Chapter 5 Conclusion and Suggestion

The test of the basic competence test for junior high school students was the first time administered since 2001 in order to replace the traditional entrance exam of high school. How is the quality of the test of the mathematics subject? Is it fair to genders? And how is the performance difference between genders? The goals of the study are to investigate the gender differences in performance on mathematics achievement items and DIF of gender relating to this system of testing.

5.1 Conclusion

In summary, the current investigation leads to several important results:

1. The female did slightly better mathematically performed in the three categories. It is significant in the algorithmic category and overall but the effect size is small.
2. The signed area and model comparison measures are likely to yield similar results in identifying items with significant DIF when statistical significance tests are applied to the area measure.
3. In general, the quality of the mathematic subject of the students' basic competence test for junior high school students is good. Only one Item 13 and 26 are identified to be DIF, and Item 24 is close the criteria of DIF. Only item 13 has to be revised after we checked the item.
4. It appears that the item b-value (after transformed) difference plays a significant role in determining whether an item is DIF or not.

5.2 Suggestion

I provide the following suggestions by the above conclusion and the process of this study.

1. The suggestions of detected DIF method

In the study, I only use two methods – area measure and model comparison measure to detect the DIF items and compare their results. In fact, M-H methods and SIBTEST are also used by public, although they are not based on IRT. In the future, we compare all the methods and compare their consistency. In addition, we can discuss the results from the four methods in order to find which is the most effective.

2. The suggestions of test application

To detect DIF items is an important part of test. The proportion of DIF items is low in the first test of the basic competence test for junior high school students. We hope the institution of the basic competence test for junior high school students can run the DIF more detail and find the real bias item after they conducted pilot-test and got data. In addition, they can run the past data of test and find the type of DIF items in order to provide the reference for the future test designer.

3. The suggestions of future research

It is hard to avoid sample bias because we only sample 5,000 from the 288,368 populations. The reason is because of the limitations of time and other factors. We can sample more samples or repeat randomly samples and compare the results and consistency.

The study only investigates the DIF of mathematics subject. We can extend to other subjects in the future research.

In the study we emphasize the detecting DIF of mathematics test. If the item is distinguished as bias, it means the item has to be logically

analyzed by the experts after it is detected as DIF. I judge whether or not the DIF items are bias by myself. If I can consult with test designers or experts to distinguish bias items, it will help to improve the quality of the test items.