

AN INVESTIGATION OF DIFFERENTIAL ITEM FUNCTIONING AND DIFFERENTIAL TEST FUNCTIONING OF SWUSAT DURING 2010-2013

Ruangdech Sirikit¹

Manida choptham²

Panwasn Mahalawalert³

Navarin Tagontong⁴

Supa Apinyapibal⁵

Abstract: The purpose of this research was to analyze the differential item functioning (DIF) and the differential test functioning (DTF) of scholastic aptitude test (SWUSAT) for groups of students and classified by gender. The data used in this research is the secondary data from Srinakharinwirot University Scholastic Aptitude Test (SWUSAT) year 2011 and 2013. SWUSAT test consists of 4 subjects and 10 tests. There are verbal ability test, number ability test, reasoning ability test and spatial ability test. The data analysis was analyzed in 2 steps. The first step was analyzing descriptive statistics. In the second step were analyzed differential item functioning (DIF) and differential test functioning (DTF) by using the DIFAS program.

The results of DIF and DTF analysis for all 10 tests in year 2011. Gender was the characteristic that found DIF all 10 tests. The percentage of item number that found DIF between 10% - 46.67%. There are 4 tests that most of items favors female group. There are 3 tests that most of items favors male group and there are 3 tests that the number of items favors female group equal favors male group. For Differential test functioning (DTF), there are 8 tests that have small level.

The results of DIF and DTF analysis for all 10 tests in year 2013. Gender was the characteristic that found DIF all 10 tests. The percentage of item number that found DIF between 6.67% - 60%. There are 5 tests that most of items favors female group and 2 tests that most of items favors male group. There are 3 tests that the number of

¹ Ph.D., Lecturer, the EPTB, Srinakharinwirot University, Thailand.
ruangdech_s@hotmail.com

² Ph.D., Lecturer, the EPTB, Srinakharinwirot University, Thailand.
manada@g.swu.ac.th

³ Lecturer, the EPTB, Srinakharinwirot University, Thailand.
panwasn@g.swu.ac.th

⁴ Lecturer, the EPTB, Srinakharinwirot University, Thailand.
navarin@g.swu.ac.th

⁵ Lecturer, the EPTB, Srinakharinwirot University, Thailand.
supa@g.swu.ac.th

items favors female group equal favors male group. For Differential test functioning (DTF), there are 8 tests that have small level.

Introduction

The fairness is a very important characteristic in the item. The items without fairness mean items which enable any group of students in the same level of competency to give correct answer differently which may result from the difference in terms of language, gender, religion, race, culture, domicile or level of intelligence. This nature of items is called the Differential Item Function (DIF). The Differential Item Functioning (DIF) is the matter on which the test developers, educational assessors and statisticians place importance and put their effort to develop guidelines or statistic techniques to investigate items with differential functioning among different groups of students in order to enable the fairness in the test results to the test results as much as possible. The test will include items with questions in the content which require the students to express what they know within the measurement framework. Therefore, the resource of difficulty depends on the construction of items which may contain a part impertinent to the group of students, such as items with vocabularies which each group of students may interpret differently. The scores obtained from the tests therefore result from the limitations of word meaning or language used in the tests (Camili & Shepard, 1994). The purpose of action relating to item bias is to investigate and indicate the test validity

SWUSAT is the standardize test that construct and develop by the Educational and Psychological Test Bureau (EPTB) for the entrance examination to the higher education in Srinakharinwirot University. It is thus important for the construction of SWUSAT to take into account the test fairness because SWUSAT is used for testing students throughout the country. Therefore, researcher wish to investigate the differential functioning of SWUSAT item and test during the academic years 2011 and 2013.

Objective

To analyze the differential item functioning (DIF) and the differential test functioning (DTF) of scholastic aptitude test (SWUSAT) for groups of students and classified by gender.

Scope of the Research

This research focuses on the differential item functioning (DIF) and differential test functioning (DTF) covering especially in the academic years 2011 and 2013 due to the limitations of date for the research.

Definitions in the Research

Differential Item Functioning (DIF) means the item characteristics which enable the students from different groups with the same level of competency to give correct answer differently. This research used the analysis of differential item functioning investigated from the student characteristics, that was; the gender factor.

SWUSAT means the scholastic aptitude test established by the Educational and Psychological Test Bureau (EPTB) for the entrance examination to the higher education in Srinakharinwirot University.

