THE DEVELOPMENT OF STATISTICAL LITERACY ASSESSMENT TOOL FOR SENIOR STUDENTS IN ELEMENTARY SCHOOLS

Chien-Ming Cheng

National Academy for Educational Research/ Research Center for Testing and Assessment (Taiwan)

Abstract

The most critical capabilities of the twenty-first century are: learning and innovation ability, information, media and technology literacy, work and life skills. Key competencies in information literacy included data interpretation and the prediction of possibility, which has a critical impact for the decision of problem solving. However, analysis, interpretation of data, prediction and judgment are statistical literacy. To cultivate the ability of data analysis, interpretation, and the prediction and judgment of event probably occurring is the prior knowledge of statistical literacy in elementary school. Therefore, this study used data and chance as item content to compile a set of crossing grade assessment with reliability and validity. The assessment and its parallel norm tests suit to evaluate the students' statistical literacy development in the fifth and sixth grade.

The research results are as follows: 1. This study applied the Rasch model to explore the validity of the "Statistical literacy assessment tool". The results found that most of the items were goodness of fit.

2. This study, by concurrent estimating the difficulty of all the items and the ability of the students, showed that the ability of the students of different grades was significantly different, and the higher the grade, the higher the ability of the students.

3. This research is the parallel design of the test items. It can be used as a longitudinal study assessment tool to explore the statistical literacy of students in the fifth and sixth grades, and can reduce the threat of validity for repeating the test.

Keywords: Statistical literacy, assessment tool, elementary school.

1. Introduction

a. The Partnership for 21st Century Skills (P21) believes that the key capabilities most needed in the 21st century are: the ability to learn and innovate, information literacy, media and technology literacy, work and life skills (Liu, 2012).

b. No matter what occupation, correctly understanding and presenting of chart information, there are already existing every corner of life and workplace.

c. There are only a few articles and project reports on the development of students' statistical literacy concepts, indicators and tests in Taiwan.

d. The survey results of PISA, TIMSS and TASA found that the statistical literacy scores of Taiwan's primary and secondary schools is relatively weak compared to the whole (Lin, Ren, Li, Lin, Chang, Cao, Yang, 2014; PISA in Taiwan, 2014).

2. Objectives

Therefore, the research is to compile a set of test and parallel test for assessing students' statistical literacy-data and chance ability and understanding the development of students' data and chance ability in the fifth and sixth grades. It is also an important tool for assessing curriculum development, teacher instruction, and students' data and chance ability. The followings are the research's objectives:

a. Administering the items of the test and all the parallel tests in order to establish a norm test and confirm the reliability and validity.

b. Administering the criterion-related test - mathematical achievement test and analyzing the validity of the criterion.

c. Analysis of the overall status of the sample students' data and chance ability.

d. Analysis of data and chance ability, and other sub-projects - the relationship between mathematical vocabulary, mathematical writing and computing ability in order to understand the important factors related to the ability of data and chance.

3. Methods

3.1. The test development process

a. Establishing statistical literacy connotation

b. Establishing a statistical literacy assessment framework
c. Setting the item type and length
d. Designing and revising procedures
e. Test compiling and data collection
3.2 Rasch model analyzing "statistical literacy test"
3.3 The research objects
For the fifth and sixth grade students, the sample statistics are as follows:

Table 1. Pre-test statistics description for the number of people in each grade.

grade	five	six	total
number	514	523	1037

Table 2. Formal test sampling description table.

grade	five		six	
area	number	proportion	number	proportion
north 1	20	0.04	137	0.27
north 2	105	0.22	85	0.16
center	209	0.44	80	0.16
south	121	0.25	178	0.34
east+distance	23	0.05	36	0.07
合計	478	1.00	516	1.00

3.2. Data analysis

3.2.1. Rasch model. Since the test includes binary and multivariate scoring items, the Rasch partial credit model (Master, 1982) is adopted.

3.2.2. Validity analysis. The validity certificate is established from the two aspects of the test content and the internal structure.

