

*Full Length Research Paper*

# Multidimensional computerized adaptive scholastic aptitude test program used for grade 9 students under different reviewing test conditions

Naruemon Khunkrai<sup>1\*</sup>, Tatsirin Sawangboon<sup>2</sup> and Jatuphum Ketchatturat<sup>3</sup>

<sup>1</sup>Educational Research and Evaluation, Faculty of Education, Mahasarakham University, Thailand.

<sup>2</sup>Department of Education Research and Development, Faculty of Education, Mahasarakham University, Thailand.

<sup>3</sup>Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University, Thailand.

Received 15 February, 2015; Accepted 22nd June, 2015

The aim of this research is to study the accurate prediction of comparing test information and evaluation result by multidimensional computerized adaptive scholastic aptitude test program used for grade 9 students under different reviewing test conditions. Grade 9 students of the Secondary Educational Service Area Office in the North-east of Thailand, in 2014 academic year were the sample used in this research. The research materials were two test programs: the test program that allows reviewing of answers and the one that does not allow reviewing of answers. The manual of the test program and evaluation form of the multidimensional computerized adaptive scholastic aptitude test program used for grade 9 students in different reviewing test conditions showed that: 1) The test program is an accurate predictor of the students' achievement, as verbal factor correlates with the students' achievement in five core subjects. The aptitude tests on number and reasoning correlate with the students' achievement in Mathematics, science and social science. The aptitude test on space correlates with the students' achievement in Mathematics and science. 2) The test program has statistical and significant difference at .05 level. 3) The evaluation result of the two test programs has statistical significance at .05 level.

**Key words:**Scholastic aptitude, multidimensional, computerized adaptive test program, test programs.

## INTRODUCTION

The work studies grade 9 adolescent students, who have many problems to deal with. Most students choose to study what their friends study, regardless of their aptitude. This corresponds with many research works that have discovered that aptitude influences one's decision making on what to study in future. Pha-on (2010) studied the

factors that affected the exciting learning of grade 9 students under Primary Education Service Area Office in Saraburi. In the study, it was found out that the highest average was obtained by the students who chose to study based on their aptitude. As a result, education management must provide content and learning activities

\*Corresponding author. E-mail: [dear.dear888@gmail.com](mailto:dear.dear888@gmail.com).

that are in line with the aptitude and differences of the individuals in order to make the students perfect all way round, be able to find a job and live happily with other people. Each school must provide flexible learning process for students to choose according to their aptitude and interests (Ministry of Education, 2010). However, aptitude is an innate ability that cannot be directly realized, so testing theory is used as a tool to scale aptitude.

Scholastic aptitude test, mostly made and developed from Multi-Factor Theory by Thurstone, is used to measure students' academic ability. This results from the processing of knowledge and experience gained by students which in turn will enable them to have success in their choice field of study as well as success in their future occupation. Moreover, the test is also used in qualifying examinations, classifying students, diagnosing their capabilities, measuring the development of learners, predicting success, comparing intelligence, evaluating school-record and research. The research showed that scholastic aptitude correlated with students' academic achievement and the ability to predict scholastic achievements, which were mostly verbal, number, reasoning and space factors (Loard and Nicely, 1997; Morton, 2004). The test where students mostly write on answer sheet is called paper and pencil test or conventional test. It had several weak points, and so theorists developed and reformed it from conventional test to modern test theory.

Currently, computerized adaptive test is based on item response theory. Multidimensional computerized adaptive test is implemented in each dimension separately, and each dimension has a high relevance, that is, using multidimensional item response theory. Segall suggested that this theory should be used for selection of limited number for each content. This is consistent with the teaching of today, which focuses on integrating more knowledge. So, measurement is intended to change in part with a focus on measurements of performance that are more complex (Junpeng, 2007). Computerized adaptive test still has issues that need to be discussed. After the test is completed, the examinees should be given the opportunity to review answers. Vispoel (1998), Olea et al. (2000) Vispole et al. (2000) and Revuelta et al. (2003) found out that most examinees like to review answers in order to reduce anxiety. After reviewing the answers, the number of correct answers will increase; but researchers, like Wainer et al. (1993) and Green et al. (1984) have proposed it, stressing that opportunities should not be given for reviewing answers due to limited time of the test process. Previously in the country, there were only a few studies on unidimensional computerized adaptive test. One of such was that of Pimsiri (2006) who found that there was no difference between the examinees who had high level of ability and those with average ability under conditions of not reviewing and reviewing answers. But there was statistical significant

difference (.05 level) between examinees with medium and low level ability and examinees with average ability under the conditions of not reviewing and reviewing answers.