Methods

The research entitled “An Investigation of Differential Item Functioning and Differential Test Functioning of SWUSAT during 2010-2013” aims to analyze the differential item functioning (DIF) and the differential test functioning (DTF) of SWUSAT for groups of students and classified by gender. for groups of students and classified by gender. The methods of research were divided to 2 sections: 1) data for research and 2) analysis of data with following details:

Section 1: Data for Research

The secondary data from SWUSAT Test Development Project which was established by the Educational and Psychological Test Bureau (EPTB) for the entrance examination to the higher education in Srinakharinwirot University. It was used in this research by analyzing the score of students in the freshmen year of Srinakharinwirot University. SWUSAT test consists of 4 subjects. The numbers of tests were 10 tests in each year (2011 and 2013). According to the details below:

1. Study Samples

The samples used to collect data for SWUSAT on years 2011 and 2013 were shown in the Tables below:

Table 1: Number of Samples to Collect Data for SWUSAT Test

| Academic Year | Test | Number of Students |
|---------------|------|--------------------|
| 2011 | V1 | 801 |
| | V2 | 699 |
| | N1 | 801 |
| | N2 | 699 |
| | R1 | 801 |
| | R2 | 698 |
| | S1 | 451 |
| | S2 | 350 |
| | S3 | 379 |
| | S4 | 379 |
| 2013 | V3 | 2,150 |
| | V4 | 961 |
| | V5 | 958 |
| | N3 | 2,150 |
| | N4 | 2,305 |
| | R3 | 2,150 |
| | R4 | 2,214 |
| | S5 | 1,100 |
| | S6 | 994 |
| | S7 | 2,305 |

2. Tools for Collecting Data

Test of SWUSAT – 10 tests in the academic year 2011 and 10 tests in the academic year 2013 according to the details below:

Table 2: Number of Items in SWUSAT by Factor

| Academic Year | Test | Number of Items |
|---------------|------|-----------------|
| 2011 | V1 | 30 |
| | V2 | 30 |
| | N1 | 30 |
| | N2 | 30 |
| | R1 | 30 |
| | R2 | 30 |
| | S1 | 30 |
| | S2 | 30 |
| | S3 | 30 |
| | S4 | 30 |
| 2013 | V3 | 30 |
| | V4 | 30 |
| | V5 | 30 |
| | N3 | 30 |
| | N4 | 30 |
| | R3 | 30 |
| | R4 | 30 |
| | S5 | 30 |
| | S6 | 30 |
| | S7 | 30 |

Characteristics of SWUSAT test are as follows:

1. Verbal test (V) is the competency assessment on understanding vocabulary, statement, poetry, conversation, circumstances relating to language, communication by means of language, and use of appropriate language.

2. Number test (N) is the competency assessment on calculating numbers; understanding relationship and meaning of numbers; and skill to use mathematical symbols of addition, subtraction, multiplication and division correctly, accurately and rapidly.

3. Reasoning test (R) is the competency assessment on thinking, analyzing and reasoning the importance and relationship in different forms; ability to categorize and classify analogically correctly; and ability to summarize circumstances and stories accurately and reasonably.

4. Spatial test (S) is the competency assessment on visualizing and understanding spatial relationships i.e. width, length, height, point, line; geometrical complexity and volume and magnitude; accurate estimation; and ability to imagine integration and disintegration of things properly.

Section 2: Analysis of Data

This research entitled “An Investigation of Differential Item Functioning and Differential Test Functioning of SWUSAT during 2010-2013” had 2 steps of analysis as follows:

Step 1 Organizing data according to the study factor

Step 2 Analyzing the differential item functioning and the differential test functioning

Step 1 Organizing Data According to the Study Factor

The data analysis of descriptive statistics was conducted with SPSS 19 analysis program for windows for analyzing fundamental data by means of descriptive statistics i.e. frequency, percentage.

Step 2 Analyzing the Differential Item Functioning (DIF) and the Differential Test Functioning (DTF)

To investigate the differential item functioning (DIF) and differential test functioning (DTF) by DIFAS 5.0 (Penfield, 2012), which was applied for analysis by Mantel-Haenszel indicated the formula as below:

Mantel-Haenszel common odds ratio for an item at score level j

$$\alpha_j = \frac{p_{Rj}}{q_{Rj}} \bigg/ \frac{P_{Fj}}{q_{Fj}} = \frac{a_j d_j}{b_j c_j}$$

Where,

p_{Rj} = Number of persons in Reference group in score interval j who answered correctly

q_{Rj} = Number of persons in Reference group in score interval j who answered incorrectly