3.2.3. Reliability analysis. This study used subject separation reliability and EAP reliability.

4. Results

4.1. Fifth and sixth grade assessment framework

Table 3.

content	Concept understanding	Procedure executing	Solving and reasoning	total
Data collation and presentation	3	2	0	5(23.8%)
Data reading and interpretation	2	4	4	10(47.6%)
possibility	3	1	2	6(28.6%)
total	8(38.1%)	7(33.3%)	6(28.6%)	21(100%)

4.2. Validity evaluation results

4.2.1. Construct validity. Rasch model goodness of fit item goodness of fit except for the 3 questions in Volume 2, the MNSQ of all the other items fell between 1 ± 0.3.

Table 4.

Volume	Ave. b	Area b	MNSQ	# outlier of MNSQ
1	-0.22	-2.04~1.00	0.74~1.17	0
2	-0.35	-1.98~1.30	0.65~1.33	3
3	-0.10	-2.00~1.20	0.79~1.21	0

Estimated difficulty of each test in each year

Using the good anchor items' parameters which were set up in the previous pre-test, the answers of all the students in the two grades are concurrently estimated, so that the student's ability and the test difficulty can be estimated under the same scale. The average difficulty of the test questions is -.22, which means that for the whole student, the test items are median and tend to easy.

Vertical scale and item calibration

The researcher used one-way ANOVA to compare differences in data and chance abilities between students in the two grades. The analysis showed significantly different between two years, sixth grade (M = .51, SD =. 79) significantly higher than the fifth grade (M = .23, SD = .70).

Expert validity

The expert validity adopted in this research, in the assessment framework part, is through expert consultation meeting. There are 5 experts and scholars, including 2 mathematics education, 1 special education, 1 mathematics cognition and learning problem diagnosis, 1 psychometrician to confirm and build the assessment framework. The validity of the test items is confirmed after discussion and revision based on 6 elementary teachers with more than five years of teaching experience, and 1 psychometrician doctor.

4.3. Reliability analysis result

Table 5. Subject separation and EAP reliability table.

grade	# items	Subject separation	EAP reliability
5	103	0.98	0.71
6	103	0.98	0.75

4.4. Simultaneously measure the validity of the data and possibility ability, and analyze the criterion-related validity

Table 6.

		5 grade	6 grade
Math vocabulary	Pearson correlation	.57**	.69**
	Number	201	200
Math words	Pearson correlation	.60**	.72**
	Number	193	217
Math calculation	Pearson correlation	.49**	.56**
	Number	184	189
Statistical literacy	Pearson correlation	1	1
	Number	478	509
Math achievement	Pearson correlation	.71**	.68**
	Number	185	205

**. p<.01

5. Discussion

Although the overall reliability and validity are good for the norm test tool. There are still some findings:

Although the parallel tests have been pre-tested and analyzed to select good items for assembling items, and take the spiral test during the test. However, the post-test analysis still finds that it will be affected by the samples, resulting in seemingly parallel items whose parameters are still different. In addition, the average difficulty of the whole test is still different. Because of the budget constraints, the samples to be administered are still limited. If a larger samples and tests can be carried out, the results of parallel tests may be better.

The reliability analysis of fifth grade is .71, which may be related to the development of the students, and because the teacher is entrusted to help the test. There is only a description of the paper; there is no training and explanation for the supervision, so the reliability is slightly lower.

References

Liu, X. H. (translated) (2012). *The future of education - the key capabilities we need*. (Original author Trilling, B. & Fadel, C. 21st Century Skills: Learning for Life in Our Times). Wild Goose Culture: Taipei City.

Lin, C. Y., Ren, Z. H., Li, Z. D., Lin, B. Z., Zhang, M. Y., Cao, B. S., & Yang, W. J. (2014). *The Trends in international Mathematics and Science Study 2011*. Taipei City: Normal University Science and Education Center.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

PISA in Taiwan (2014). *Taiwan PISA2012 streamlined report*. Retried from http://pisa.nutn.edu.tw/download_tw.htm