The researcher is interested in developing the multi-dimensional computerized adaptive scholastic aptitude test program under different reviewing test conditions, focusing on verbal, number, reasoning and space factors. In addition, this program will guide further study of both general and vocational education consistent with testing the ability of individual examinees. It will be used as information for further study and as a guide for students to realize their aptitude. It will also be useful for students who are going to graduate from grade 9 and as a method for developing multidimensional computerized adaptive test in details.

### Research questions

1. How are the test program that allows reviewing of answers, and the one that does not allow reviewing of answers accurate predictors ?
2. Are they any differences between information of the test program that allows reviewing of answers and the one that does not allow reviewing of answers?
3. Do the examinees have different opinions about the test program that allows reviewing of answers and the one that does not allow reviewing of answers?

### The purpose of the study

1. To study the accuracy of Multidimensional Computerized Adaptive Scholastic Aptitude Test Program to predict
2. To compare the test information of multidimensional computerized adaptive scholastic aptitude test program used for grade 9 students under different reviewing test conditions.
3. To compare the evaluation of multidimensional computerized adaptive scholastic aptitude test program used for grade 9 students under different review test conditions.

### METHODOLOGY

#### Materials

The materials used include 1, multidimensional computerized adaptive scholastic aptitude test program for grade 9 students; it has 2 programs: program that allows reviewing of answers and program that does not allow reviewing of answers.

2. Manual of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test conditions.
3. Evaluation form of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under varying reviewing test conditions.



Figure 1. Developing the test bank of scholastic aptitude.

**Development of materials used for data collection**

Multidimensional computerized adaptive scholastic aptitude test program for grade 9 students: this program selects item based on the ability of each examinees; that is, each examinee is given different items. The examinee gets 1 score if the answer is right and 0 score if the answer is wrong. See the procedures below:

1. Program design includes structure and application component.
2. Building program flowchart includes input, process and output.
3. In selecting language for development test program, the researcher selected Visual Basic .Net 2010 and Microsoft SQL Server 2008.
4. Coding
5. Trial program was used to check authenticity primarily by the researcher.
6. Experts verify the quality of the test program.
7. The recommendation of experts was improved and a complete test program was developed.

Manual of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test condition: The procedures are as follows:

1. Manual of program design includes information about the program, objective and utilization of the program, preliminary agreement, installations, running of the program, and definitions of specific terms.
2. Preparing the manual of the program
3. Experts verify the quality of the manual
4. The recommendation of the experts was improved and the complete manual was developed.

Evaluation form of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test condition: The procedures are as follows:

1. Learning and synthesizing how to create evaluation form related documents.
  2. Building evaluation form using psycho-social criteria given by Sympon; they consist of 4 aspects : 1) the statement and how to perform the exam, 2) Attraction of test program, 3) The anxiety in the test and 4) The general opinion about test using computer program
  3. Experts verify the quality of evaluation form
  4. The recommendation of the experts was improved and a complete evaluation form was developed.
- Data collection is divided into 4 steps as follows:

- Step 1: Developing the test bank of scholastic aptitude (Figure 1).
- Step 2: Developing multidimensional computerized adaptive scholastic aptitude test program (Figure 2).
- Step 3: Using the multidimensional computerized adaptive scholastic aptitude test program (Figure 3).
- Step 4: Evaluating the multidimensional computerized adaptive scholastic aptitude test program (Figure 4).