Notation F relates to the focal group

Mantel-Haenszel common odds ratio for item i

- For the slice j $\alpha_j = \frac{a_j d_j}{b_j c_j}$
- Across all slice $\hat{\alpha}_{MH_i} = \frac{\sum_j^s a_j d_j / n_j}{\sum_j b_j c_j / n_j}$
- The logarithm of common log-odds ratio is normally distributed and is used as effect size measure

$$\lambda_{MH} = \log(\hat{\alpha}_{MH})$$

As the analysis of Mantel-Haenszel common log-odds ratio was to compare the response of Reference group with Focal group, criteria for considering differential functioning of items were obtained from LOR Z value and MH LOR value according to the process of interpreting as below:

- 1) Considering LOR Z value

- 1.1 LOR Z value > 2 or LOR Z value < -2 indicates that can consider evidence of presence of DIF and MH LOR value in the step 2 shall be considered.
 - 1.2 $-2 \leq \text{LOR Z} \leq 2$ indicates no evidence of presence of DIF.
 - 2) Considering MH LOR value for the question with differential functioning between groups
 - 2.1 Positive (+) MH LOR value indicates that item favor male students.
 - 2.2 Negative (-) MH LOR value indicates that item favor female students.
- Examining differential test functioning were obtained from Tau^2 value. (The psychometrics center, 2010)
- 3) Considering Tau^2 value for the differential test functioning between groups
 - 3.1 $\text{Tau}^2 < 0.07$ indicates that find DTF in the low level.
 - 3.2 $0.07 < \text{Tau}^2 < 0.14$ in indicates that find DTF in the medium level.
 - 3.3 $\text{Tau}^2 > 0.14$ indicates that that find DTF in the large level.

Results

Results of analyzing the differential item functioning (DIF) and the differential test functioning (DTF) of SWUSAT for groups of students by gender were summarized as follows:

Data of Year 2011

The result of differential item functioning (DIF) of verbal test 1 (V1) found 5 from 30 items that can consider evidence of presence of DIF. There are 4 items indicates that item favor female students and only 1 item indicates that item favor male students. The results of differential test functioning (DTF) for verbal test 1 (V1) showed that $\text{Tau}^2 = .035$ which was less than $.07$ indicates that found DTF in the low level.

The results of differential item functioning (DIF) of verbal test 2 (V2) found 4 from 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor male students and 2 items indicates that item favor female students. The results of differential test functioning (DTF) for verbal test 2 (V2) showed that $\text{Tau}^2 = .062$ which was less than $.07$ indicated that found DTF in the low level.

The results of differential item functioning (DIF) of number test 1 (N1) found 3 from 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor female students and only 1 item indicates that item favor male students. The results of differential test functioning (DTF) for number test 1 (N1) showed that $\text{Tau}^2 = .011$ which was less than $.07$ indicated that found DTF in the low level.

The results of differential item functioning (DIF) of number test 2 (N2) found 5 from 30 items that can consider evidence of presence of DIF. There are 3 items indicates that item favor male students and 2 items indicates that item favor female students. The results of differential test functioning (DTF) for number test 2 (N2) showed that $\text{Tau}^2 = .029$ which was less than $.07$ indicated that found DTF in the low level.

The results of differential item functioning (DIF) of reasoning test 1 (R1) found 8 from 30 items that can consider evidence of presence of DIF. There are 5 items

indicates that item favor female students and 3 items indicates that item favor male students. The results of differential test functioning (DTF) for reasoning test 1 (R1) showed that $Tau^2 = .042$ which was less than .07 indicated that found DTF in the low level.

The results of differential item functioning (DIF) of reasoning test 2 (R2) found 14 from 30 items that can consider evidence of presence of DIF. There are 7 items indicates that item favor male students and 7 items indicates that item favor female students. The results of differential test functioning (DTF) for reasoning test 2 (R2) showed that $Tau^2 = .123$ which was more than .07 but less than .14 indicated that found DTF in the medium level.

The results of differential item functioning (DIF) of spatial test 1 (S1) found 3 from 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor male students and only 1 item indicates that item favor female students. The results of differential test functioning (DTF) for spatial test 1 (S1) showed that $Tau^2 = .053$ which was less than .07 indicated that found DTF in the low level.

The results of differential item functioning (DIF) of spatial test 2 (S2) found 4 from 30 items that can consider evidence of presence of DIF. There are 3 items indicates that item favor male students and 1 item indicates that item favor female students. The results of differential test functioning (DTF) for spatial test 2 (S2) showed that $Tau^2 = .053$ which was less than .07 indicated that found DTF in the low level.

The results of differential item functioning (DIF) of spatial test 3 (S3) found 5 from 30 items that can consider evidence of presence of DIF. There are 3 items indicates that item favor female students and 2 items indicates that item favor male students. The results of differential test functioning (DTF) for reasoning test 3 (S3) showed that $Tau^2 = .091$ which was more than .07 but less than .14 indicated that find DTF in the medium level.

