

A Study of the Quality of Scholastic Aptitude Test by Applying Modern Test Theories

Ruangdech Sirikit^a, Panwasn Mahalawalert^b

^{a,b} Educational and Psychological Test Bureau, Srinakharinwirot University, Bangkok, Thailand

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: The objective of this research was to study the quality of the scholastic aptitude test by applying modern test theories through Item Response Theory, test parameters, reliability, and differential item functioning. The secondary data were collected from the SWUSAT Test Development Project 2014, which consists of a test for all 4 subjects: Verbal Factor, Number Factor, Reasoning Factor and Spatial Factor. The data analysis was divided into 3 steps as follows: Step 1: Data preparation according to the studied factors, the researcher performed the analysis of the basic statistics of the data by analyzing the preliminary data with descriptive statistics, i.e. frequency, percentage. Step 2: Analyze test parameters using a 2-parameter test response model based on parameters, discriminant power (a), difficulty (b) parameter, and reliability. Step 3: Analysis of data with differential item functioning.

The results of the research were as follows:

The results of the analysis of the individual parameters of the 4-subject aptitude test with the 2-parameter test response analysis model were determined from the discriminant power parameter (a) and the difficulty parameter. (b) When considering the results of the course analysis, it was found that all subjects had a predominantly number of tests with a parameter that met the quality criteria, including a relatively high reliability level.

The results of analysis of the differential item functioning and the SWUSAT test, which measures Verbal Factor, Number Factor, Reasoning Factor and Spatial Factor. The difference of the test was in the range of 6.67% - 43.33, with two tests having a number of questions that performed differently, with males being more likely to get the correct answers than females, while the other two had the same number of questions. And all tests performed at a low level of different functions

Keywords: Aptitude Test, Modern Test Theories

1. Introduction

The multiple-choice test is a widely used test tool, being used for a wide variety of important test situations, such as the placement test in schools or the entrance examination for students to study at the university level, etc. The reason why the multiple-choice test is so popular to use is due to its distinctive features: It can be used for a large number of test takers within a limited time. And the latent trait of the candidate can be examined to meet the needs of the test taker as the multi-choice test is used in a wide range of situations. Therefore, the quality control of the test is very important.

Psychologists and measurers strive to formulate theories so that measurement results truly represent individual attributes or abilities. Modern Test Theory is classified as a reference theory in testing; Generalizability Theory: G and Item Response Theory: IRT (Sirichai Kanjanawat, 2007).

A good measuring tool requires good construction design and quality checks to give reliability or the assurance that the tools they make are able to provide accurate information. For the key features of the Item Response Theory (IRT), a theory that describes the relationship between a person's inherent traits or competence and each test's response behavior as having a probability of correct response. It is a theory that solves many limitations of the Classical Test Theory. Test quality according to the test response theory consists of test parameters: difficulty (b), discriminant power (a), and guessing (c). They are values that do not vary depending on the group of test takers. In addition, Differential Item Functioning; DIF) is another type of quality examination of the validity of an test (Camilli & Shepard, 1994). It can be seen that the study of the audit of the different functions of the present grader test has a lot of interest. This is partly because in today's society there are quite a lot of competition using standardized examinations as a tool in judging. Therefore, it is important to determine whether the test in various situations is fair to stakeholders.

Therefore, the researcher is interested in studying the quality of the scholastic aptitude test by applying modern test theories which will provide useful information for future development of the scholastic aptitude test. SWUSAT is a standardized test development project that the Office of Educational and Psychological Testing is responsible to create for use in the entrance examination for higher education students at Srinakharinwirot University.

2. Aim

To apply modern test theories to study the quality of the scholastic aptitude test using the Item Response Theory: item parameters, reliability, differential item functioning

3. Definitions

Differential Item functioning: DIF refers to the nature of a test that provides different groups of test takers of the same level of examiner but with different opportunities to answer the test correctly. In this research, an analysis of different functions examined the students' characteristics, namely gender variables.

