Designing and Validating an On-line Dynamic Assessment for Improving Reading Comprehension on Science

Jing-Ru WANG

Graduate Institute of Mathematics and Science Education, National Pingtung University of Education, Taiwan

mail100@mail.npue.edu.tw

Abstract: This study designed and validated a dynamic assessment e-learning environment for enhancing elementary students' science reading comprehension. The "graduated prompt approach" was used to develop the assessment module. The number of participants for the formal administration was 480 grade 5 and 6 students. Statistical results indicated the two versions of DARCS had concurrent validity. The items had good discrimination power and difficulty appropriate for the majority of the students across grade 5 and 6 level and no statistical difference between these two versions. The implications of the study were discussed.

Keywords: Graduated dynamic assessment, dynamic assessment, science reading comprehension

Introduction

The alignment of assessment, learning, and instruction implies the effectiveness of how well the lesson was designed and taught, and how well students achieved the full range of desired educational goals [1]. However, without providing appropriate feedback, traditional static assessment only labels students' level of performance without providing instant supports for moving up to the next levels of performance. Different from static assessments, dynamic assessment (DA) integrates assessment and support for raising learner's abilities[2].

Brown's graduated prompt approach is one model of DA that used a set of standardized hints and guiding questions arranging from most implicit to most explicit and ends with the correct answer [3]. Based on the whole learning sessions, The researcher can examine the metric of learning efficiency that indicate the time period they can learn the new patterns and how far they can transfer this knowledge to novel tasks [3]. According to their definition, The metric of learning efficiency refers to the number of hints needed to reach the learning criteria. As the smartphone has morphed into instant messenger and has become prevalent, more and more digital devises are available to most students, there is a need to apply the potential benefit of e-learning environment to improve students' reading ability by designing learning tasks that take advantage of what technology can efficiently and well graphing, visualizing and computing [4]. Hence, the present study used Brown's graduated prompt approach to design on-line reading tasks aiming to raise students' reading proficiency on science texts.

Methods

1. Designing On-line DARCS Test

The process of designing and pilot study of the DARCS was reported in author's previous report in the 2012GCCCE conference [5]. The validated draft of the instrument involved 14 units with 69 items which was equally divided into two versions: test A for grade 5 students and test B for grade 6 students. In addition, there were two units with 8 items serviced as common items. Each textual passage was followed by connected items which were followed by the aid of three sequential prompts: promp1 rephrased the question in an oral form, prompt 2 showed the reading strategy, and prompt3 presented the correct answer derived from the reading strategy. If an examinee chooses the incorrect answer, the prompt will appear from implicit hint to explicit instruction. All of the questions were developed to reflect the four components of science reading comprehension [6]. The examinee can choose the option of sound reading to help comprehend the reading content. After finishing the text items, an assessment report will be presented on the screen showing the result of the student's reading performances, including the length of the time spending on reading the text and textual items, the number of instruction prompts used for each item, and the overall diagnostic results and comments. DARCS Test was installed under the website called Web-Based Diagnostic Instruction (WBDI) for Science Literacy, which was designed using PHP and HTML as frameworks for web users. The DARCS system was designed using Adobe Flash CS5.5 that automatically stored student's individual data and test responses e in a database service called My Ess Que Ell (MySQL). The MySQL automatically transformed and stored the data as excel file for latter retrieval.

2. The Student Samples, Administration and Data Analysis

For the formal administration, randomly selected reflected the proportion of 10 elementary schools was used for the formal administration. The number of schools in the two southern Taiwan school districts; seven schools were from Kaohsiung city and three schools from Pingtung county. One class from both grade 5 and grade 6 was randomly selected for each school, and all students of those classes were administered the DARCS. The total number of students was 480 (253 males, 227 females).

Figure 1. Text, item and prompts of one unit of DARCS

Figure 2. Performance reports on one unit of DARCS DARCS

Test A was administered to 233 (122 males, 111 females) grade 5 students, and test B without IP was administered to 247 (131 males, 116 females) grade 6 students. A retest was conducted 2 weeks later with subsamples of original students. Four classes of students from each Grades 5 and 6 group ($n = 128$, 66 males & 62 females; $n = 138$, 74 male & 64

females, respectively) were randomly selected as the subsamples for retest. Both classical test theory (CTT) and item response theory (IRT) approaches were used in the study to analyze the characteristics of the test and features of the items. In order to explore the criterion validity of the instrument, all the students completed the RCST [6] measuring their comprehension of science information. T-test was used to explore the statistical differences between DARCS and RCST, and differences across grade levels.