The results of differential item functioning (DIF) of spatial test 4 (S4) found 4 from 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor male students and 2 items indicates that item favor female students. The results of differential test functioning (DTF) for reasoning test 4 (S4) showed that $Tau^2 = .443$ which was more than .14 indicated that found DTF in the large level.

Data of Year 2013

The result of differential item functioning (DIF) of verbal test 3 (V3) found 9 from 30 items that can consider evidence of presence of DIF. There are 6 items indicates that item favor female students and 3 items indicates that item favor male students. The results of differential test functioning (DTF) for verbal test 3 (V3) showed that $Tau^2 = .021$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of verbal test 4 (V4) found 3 of 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor male students and 1 item indicates that item favor female students. The results of differential test functioning (DTF) for verbal test 4 (V4) showed that $Tau^2 = .018$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of verbal test 5 (V5) found 2 of 30 items that can consider evidence of presence of DIF. There are 1 item indicates that item favor male students and 1 item indicates that item favor female students. The results of differential test functioning (DTF) for verbal test 5 (V5) showed that $Tau^2 = .007$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of number test 3 (N3) found 5 of 35 items that can consider evidence of presence of DIF. There are 3 items indicates that item favor female students and 2 items indicates that item favor male students. The results of differential test functioning (DTF) for number test 3 (N3) showed that $Tau^2 = .020$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of number test 4 (N4) found 6 of 35 items that can consider evidence of presence of DIF. There are 4 items indicates that item favor female students and 2 items indicates that item favor male students. The results of differential test functioning (DTF) for number test 4 (N4) showed that $Tau^2 = .011$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of reasoning test 3 (R3) found 16 of 30 items that can consider evidence of presence of DIF. There are 8 items indicates that item favor male students and 8 items indicates that item favor female students. The results of differential test functioning (DTF) for reasoning test 3 (R3) showed that $Tau^2 = .081$ which was more than .07 but less than .14 indicated that found DTF in the medium level.

The result of differential item functioning (DIF) of reasoning test 4 (R4) found 9 of 30 items that can consider evidence of presence of DIF. There are 5 items indicates that item favor female students and 4 items indicates that item favor male students. The results of differential test functioning (DTF) for reasoning test 4 (R4) showed that $Tau^2 = .036$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of spatial test 5 (S5) found 3 of 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor female students and 1 item indicates that item favor male students. The results of differential test functioning (DTF) for spatial test 5 (S5) showed that $Tau^2 = .021$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of spatial test 6 (S6) found 3 of 30 items that can consider evidence of presence of DIF. There are 2 items indicates that item favor male students and 1 item indicates that item favor female students. The results of differential test functioning (DTF) for spatial test 6 (S6) showed that $Tau^2 = .017$ which was less than .07 indicated that found DTF in the low level.

The result of differential item functioning (DIF) of spatial test 7 (S7) found 18 of 30 items that can consider evidence of presence of DIF. There are 10 items indicates that item favor male students and 8 items indicates that item favor female students. The results of differential test functioning (DTF) for spatial test 7 (S7) showed that $Tau^2 = .092$ which was more than .07 but less than .14 indicated that found DTF in the medium level.

According to the details below:

Table 3: The Result of DIF and DTF Analysis Classified by Tests

| Year | Test | Number of Item | Number of item that found DIF | | | % | Level of DTF |
|------|------|----------------|-------------------------------|--------------|-------|-------|--------------|
| | | | Favor male | Favor female | Total | | |
| 2011 | V1 | 30 | 1 | 4 | 5 | 16.67 | Low |
| | V2 | 30 | 2 | 2 | 4 | 13.33 | Low |
| | N1 | 30 | 1 | 2 | 3 | 10.00 | Low |
| | N2 | 30 | 3 | 2 | 5 | 16.67 | Low |
| | R1 | 30 | 3 | 5 | 8 | 26.67 | Low |
| | R2 | 30 | 7 | 7 | 14 | 46.67 | Medium |
| | S1 | 30 | 2 | 1 | 3 | 10.00 | Low |
| | S2 | 30 | 3 | 1 | 4 | 13.33 | Low |
| | S3 | 30 | 2 | 3 | 5 | 16.67 | Low |
| | S4 | 30 | 2 | 2 | 4 | 13.33 | Large |
| 2013 | V3 | 30 | 3 | 6 | 9 | 30.00 | Low |
| | V4 | 30 | 2 | 1 | 3 | 10.00 | Low |
| | V5 | 30 | 1 | 1 | 2 | 6.67 | Low |
| | N3 | 35 | 2 | 3 | 5 | 14.29 | Low |
| | N4 | 35 | 2 | 4 | 6 | 17.14 | Low |
| | R3 | 30 | 8 | 8 | 16 | 53.33 | Medium |
| | R4 | 30 | 4 | 5 | 9 | 30.00 | Low |
| | S5 | 30 | 1 | 2 | 3 | 10.00 | Low |
| | S6 | 30 | 2 | 1 | 3 | 10.00 | low |
| | S7 | 30 | 10 | 8 | 18 | 60.00 | Medium |