The SWUSAT means the scholastic aptitude test that the Office of Educational and Psychological Testing was created for use in the entrance examination for students to study at the tertiary level, Srinakharinwirot University.

4. Methods

Research paper on a quality study of the scholastic aptitude test by applying modern test theories has the objectives of research which are: To apply modern test theories to study the quality of the study the scholastic aptitude test by using the Item Response Theory, test parameters, reliability, and performing different functions of examination. The method of research is divided into 2 parts: part 1, data used in research, part 2, data analysis. With details as follows:

Part 1 Research Data

In this research, the researcher used the secondary data obtained from the SWUSAT scholastic aptitude test Development Project, a project organized to develop the scholastic aptitude test for use in the examination for the selection of students to study at the tertiary education level of Srinakharinwirot University, fiscal year 2014. The first-year students were tested at Srinakharinwirot University for the academic year 2015 as follows:

1) The sample

Data collected for the development of the SWUSAT test year 2014 are detailed in Table 1.

Table 1: Number of samples for data collection for SWUSAT test development

Year	Subjects	Number of Students		
		Male	Female	Total
2014	Verbal	977	1,812	2,789
	Number	978	1,815	2,793
	Reasoning	925	1782	2,707
	Spatial	940	1851	2,791

2) The tools for data collection include:

2.1) SWUSAT Project Test for Academic Year 2015, totaling 30 items each, with details as follows

The characteristics of the SWUSAT test are as follows:

1. Verbal Factor is a measure of ability to understand words, texts, words, words, conversations, stories in language, communicating using language, as well as choosing the right language.
2. Number Factor is a measure of numerical performance, understanding of relationships and meaning of numbers, and skills in using the addition, subtraction, multiplication and division skills accurately and quickly.
3. Reasoning Factor is a measure of ability to think, analyze, reason, importance and relationships in different ways, and have the ability to infer from events or stories in a reasonable manner.
4. Spatial Factor is a measure of visual performance and understanding of the relationship of dimensions between width, length, height, depth, point, line, complexity, concealment of the geometric figure. As well as the volume and the size of the distance can be accurately imagined to combine or separate things well.

Part 2 Data Analysis

Quality Study of the scholastic aptitude test by using Item response theory, the test parameters, the reliability and the differential item functioning are analyzed in 3 steps as follows:

Step 1: Data Acquisition according to the Study Factors, the researcher performed the fundamental statistical analysis of the data by analyzing the preliminary data with descriptive statistics, i.e. frequency, percentage.

Step 2: Analyze test parameters using a 2-parameter test response model based on discriminant power parameters (a), preferably between +0.50 and +2.50. And the difficulty parameter (b) should be between -2.50 and +2.50 Sirichai Kanjanawat (2012).

Step 3 Data analysis by performing Differential Item Functioning, and the differential item functioning

An examination of the Differential Item functioning (DIF) was performed by using DIFAS 5.0 (Penfield, 2012), a program that used mantel-Hanschel analysis with odds ratio. The analysis was as follows:

$$\hat{\alpha}_{MH_i} = \frac{\sum_{s=1}^s R_{1s} F_{0s} / n_s}{\sum_{s=1}^s R_{0s} F_{1s} / n_s} \tag{1}$$

and Mantel-Hanschel, together with a copy of the odds ratio, the formula for the analysis was shown as

$$\hat{\lambda}_j = \ln(\hat{\alpha}_j) \tag{2}$$

This is because the Mantel-Hanschel analysis, together with the odds ratio copy, is a comparison of the response ratio of the reference group (reference) to the comparison group (Focal) with different criteria for determining the function. The results of the test were obtained from the LOR Z and MH LOR values with the following interpretive steps (Penfield, 2012).

1.) Consider the LOR Z

1.1 Value LOR Z > 2 or Value LOR Z < -2 They perform significantly different functions between the groups. Consider the MH LOR in step 2

1.2) -2 ≤ Value LOR Z ≤ 2 indicates that the question does not act differently between the groups.