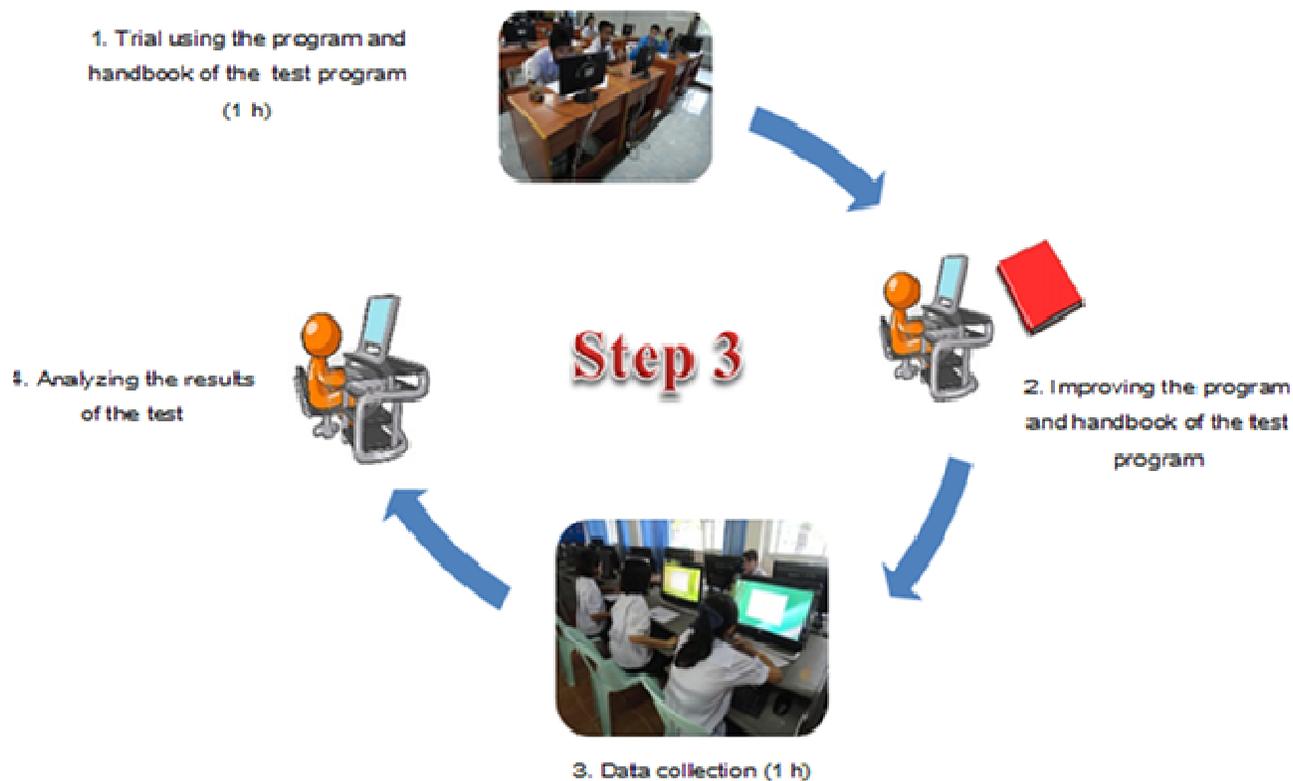
**Data analysis**

To study the predictive accuracy of the test programme, Pearson Product moment correlation was used to analyze the relationship between the mean of the examinees' total ability from the test program and mean of the examinees' average ability from school record.