Results and Discussion

The results of psychometrics for the DARCS test indicate the internal consistency of the whole test, estimated by the Cronbach α coefficient, was reliable (A = 0.84, B = 0.85); also, the test-retest reliability revealed the test was 82% and 83%, respectively, stable. The correlation between DARCS and RCST was 0.75 ($p<$ 0.01) for grade 5 version and 0.67 ($p<$ 0.01) for grade 6 version, indicating the efficacy of DARCS for measuring the students' comprehension of science information.

Diagnostic tests need to consider the difficulty of an item and whether it allows the more knowledgeable students to perform better than those who are less knowledgeable. The results of CCT showed that the item difficulties for the whole test averaging 0.55 (range 0.20~0.85) and discrimination averaging 0.45 (range 0.20~0.69). For the results of IRT, the estimated discrimination indices (a) ranged from 0.32 to 2.02 with a mean value of 1.11, the estimated difficulty indices (b) ranging from -1.49 to 2.22 with a mean value of 0.52. The guessing parameters (c) ranged from 0.15 to 0.35 with a mean of 0.23. These results indicated the items had good discrimination power and had difficulty appropriate for the majority of the students across grade 5 and 6 level. In IRT, a test information function (TIF) is developed to show the degree of precision in parameter estimation at various points along the score distribution; it is an average of item information at each trait level. The TIF curve implies that the test was appropriate for most examinees whose ability was at 0.8.

Figure 4 shows the comparison of the TIF between grade 5 and grade 6, in which more information was available on the midpoint and upper level of the score scale to make an accurate parameter estimate than was available at the lower point of the distribution (Figure 3). The comparison of the two versions of the DARCS indicated there is no difference between the two versions. In other words, for this age of students, science reading ability was not affected by grade level.

The draft of the online DARCS would benefit both learning and teaching of science reading comprehension because it provides instant feedback guiding reader toward the aims of reading comprehension, as well as comments and suggestions to the reader regarding his or her relative ability compared to that of the population. In addition, the changing patterns of time laps of reading text and items, and the number of instructional prompts for each components of reading comprehension would provide the evidences of mental models regarding science reading process.
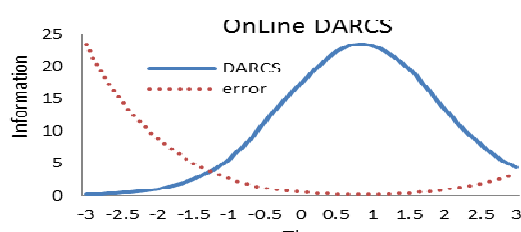

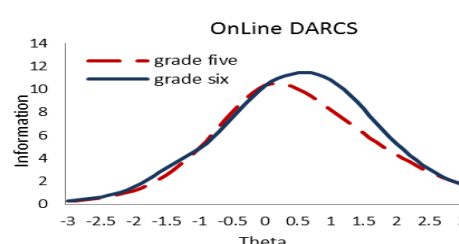**Figure 3. Test information function and standard error of the DARCS (based on 69)**

Figure 4. Comparison of Test information function for the two versions of the DARCS

Acknowledgment

References

[1] Bransford, J. D., & Donovan, M. S. (2005). Scientific inquiry and how people learn. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 397-419). Washington, D.C.: The National Academies Press.

[2] Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin, Germany: Springer.

[3] Brown, A., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 273-305). Cambridge: Cambridge University Press.

[4] National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

[5] Wang, J.-R., & Lin, S.-W. (2012). *Pilot study of using formative assessment e-learning environment for science reading comprehension.* Paper presented at the The 16th Global Chinese Conference Computer in Education, Kenting: Taiwan.

[6] Wang, J.-R., Chen, S.-F., Tsay, R.-F., Chou, C.-T., Lin, S.-W., & Kao, H.-L. (2012). Developing a test for assessing elementary students' comprehension of science texts. *International Journal of Science and Mathematics Education, 10*(4), 955-973.