2) Consider the MH LOR for questions that perform different functions between groups H:

2.1) Positive MH LOR value (+) indicates that the question is favorable to reference group

2.2) MH LOR value has a negative value (-) indicates that the question is in favor of the Focal group.

with the criteria for determining the different functions of the test from Tau²

3) Consider the Tau² value for the test with different functions between the groups:

3.1 Tau² < 0.07 indicates that the test has different functions of the test at a low level

3.2 0.07 < Tau² < 0.14 indicates that the test The differential item functioning are at an intermediate level.

3.3 Tau² > 0.14 indicates that the examination results in differential item functioning at a high level.

5. Results

Part 1 Fundamental analysis results

Presentation of fundamental analysis results is a presentation of the basic data of a sample to characterize the distribution of the data. It contains descriptive statistics such as frequency, percentage.

1. Basic information of the sample

Analysis of fundamental data of the sample group, the researcher led the variables, basic characteristics, students of Srinakharinwirot University. Available in the database of the SWUSAT test project for academic year 2014 to be analyzed to determine the characteristics of the sample.

Table 2 shows the number of students who have taken the SWUSAT test by gender and course

Subjects	Sex		Total
	Male	Female	
Verbal	977 (35.03%)	1,812 (64.97%)	2,789 (100%)

Number	978 (35.02%)	1,815 (64.98%)	2,793 (100%)
Reasoning	925 (34.34%)	1,782 (65.83%)	2,707 (100%)
Spatial	940 (33.68%)	1,851 (66.32%)	2,791 (100%)

Table 2 shows the number of data samples of students in the first year of the academic year 2015 at Srinakharinwirot University by gender and course. The number of females taking the test is greater than that of males and each subject has the same number of students taking the exam.

Part 2: Analyze test parameters using a 2-parameter test response model based on discriminant power parameters (a), the difficulty parameter (b), and reliability

The test parameters were analyzed by a 2-parameter test response model based on the discriminant power parameter (a) and the difficulty parameter (b). The analysis results are detailed in Table 3.

Table 3: Discriminant power parameters (a) and the difficulty parameter (b) were analyzed by a 2-parameter question response model.

Item	Verbal (V)		Number (N)		Reasoning (R)		Spatial (S)	
	a	b	a	b	a	b	a	b
1	0.13	1.76	1.39	-0.45	0.71	0.37	0.57	-1.39
2	0.39	0.32	1.26	0.47	1.04	0.54	0.38	2.26
3	0.22	3.75	1.12	0.43	1.05	0.74	0.51	-1.6
4	0.08	8.13	0.51	0.54	0.9	0.41	0.43	-1.12
5	0.43	-2.78	0.9	2.02	1	-0.16	0.13	7.14
6	0.05	8.97	0.69	0.2	0.76	-0.07	0.42	-3.58
7	0.19	-0.65	1.39	-0.19	0.87	0.71	0.49	0.22
8	0.2	-1.33	1.91	0.24	0.81	1.31	0.57	-0.63
9	0.03	43.1	0.72	0.34	0.83	0.9	0.69	-1.83
10	0.65	-2.63	1.46	0.35	0.86	1.12	0.56	-0.49
11	0.53	-2.38	0.63	1.51	0.76	0.74	0.47	-1.39
12	0.52	1.64	1.44	0.49	0.8	0.86	0.54	1.37
13	0.28	0.27	0.36	3.05	0.8	0.04	0.54	-0.24
14	0.1	8.4	1.08	0.48	1.01	0.31	0.6	-1.44
15	0.67	-0.03	1.97	-0.57	0.85	1.96	0.66	0.26
16	0.42	-2.21	1.21	0.01	0.64	-0.6	1.02	-1.14
17	0.53	0.95	0.7	0.58	0.69	-0.14	1.04	0.17
18	0.41	1.36	1.32	0.04	0.53	0.71	1.41	-0.76
19	0.5	1.19	0.76	2.51	0.67	1.45	1.14	-0.22
20	0.91	-0.07	0.68	0.1	0.51	0.47	0.59	0.97
21	0.04	1.13	0.6	2.4	0.48	-1.02	1.24	0.78
22	0.23	-0.71	0.7	0.48	0.71	0.15	0.88	0.64
23	0.33	-0.19	1.03	-0.32	0.34	-1.23	1.82	-0.55
24	1.11	-0.85	0.66	1.69	0.53	1.36	1.81	-0.41
25	0.71	-1.2	1.27	-0.21	0.46	-0.74	2.11	-0.19
26	0.39	1.25	1.22	-0.1	0.4	1.91	1.77	-0.57
27	0.52	-0.26	0.44	3.3	0.53	-1.23	0.83	0.47
28	0.39	0.14	0.8	0.31	0.49	-1.22	1.39	-0.07
29	0.63	0.41	0.32	3.27	0.19	5.14	0.46	2.3
30	0.68	0.42	0.56	2.49	0.28	-0.94	1.07	-0.27