Discussion

The Investigation of Differential Item Functioning and Differential Test Functioning of SWUSAT during 2010-2013” was concluded as follows:

The results of differential item and test functioning of SWUSAT for verbal tests, number tests, reasoning tests and spatial tests all 10 tests in year 2011. The percentage of item number that found DIF between 10% - 46.67%. There are 4 tests that most of items favors female group. There are 3 tests that most of items favors male group and there are 3 tests that the number of items favors female group equal favors male group. For Differential test functioning (DTF), there are 8 tests that have small level, 1 test for medium level and large level. For all 10 tests in year 2013, The percentage of item number that found DIF between 6.67% - 60%. There are 5 tests that most of items favors female group and 2 tests that most of items favors male group. There are 3 tests that the number of items favors female group equal favors male group. For Differential test functioning (DTF), there are 8 tests that have small level and 1 test for medium level.

In addition, these findings conformed to the study of Chutima Saengdararat (2002) which compared the results of differential item functioning detection of the scholastic aptitude test in accordance with the familiarity, interest and pleasantness

in tests by different detecting methods. The results showed that the number of items with differential functioning of reasoning scholastic aptitude test through different detecting method was able to indicate the items with differential functioning. There was the statistic significant difference at the level of 0.5 among the criterion group of gender and interest whereas there was no statistic significant different among the criterion group of familiarity and pleasantness.

Furthermore, Giray (1995) analyzed the differential item functioning for entrance examination test with calculation, vocabulary and geometry by gender and socioeconomic status. The results showed that items with different functioning were those items in mathematics in case of both gender and socioeconomic status. For gender, the results indicated that male students had advantage in items with calculation questions whereas female students had advantage in items with vocabulary and geometry questions. It was showed that female students had better language or vocabulary skills than male students who had better calculation skills than female students.

Therefore, each of scholastic aptitude test for verbal factor, number factor, reasoning factor and spatial factor contained the items with differential functioning by characteristics of gender.

Suggestion

According to the Investigation of Differential Item Functioning and Differential Test Functioning of SWUSAT during 2010-2013”, the results can be applied in different situations and the suggestions for application of results and for further research were as below:

1. Suggestion for Application of Results are Divided into 2 Parts as Follows:

1.1 The analysis of differential item functioning and differential test functioning can be applied in the construction and development of items and tests to ensure good standard and enhance the reliability of assessment tools in the educational institution.

1.2 The test bank is another channel for development of measurement and assessment that many educational institutions are working on. The difficulty index and discrimination index are normally considered for the selection of tests which are to be kept in the test bank. The analysis of differential item functioning is another method to create reliability to the test bank of educational institution.

2. Suggestion for Further Research

The data obtained from this research is the secondary data from the SWUSAT project in the academic years 2010-2013 with focus on the differential item functioning (DIF) and differential test functioning (DIF) of correct choices only. The suggestion for further research is the analysis of differential distractor functioning (DDF) in order to increase more information for construction and development of standard tests.

References

Camilli G., & Shepard L. A. (1994). *Method for identifying biased test items*. California: SAGE.

- Chutima Saengdararat (2003). *A comparison of scholastic aptitude test of differential item functioning detection by familiarity interest and pleasantness in tests using by different detecting methods*. Master thesis. Graduate school. Srinakharinwirot University.
- Giray B. (1995). Differential Item Functioning (DIF) Analysis of Computation, Word Problem and Geometry Questions across Gender and SES Groups. *Studies in Educational Evaluation*. 21(4): 439–456.
- Penfield R. D. (2008). An Odds Ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement*. 45(3): 247–269.
- Penfield, R. D. (2012). *DIFAS 5.0 Differential item functioning analysis system: User's manual*, unpublished manuscript.
- Sirichai Kanjanawasi. (2012). *Modern Test Theories*. 4rd Edition. Bangkok: Chulalongkorn University Press.
- The Psychometrics Centre (2010). *Summer School in Applied Psychometrics Principles*. University of Cambridge.