$$R_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$



**Figure 2.** Development of multidimensional computerized adaptive scholastic aptitude test program.



**Figure 3.** Using the multidimensional computerized adaptive scholastic aptitude test program.

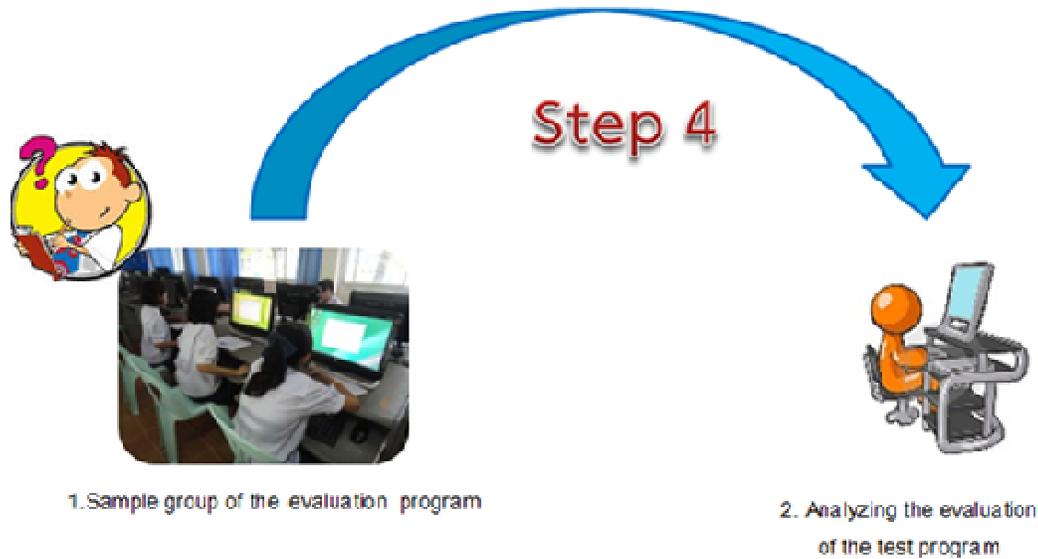


Figure 4. Evaluating the multidimensional computerized adaptive scholastic aptitude test program.

where  $R_{xy}$  is correlation coefficients between  $X$  and  $Y$ ,

$X$  = the examinees' total ability from test program

$Y$  = the examinees' ability from average school record

$\sum XY$  = sum of product between  $X$  and  $Y$

$\sum X$  = sum of each examinee's total ability from test program

$\sum X^2$  = sum of squares of each examinee's total ability from test program

$(\sum X)^2$  = sum of squares of each examinee's total ability from test program

$\sum Y$  = sum of each examinee's total ability from average school record

$\sum Y^2$  = sum of squares of each examinee's total ability from average school record

$(\sum Y)^2$  = sum of squares of each examinee's total ability from average school record

$N$  = all of the examinees

Independent sample t-test was used to compare and analyze the test information; its formula is given below (Hambleton and Cook, 1977 : 66):

$$I(\theta) = \sum_{i=1}^k I_i(\theta) \quad , i = 1, 2, \dots, k$$

$$I_i(\theta) = \frac{2.89a_i^2(1 - c_i)}{\left[ c_i + e^{1.7a_i(\theta - b_i)} \right] \left[ 1 + e^{-1.7a_i(\theta - b_i)} \right]^2}$$

where  $I(\theta)$  is test information at ability  $\theta$

$I_i(\theta)$  = item information at ability  $\theta$

$k$  = number of questions in the test

$a_i$  = discrimination parameter of item  $i$

$b_i$  = difficulty parameter of item  $i$

$c_i$  = guessing parameter of item  $i$

$e$  = natural logarithm ( $e = 2.72$ )

One-way MANOVA was used to compare and evaluate the test program as follows.

Mean

$$\bar{X} = \frac{\sum X}{N}$$

where  $\bar{X}$  is mean

$\sum X$  = sum of data

$N$  = number of all data

Standard Deviation

$$S.D. = \sqrt{\frac{N\sum x^2 - (\sum x)^2}{N(N-1)}}$$

where  $S.D.$  is standard deviation

$x$  = each data

$x^2$  = squares of data

$N$  = number of all the data

**Table 1.** Results of the correlation between mean of the examinees' total ability from test program with mean of the examinees' ability from average school record.

Academic achievement	Test program	Aptitude component				
		Verbal	Number	Reasoning	Space	Total
Thai	Allows reviewing of answer	.859**	.173	.058	.270	.796**
	Does not allow reviewing answers	.680*	.147	.148	.177	.399*
Mathematics	Allows reviewing of answers	.766*	.856**	.802*	.858*	.759*
	Does not allow reviewing of answers	.644*	.774*	.596*	.600*	.372*
Science	Allows reviewing of answers	.676*	.852*	.755*	.872*	.729*
	Does not allow reviewing of answers	.501*	.605*	.676*	.899*	.342*
Social	Allows reviewing of answers	.662*	.604*	.689*	.130	.618*
	Does not allow reviewing of answers	.420*	.430*	.278**	.076	.300*
English	Allows reviewing of answers	.860*	.152	.155	.257	.599*
	Does not allow reviewing of answers	.362*	.098	.067	.080	.285*
Grade average	point Allows reviewing of answers	.732*	.855*	.835*	.848*	.757*
	Does not allow reviewing of answers	.733*	.679*	.775*	.658*	.608*

\* statistical significance at .05 level; \*\* statistical significance at .01 level

**Table 2.** The comparison result of information test form of multidimensional computerized adaptive scholastic aptitude testing program under varying review test conditions.