From Table 3, it is the result of the individual parameter analysis of the four scholastic aptitude test subjects with the 2-parameter test response analysis model based on the discriminant power parameter (a) and the difficulty parameter (b). When considering the results of the analysis as a course, it was found that

Verbal Factor (V) had 12 questions with discriminant power parameter (a) from the total of 30 questions, representing 40% of the total exam. For the difficulty parameter (b), there were 23 questions on the test parameters ranging from -2.50 to 2.50, accounting for 76.67% of the total number of the exams.

Number Factor (N) had the test with discriminant parameters (a) between 0.50 and 2.50, 27 questions out of a total of 30 questions, accounting for 90 percent of the total exam. For the difficulty parameter (b), there were 26 questions in the test that ranged from -2.50 to 2.50, representing 86.67 percent of the total.

Reasoning Factor (R): There are 23 tests with the discriminant power parameter (a) between 0.50 and 2.50 out of 30 questions, accounting for 76.67% of the total. For the difficulty parameter (b), 29 questions on the test parameters ranged from -2.50 to 2.50, representing 96.67% of the total exam.

Spatial Factor (S): There are 23 tests with the discriminant power parameter (a) between 0.50 and 2.50 out of 30 questions, accounting for 76.67% of the total. For the difficulty parameter (b), there were 28 questions in the test that ranged from -2.50 to 2.50, accounting for 93.33% of the total exam.

Table 4 Reliability of the scholastic aptitude test. The results were analyzed by the 2-parameter test response model

Subjects	Reliability
1. Verbal (V)	0.59
2. Number (N)	0.84
3. Reasoning (R)	0.75
4. Spatial (S)	0.82

From Table 4, the results of the reliability analysis of the scholastic aptitude test were classified by subject. It was found that Number Factor (N) had the highest reliability at 0.84, followed by the Spatial Factor (S) had 0.82 reliability. The subject with the lowest reliability, Verbal Factor (V), had a reliability factor of 0.59.

Part 3 The results of validating differential item functioning and differential test functioning

According to the data analysis in this section, the researchers used the scores from the SWUSAT test for the academic year 2014 of Srinakharinwirot University students to analyze the differential item functioning in gender (male, female) using the Mantel-Hanssel method together with Log Odd Ratio method with the DIFAS program (Penfield, 2012) with the male reference group and the female focal group as follows:

1. Consider the LOR Z:

1.1) $LOR Z > 2$ or $LOR Z < -2$, indicating that the question has a significantly different function between the groups. The MH LOR value should be considered in step 2)

1.2) $-2 \leq LOR Z \leq 2$, indicating that the question does not function differently between the groups.

2) Consider the value of MH LOR for questions that perform different functions between groups

2.1) The MH LOR value is positive (+), indicating that the question will be in favor of the Reference group.

2.2) The MH LOR value is negative (-), indicating that the question is in favor of the Focal group.

3) Tau2 is considered for the examination with different functions between the groups.

3.1 $Tau^2 < .07$ indicates that a low level of function of the test occurred.

3.2 $.07 < Tau^2 < .14$ indicates that the function was performed. Intermediate level differences of the test.

3.3 $Tau^2 > .14$ indicate that different functions of the test were performed at high levels.

