the qualitative data, all interview data were transcribed verbatim, and were categorized for inferring the impacts of the assessment formats from students' perspective (inter-rater reliability was .98).

Results

Alternative conceptions or knowledge deficiencies? As Peterson, Treagust, and Garnett (1989) recommended, a response combination whose percentage greater than 20 is a typical alternative conception. Table 1 displays 52 students' representative incorrect responses (excluded those students' data that confidence is equal in different items), their percentages and average confidences. Six typical responses were identified, and four in six (responses A, C, D and F) were with the higher confidence (greater than 3). It suggested these four responses might be robust alternative conceptions, and responses B and E might be knowledge deficiencies only. It deserves to be mentioned that the pattern of 62 students is identical with Table 1.

Table 1. Fifty-two students' typical incorrect responses and their confidences

Items	Representative incorrect responses	%	Confidence (Mean±SD)
1.The electricity current in a series connection	**A.** The direction of the electricity current is correct, but the first bulb takes up most of the electric current.	36.5	3.4±0.8
2.Open circuit in a series connection	B. One bulb lights in an open circuit. The electric current can flow from one of the battery poles to a bulb, but it cannot continue to flow to the other bulb.	25.0	2.9±0.9
	C. No bulb lights in an open circuit. The electric current can flow from one of the battery poles to a bulb, but it stops at the disconnected point.	23.1	3.8±1.0
3.The relationship between voltage and ampere	**D.** The size of the battery is bigger, and the bulb is brighter.	40.4	3.4±1.0
4.The relationship between	E. The length of wire influences the brightness and ampere in a circuit.	36.5	2.8±0.9

ampere and ohm	F. Bulbs take up the current. The more bulbs that the current passes through, the weaker of its strength.	36.5	3.3±1.1

Judging answers. For the judging answers, we used scores as an index. Table 2 indicates' scores in the group of dynamic representations are almost higher than the other group. Besides, there is a significant effect for the second-tier (reasons) in H group. Because this tier is full of dynamic representations, we presumed the students in H group might get more clues from voices and animations to judge answers. The result is identical to our prediction.

Table 2. Students' scores in the two-tier test with different representations

	Higher score group (H group)				Lower score group (L group)			
	Static (HS, 15)	Dynamic (HD, 15)	Z	*p-value*	Static (LS, 16)	Dynamic (LD, 16)	Z	*p-value*
Tiers	Mean Rank	Mean Rank			Mean Rank	Mean Rank		
Content	16.8	14.2	-0.908	.364	16.3	16.8	-0.163	.870
Reasons	12.5	18.5	-2.015	.044*	16.5	16.5	0.000	1.000
Total	13.9	17.1	-1.045	.296	16.1	16.9	-0.273	.785

Confidences. Table 3 shows a pattern that the HS group gets higher confidence than the HD group. Besides, it reveals significant differences on the third item and the total ranking.. However, the pattern reverses in the L group. That is, LD group gets the higher confidence than the LS group, and it shows a significant difference on the first item. Accordingly, we inferred that dynamic representations might reinforce the confidence of L group but discourage the one of H group.

Table 3. Fifty-two students' answering confidence about the test with different representations

	Higher score group (H group)				Lower score group (L group)			
	Static (HS, 12)	Dynamic (HD, 13)	Z	*p-value*	Static (LS, 13)	Dynamic (LD, 14)	Z	*p-value*
Items	Mean Rank	Mean Rank			Mean Rank	Mean Rank		
1	14.8	11.4	-1.216	.224	10.6	17.2	-2.341	.019*
2	15.4	10.8	-1.652	.099	13.1	14.9	-0.619	.536
3	16.0	10.2	-2.030	.042*	13.7	14.3	-0.205	.838
4	15.6	10.6	-1.782	.075	12.4	15.5	-1.069	.285
Total	16.4	9.9	-2.220	.026*	12.0	15.9	-1.272	.203

The impact of dynamic representations on students' understanding the context of test items, and the consistency between internal and external representations. In this section, we explored students' answering behavior further from their interview data. The results are as Table 4. As for the understanding the context of test items, students in the LD group held the most positive attitude. This result echoed with our predication, and we inferred that might be the reason for the students in the LD group held the higher confidence than the LS group (Table3). As a regard to the impact of the consistency of internal and external representations, it revealed the students in the HD group had the lowest consistency. We inferred that might be the reason for the students in the HD group held the lowest confidence.

Table4. The impact of H and L group with different test formats on several answering behavior

	Higher group (H group)						Lower group (L group)					
	Static (HS, 15)			Dynamic (HD, 15)			Static (LS, 16)			Dynamic (LD, 16)		
	Yes	No	Not mention	Yes	No	Not mention	Yes	No	Not mention	Yes	No	Not mention
Understanding meaning (%)	20.0	6.7	73.3	51.7	3.3	45.0	35.9	10.9	53.1	62.5	14.1	23.4
Consistency (%)	71.7	28.3	0.0	66.7	33.3	0.0	70.3	29.7	0.0	73.4	25.0	1.6

Conclusions and Implications

Several research questions were addressed in this study, and the principal findings identified six typical responses, and four in six were deep-rooted alternative conceptions in high confidence. Besides, dynamic representations might be helpful for the lower score students to understand the context of test items and the higher score students to obtain more clues in judging answers. Furthermore, according to Lin & Wu's study (2011), we referred that dynamic representations might enhance the lower score students' answering confidence for increasing the understanding the context of test items, but discourage the higher score students' one for decreasing the consistency of students' internal and external representations. These results could be taken into account on employing dynamic representations in computer-based assessment design. And using confidence rating in a two-tier test was more powerful in diagnosing alternative conceptions.

Bibliography

1. Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves, *International Journal of Science Education, 32*(7), 939 -961.
2. Hasan, S., Bagayoko, D., & Kelley, E. L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education*, *34*, 294-299.
3. Lin, J. W., & Wu, Y. L. (2011). Investigating the impact of dynamic representations with interactive response system on diagnosing students' alternative conceptions of simple and series connections. *Journal of Education & Psychology, 34*(1), 79-107.
4. Liu, Y. C. (2010). *The impact of multiple representations of examination to diagnoses the student's simple and the series circuit mental model* (Unpublished master's thesis). Taipei Municipal University of Education, Taipei, Taiwan.
5. Peterson, R. F., Treagust, D. F., Garnett, P. J. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching, 26*, 301-314.
6. Skopeliti, I., & Vosniadou, S. (2007). Reasoning with external representations in elementary astronomy. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the European Cognitive Science Conference 2007* (pp. 244-249). London: Taylor and Francis.
7. Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 10*, 159-169.