Program	N	$\bar{X}$	S.D.	t	Sig.
Allows reviewing of answers	79	637.82	199.72	4.685*	.000
Does not allow reviewing of answers	79	511.35	132.98		

\* statistical significance at .05 level

## RESULTS

Results on the predictive accuracy of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test conditions showed that the test program was accurate in predicting the students' achievement (Table 1). Verbal factor was related with the students' achievement in 5 core subjects. Number factor and reasoning factor had relation with the students' achievement in Mathematics, Science and Social Science but had no relation with the students' achievement in Thai and English.

Space factor was connected to the students' achievement in Mathematics and Science but had no relation with the students' achievement in Thai, Social and English subjects.

The results of comparison in test information form from multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test conditions: The researcher did a random sampling of the two research programs, which have the

same amount. The result showed that multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different reviewing test conditions had different test information form with statistical significance at .05 level. The test program that allows reviewing test condition has average test information higher than the test program that does not allow reviewing test condition (Table 2).

Results of comparing evaluation of multidimensional computerized adaptive scholastic aptitude test program for grade 9 students under different review conditions: the result of the two programs was statistically significant at .05 level. When analyzing the issue of evaluation in 4 parts, the result of explanation and test operation in terms of general opinion about using computer program for test was statistically significant at .05 level. In three sections, the test program that allows reviewing of answers has high average greater than the test program that does not allow reviewing of answers. For assurance of the test program, there is no difference in the evaluation as shown in Tables 3-4.

**Table 3.** The results using one-way MANOVA for comparison of evaluation of the test program

Statistic	Value	F	Hypothesis df	Error df	Sig.
Pillai's Trace	.157	7.10*	4.00	153.00	.000
Wilks' Lambda	.843	7.10*	4.00	153.00	.000
Hotelling's Trace	.186	7.10*	4.00	153.00	.000
Roy's Largest Root	.186	7.10*	4.00	153.00	.000

\*statistical significance at .05 level

**Table 4.** The results of using one-way MANOVA to make comparison in evaluation test program.

Issues of evaluation	Test program	N	$\bar{X}$	S.D.	F	Sig.
The statement and how to perform the exam	Allows reviewing of answers	79	4.27	.56	16.198*	.000
	Do not allow reviewing of answers	79	3.94	.64		
Attraction of test program	Allows reviewing of answers	79	4.20	.61	14.514*	.000
	Do not allow reviewing of answers	79	3.86	.71		
Anxiety in the test	Allows reviewing of answers	79	2.79	.83	.722	.397
	Do not allow reviewing of answers	79	3.13	.83		
The general opinion about testing by computer program	Allows reviewing of answers	79	4.09	.57	22.746*	.000
	Do not allow reviewing of answers	79	3.65	.67		
Overall	Allows reviewing of answers	79	3.95	.42	24.429*	.000
	Do not allow reviewing of answers	79	3.64	.46		

\* Statistical significance at .05 level

## DISCUSSION

The test program was an accurate predictor of the students' achievement. The computerized adaptive scholastic aptitude test was appropriate for testing the examinees' ability. So, this result is consistent with the reality that verbal factor affects learning of general communication, understanding the meaning of conversation, listening to explanation and reading of the main idea for comprehension of each subject. Chinese aptitude is related with 5 core subjects, while quantity aptitude is related to number, calculation and reasoning factors, which are the basic characteristics necessary for the students. Therefore, these two aptitude factors correlate with Mathematics, Science and Social Science. Space factor is the basic factor that affects imagination and vision, which are connected to Mathematics and Science. Moreover, the sample in this research had moderate to high ability and cooperated in the test as well. So, item response showed a careful reflection. The result of this study is in line with the study of Philatun and On-uma (2008: 77) who found that the good predictor of students' Mathematics achievement involved the students' aptitude such as number, space, reasoning and verbal factors.

Multidimensional computerized adaptive scholastic aptitude test program for Grade 9 students under different reviewing test conditions has different test information, which is statistically significant at .05 level. The computerized adaptive scholastic aptitude test program that allows reviewing of answers made the mean ability of the examinees higher than test program that does not allow reviewing of answers. In calculating the information test, the mean ability of the examinees were calculated. The test program that allows reviewing of answers and the test program that does not allow reviewing of answers had different information. This is connected to the study of Tienoraset (2006) that compared the average examinees' mean ability under the condition of not allowing reviewing of answers and allowing reviewing of answers. The examinees with medium and low ability and those with average ability under the condition of not allowing reviewing of answers and allowing reviewing of answers had statistical and significant difference at .05 level. However, there was no difference between the examinees with high ability and those with average ability in the condition of not allowing reviewing of answers and allowing reviewing of answers. From the study of Vispoel et al. (2000) on computerized

adaptive testing which allows reviewing of answers, it is found that the mean ability of the examinees slightly increased after reviewing of answers. Therefore, in this research, the examinees from allowing reviewing of answers test program had higher mean than the program that does not allow reviewing of answers; so, allowing of reviewing of answers test program has higher information test mean than the test program that does not allow reviewing of answers.