The results of the analysis of the different functions of the questions were:

Table 5, the results of the analysis of the differential item functioning and the Verbal Aptitude test (V).

Item	MH LOR	LOR SE	LOR Z
1	-0.0194	0.0816	-0.2377
2	0.1119	0.0828	1.3514
3	-0.0517	0.0883	-0.5855
4	-0.008	0.0847	-0.0945

5	-0.0723	0.0962	-0.7516
6	0.0815	0.0822	0.9915
7	-0.0008	0.0813	-0.0098
8	0.0206	0.0824	0.250
9	0.1155	0.094	1.2287
10	-0.0702	0.1089	-0.6446
11	0.0086	0.0974	0.0883
12	-0.0684	0.0902	-0.7583
13	0.0949	0.0816	1.163
14	-0.029	0.0884	-0.3281
15	-0.2202	0.0855	-2.5754**
16	0.0862	0.0909	0.9483
17	-0.154	0.0858	-1.7949
18	0.2176	0.0845	2.5751**
19	-0.088	0.0865	-1.0173
20	-0.0084	0.0863	-0.0973
21	0.1498	0.0806	1.8586
22	-0.1059	0.0817	-1.2962
23	0.0904	0.0821	1.1011
24	-0.0608	0.0936	-0.6496
25	0.0916	0.0919	0.9967
26	-0.046	0.0851	-0.5405
27	0.0348	0.084	0.4143
28	-0.1079	0.0825	-1.3079
29	-0.0485	0.0846	-0.5733
30	-0.0006	0.0857	-0.007
Statistic	Value	SE	Z
Tau²	.001	.002	.500

From Table 5, the results of the analysis of the different functions of the DIFs (V) for the Verbal Aptitude test (V) were found that there were 2 out of 30 questions in total that caused differential item functioning s. Females were more likely to get correct answers than males, with 1 question being number 18, while males were more likely to get correct answers than females are number 15.

The results of the analysis of the different functions of the DTF) were found that Tau2 = .001, which was less than .07, showed that different functions of the test were performed at a low level.

Table 6 Results of analysis of differential item functioning and Number Aptitude test (N)

Item	MH LOR	LOR SE	LOR Z
1	0.2142	0.0966	2.2174**
2	-0.113	0.0974	-1.1602
3	0.1971	0.0937	2.1035**
4	0.1797	0.0853	2.1067**
5	0.1726	0.1146	1.5061
6	0.0282	0.0869	0.3245
7	0.0888	0.0949	0.9357
8	0.3672	0.1021	3.5965**
9	0.0525	0.0874	0.6007
10	-0.0279	0.0988	-0.2824
11	-0.0593	0.0938	-0.6322
12	0.3563	0.0994	3.5845**
13	0.1031	0.0933	1.105
14	0.0705	0.0935	0.754
15	-0.1434	0.1035	-1.3855
16	0.1239	0.0925	1.3395
17	-0.0937	0.0886	-1.0576
18	0.0943	0.0938	1.0053
19	-0.0321	0.1212	-0.2649
20	0.0353	0.0862	0.4095
21	-0.0827	0.1063	-0.778

22	-0.1847	0.089	-2.0753**
23	-0.1423	0.0908	-1.5672
24	0.0308	0.0976	0.3156
25	-0.314	0.0944	-3.3263**
26	-0.3045	0.0933	-3.2637**
27	-0.1362	0.1052	-1.2947
28	-0.1983	0.0886	-2.2381**
29	0.01	0.0926	0.108
30	-0.2395	0.1055	-2.2701**
Statistic	Value	SE	Z
Tau²	.020	.008	2.500

From Table 6, the results of the DIF's Analysis of the Number Aptitude Test(N) were found that there were 10 of the 30 different functions of the female exam. There are 5 questions more likely to be correct than males: 22, 25, 26, 28 and 30, while males are 5 more likely to get correct answers than females, namely 1, 3, 4, 8 and 12.

The results of the analysis of the different functions of the number aptitude test (DTF) (N1) showed that Tau2 = .02, less than .07, meaning that the different functions of the test were performed at low levels.

Table 7: Results of Functional Analysis for Different Tasks and Reasoning Aptitude Test (R)

Item	MH LOR	LOR SE	LOR Z
1	0.0498	0.0878	0.5672
2	0.1647	0.0929	1.7729
3	0.0001	0.0952	0.0011
4	-0.1169	0.0907	-1.2889
5	-0.091	0.0916	-0.9934
6	-0.1295	0.0883	-1.4666
7	-0.1323	0.0921	-1.4365
8	-0.0083	0.0972	-0.0854
9	-0.1903	0.0948	-2.0074**
10	0.0213	0.0959	0.2221
11	-0.2929	0.0921	-3.1802**
12	-0.1239	0.0931	-1.3308
13	-0.2779	0.0892	-3.1155**
14	-0.1557	0.0921	-1.6906
15	-0.2392	0.1134	-2.1093**
16	0.2422	0.09	2.6911**
17	0.1921	0.0883	2.1755**
18	-0.0288	0.088	-0.3273
19	-0.0325	0.0959	-0.3389
20	0.4601	0.0862	5.3376**
21	-0.085	0.0882	-0.9637
22	0.3977	0.0893	4.4535**
23	-0.0388	0.0867	-0.4475
24	0.1086	0.0903	1.2027
25	-0.1299	0.0877	-1.4812
26	-0.0448	0.0908	-0.4934
27	0.2011	0.0925	2.1741**
28	0.0263	0.0906	0.2903
29	-0.0291	0.0925	-0.3146
30	0.12	0.0857	1.4002
Statistic	Value	SE	Z
Tau²	0.023	0.008	2.875

From Table 7, the results of the DTF's Analysis of the Reasoning Aptitude Test (R) were found that there were 9 of the 30 different functions of the female test. There are four more chances of getting right than males: 9, 11, 13, and 15, while males are more likely to get correct answers than females. There are 5 questions, namely, clauses 16, 17, 20, 22, and 27.

For the analysis of the different functions of the DTF), the Reasoning Aptitude Test (R), it was found that Tau2 = .023, which is less than .07, showed that different functions were performed in the Reasoning Aptitude Test. Low level

Table 8: The results of analyzing the differential item functioning and the Spatial Factor (S)

Item	MH LOR	LOR SE	LOR Z
1	0.304	0.096	3.180**
2	-0.136	0.092	-1.481
3	0.125	0.095	1.324
4	-0.176	0.089	-1.970
5	0.072	0.091	0.791
6	-0.366	0.108	-3.385**
7	-0.108	0.086	-1.245
8	-0.131	0.090	-1.462
9	-0.283	0.104	-2.712**
10	-0.179	0.089	-2.019**
11	-0.283	0.091	-3.094**
12	-0.058	0.092	-0.634
13	-0.393	0.088	-4.457**
14	-0.392	0.095	-4.117**
15	-0.172	0.091	-1.896
16	-0.086	0.102	-0.843
17	0.360	0.092	3.910**
18	0.160	0.104	1.536
19	-0.007	0.095	-0.069
20	0.328	0.088	3.722**
21	0.451	0.099	4.543**
22	0.288	0.091	3.177**
23	0.001	0.104	0.013
24	0.219	0.101	2.171**
25	0.153	0.101	1.516
26	-0.140	0.102	-1.370
27	0.196	0.089	2.202**
28	-0.033	0.094	-0.353
29	0.185	0.095	1.940
30	0.119	0.091	1.312
Statistic	Value	SE	Z
Tau²	0.044	0.014	3.143

From Table 8, the Analysis of differential item functioning and the Spatial Aptitude test (S): From Table 9, the Analysis of differential item functioning (DIF), the Spatial Aptitude test (S) There were a total of 13 out of 30 exams that performed significantly different functions. Females are more likely to get correct answers than males. There are 6 questions: No. 6, 9, 10, 11, 13 and 14, while males are more likely to get correct answers than females. 17, 20, 21, 22, 24, and 27