The result of the two evaluation programs is statistically significant at .05 level. The test program that allows reviewing of answers is greater than the test program that does not allow reviewing of answers. The examinees may likely want to review answers in order to be confident in the test; so, the opinions about the test program that allows reviewing of answers and the one that does not allow reviewing of answers differ. The research sample tested with the program that allows reviewing of answers has more satisfaction than the ones tested with the test program that does not allow reviewing of answers. However, for test program, it is necessary to use interesting and fashionable program as well as computer, which is not related to reviewing or not reviewing of answers. Therefore, in the research result, there is no difference between these two programs with the use of computer. Vispoel (1998) found that mostly of the examinees need to review answers and insisted that the examinees had positive opinion and satisfaction towards the test program that allow reviewing of answers more than the one that does not allow reviewing of answers.

### Conflict of Interests

The authors have not declared any conflicts of interest.

### ACKNOWLEDGEMENT

The research project was supported by funds from income budget study in the year 2015, from the Faculty of Education, Mahasarakham University and funds received from National Research Council of Thailand in the year 2015.

### REFERENCES

- Green BF, others. (1984). "Technical Guidelines for Assessing Computerized Adaptive Tests," *J. Educ. Meas.* 21:347-360.
- Hambleton RK, Cook LL (1977). "Latent Trait Model and their Use in the Analysis of Educational Test Data," *J. Educ. Measure.* 14(2):75-96.
- Junpeng P (2007). "A Comparison of Quality of Multidimensional Item Response Theory Linking Methods under the Differences of Rotation, Dimensional Structure, and Correlation Coefficient." Dissertation for the Degree of Doctor of Education Program. Bangkok : Chulalongkorn University.
- Loard T, Nicely G (1997). "Does spatial aptitude influence science-math subject preferences of children?," *J. Elementary Sci. Educ.* 9:67-81.
- Ministry of Education. National Education Act 2010 (No.3). Bangkok : Express Transportation Organization of Thailand Printing (E.T.O.),010.
- Morton SL (2004). "The Relationship between Language Learning Aptitude and the Perception and Production and Production of Foreign Speech Sounds," *Masters Abstracts International* 42(2):407.
- Olea J, others. (2000). "Psychometric and psychological effects of review on computerized fixed and adaptive tests," *Educ. Psychol. Measure.* 21:157-173.
- Pha-on J (2010). "Factors Affecting Happy Learning of Matthayomsuksa 3 Students under Saraburi Educational Service Area Office." Dissertation for the Degree of Master of Education Program. Lopburi: Thepsatri Rajabhat University.
- Philatun W, On-uma W (2008). "Factors Affecting Mathematics Achievement of Second Year of Vocational Certificate Students at Nakhonsawan Vocational College." Independent Study for the Degree of Master of Science in Statistics Apply Program. Phitsanulok: Naresuan University.
- Pimsiri T (2006). "A Comparison of Examinees' Abilities, Answer-Changing Characteristics and Time Spent in Computerized Adaptive Testing under Varying Testing Conditions and Examinees' Ability Levels." Dissertation for the Degree of Master of Education Program. Bangkok: Chulalongkorn University.
- Revuelta J, Ximenez MC, Olea J (2003). "Psychometric and psychological effects of item selection and review on Computerized Adaptive Testing," *Educ. Psychol. Measure.* 63 (5): 791-808.
- Vispoel WP, Hendrickson AB, Bleiler T (2000). "Limiting answer review and change on computerized adaptive vocabulary tests : Psychometric and attitudinal results," *J. Educ. Measure.* 37:21-38.
- Vispoel WP (1998). "Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests," *J. Educ. Measure.* 35:328-345.
- Wainer H, Dorans NJ, Flaugher R, Green BF, Mislevy RJ (1993). "Test Validity Measurement," Research Report. Psychometric Methods Program, Department of Psychology. University of Minnesota.