Table 9: Summary of the Analytic results of Differential Item Functioning and Differential Test Functioning, classified by Subject

year	Subject	Number of Item	Number of DIF				Level of DTF
			M	F	Total	%	
2014	V	30	1	1	2	6.67	Low
	N	30	5	5	10	33.33	Low
	R	30	5	4	9	30.00	Low
	S	30	7	6	13	43.33	Low

From Table 9, Results of Analysis of Differential Item Functioning and Differential Test Functioning of the SWUSAT in four factors in the academic year 2014, there were a total of 4 tests, each of which performed differential item functioning in the range of 6.67 - 43.33 percent. It performed differently, with males more likely to get correct answers than females, while the other two have the same number of questions. And all tests performed at low levels of different functions.

6. Discussion

A quality study of the scholastic aptitude test by applying modern test theories, the objective of research is to apply modern test theories to study the quality of the scholastic aptitude test by using Item response theory, test parameters, reliability, and differential item functioning can be discussed in the following ways.

The results of the analysis of the individual parameters of the 4-subject aptitude test with a 2-parameter test response analysis model determined from the discriminant power parameter (a) and the difficulty parameter (b). It was found that all subjects had a number of tests with parameters that met the quality criterion, including a relatively high Reliability, which was consistent with the quality criterion described by Sirichai Kanjanawasee (2012) that the discriminant power parameter (a) should be between +0.50 and +2.50 while the difficulty (b) parameter should be between -2.50 and +2.50.

Analysis of the different functions of the test and the SWUSAT test, which measures verbal, number, reason, and spatial relationship. The difference of the test ranges from 6.67% - 43.33%. There were two tests with different number of questions that performed different functions, with males more likely to get correct answers than females, while the other two had the same number of questions. And all tests performed at low levels of different functions. The above findings were consistent with the research of Chutima Sangdararat (2002), comparing the results of examining the different functions of the scholastic aptitude test based on familiarity, interest and satisfaction in the test with different examination methods. The results of the important study found that the number of tests that perform different functions of the Reasoning Aptitude test with different methods can indicate the verses that perform different functions from the gender reference group and feeling of interest. The number of differences was statistically significant at the .05 level, while the reference group showed no statistically significant difference in familiarity and satisfaction.

Giray (1995) also analyzed the differential item functioning inations for the University Admission Examination, which are vocabulary and geometry exams. The test takers were divided according to the variables studied as follows: Gender and economic status; According to the research results, it was found that the test that performed different functions was in mathematics, both in the case of grouping of test takers according to gender and economic status variables. For gender variables, the results of the analysis revealed that men had advantages in taking the computational exam, while women had advantages in taking the vocabulary and geometry exam. This suggests that women have better language or vocabulary proficiency, and that males have better computational skills than females.

Therefore, it can be seen that both the verbal proficiency test, number, reason and spatial test have differential item functioning classified by gender.

Suggestions

From the research on the differential item functioning and the SWUSAT scholastic aptitude test for 2014, there are many useful research results. The situation by the researcher has presented recommendations for the application of the research results. The suggestions for the next research are as follows:

1. Suggestions for applying the research results are divided into 2 parts as follows:

1.1 From the research results, different analyzes of the test and the examination can be applied in the construction and development of both the test and the test in order to achieve a standardized, thereby increasing the reliability of the tools used in measurement results within the school as well.

1.2 The test library is an alternative development of measurement and evaluation work in which many basic education institutions are developing their own test libraries. Usually, the test is selected to be kept in the test library, with difficulty and discriminant power as an analysis principle. The differential item functioning are another way to help build trust in the school library.

Suggestions for the next research

The data used for analysis in this research is the secondary data obtained from the SWUSAT Project of Academic Year 2014. In this study, the researcher performed a specific study for differential item functioning (DIF) and the differential item functioning (DTF) by studying only the correct options. Suggestions for further research; The researcher should conduct further analytical studies on the differential distractor functioning (DDF) to help add more information to the construction and development of a standardized test

References