Differential Item Functioning in the Aspect of Gender Differences in Reading Literacy

Andrejs Geske, University of Latvia, andrejs.geske@lu.lv

Antra Ozola, University of Latvia, antra.ozola@lu.lv

Abstract

The purpose of the study is to analyze the IEA PIRLS 2006 set of items in order to check if a differential item functioning (DIF) can be noticed. The groups of comparison are based on gender since PIRLS 2006 study has showed that Latvia has the 6th largest gender gap in reading literacy scores. So the DIF analysis is expected to show if an instrument could be one of the reasons of the gender gap in achievement.

Items easiest and hardest for each sex are found as well as items with the biggest and smallest DIF values to check if an item format gives any advantages or disadvantages to any of the gender groups of Latvian students.

The results show that the set of PIRLS 2006 items does not have any gender bias on Latvian students since there are items with no DIF at all and there is almost equal number of items with DIF in favor of each of the sexes.

The length of the item stem does not have any correlation with the item difficulty or differential item functioning between sexes. There is no difference in DIF when the items of literary and informative texts are compared and the same is true for multiple choice and constructed response items.

The main conclusion is that the achievement gap between boys and girls in Latvia is caused by the differences in their skills/abilities and not by a differential item functioning.

Keywords: *secondary analysis, differential item functioning, PIRLS, reading literacy, gender differences*

Introduction

IEA PIRLS 2006 study in Latvia has showed that it has the 6th largest gender gap in reading literacy scores. Although IEA PIRLS tests are absolutely professionally made and reliable, validation of the test as an instrument is extremely important before addressing the research problem itself. So, before analyzing gender differences in reading it is valuable to check if the

instrument of the measurement is free of gender biases. Only after proving that differences in abilities (and not tests) are the ones which cause the gender gap in achievement it is possible to start a proper analysis of factors behind these abilities.

The purpose of the paper is to find out if an instrument (PIRLS 2006 item set) could be partly blamed for the achievement gap between boys and girls in Latvia. The proposed hypothesis is that the gender differences in achievement scores is caused by the differences in average abilities of boys and girls in Latvia and not by differential item functioning. The research question is: is there a differential item functioning in the PIRLS 2006 study comparing boys and girls and does it give an advantage to any of genders?

Theoretical framework

When speaking about the gender differences in achievement usually issues about the measurement tests are raised. Like: If boys do better in multiple choice items, does a test with no constructed responses gives a smaller gender differences in scores? If girls tend to write longer constructed responses, do they get higher scores than boys who are more lazy writers? So the doubts about the way reading literacy is measured show up.

Michael Gurian declares that girls do better in essay like tasks but boys in multiple choice items since males are generally are better at deduction skills which are necessary for picking one answer from a set (Gurian, 2001).

This thought is also supported by Nancy Cole and Warren Willingham who have found that female respondents score higher in tasks which involve a longer writing and explanation (Willingham, Cole, 1997).

In a study where the data of OECD PISA 2000 study were analyzed also revealed that the gender difference in literacy scores was bigger for constructed response items than for multiple choice items. The item type has a stronger influence on the gender gap in achievement for the items of higher cognitive domains (Lafontaine, Monseur).

Differential item functioning (DIF) methods allow one to judge whether items (and ultimately the test they constitute) are functioning in the same manner in various groups of examinees (Zumbo, 2007). Or in other words, using the DIF analyses it is possible to check if there is a difference in item difficulty for people of two groups with the same level of ability.

So the DIF analysis can help to solve the problem about possible gender biases in item formats.

Methodology

In this study the Latvian data form PILRS 2006 study have been used. The performed analysis of Differential Item Functioning (DIF) is based on the one parameter (Rasch) model of Item Response Theory (IRT). So only a difficulty of the item and students abilities are taken into account in this model.

For scaling and analysis ACER ConQuest 2.0 software has been used.

All items initially being coded in a scale from 2 to 4 points have been recoded to dichotomous. All incorrect answers have been left a score 0 but any level of correct answer has been recoded to 1. This has been done mainly with a purpose to avoid issues mentioned in the Theoretical framework above about girls tending to write more complete answers than boys. Besides a study performed by Schwippert, Bos and Lankes (2004) has proved a correlation 0.88 for item difficulty parameters when 1) item set contained dichotomous and partial credit items and 2) all partial credit items in the set were recoded as dichotomous.

Finding and Discussion

Item Difficulty and Student Ability

As it might have been expected knowing the gender gap in reading achievement in Latvia, a statistically significant difference in abilities between boys and girls was found. The gender difference in ability estimates for the whole PIRLS 2006 item set was 0.38 (0.008) measured in the traditional Rasch scale with average being a 0 and ability values usually varying form -3 to 3.

The difficulty of the PIRLS 2006 items for Latvian students was between -2.32 (0.07) and 2.42 (0.06). So it can be stated that the item set covers a wide range of abilities.

The easiest item for all Latvian students was an extended response three point item:

Give three ways penguins are able to keep warm in Antarctica.

The most difficult item for all Latvian students was 1 point item where a sequence of operations has to be provided:

Number the steps in the order you would follow to make a wormery. The first one has been done for you.

put the bottle in the shoebox

1 *poke holes in the top of the shoebox*
drop in the worms
add potato and onion
fill the bottle with soil and sand

These results lead to a conclusion that for Latvian students it has been easier to give a correct answer to more complicated questions that to easy ones. This is corresponds with the information published in the PIRLS 2006 International Report where Latvian students have scored higher in items of interpreting, integrating, and evaluating cognitive domains than in items of retrieving and straightforward inferencing cognitive domains (Mullis et al., 2007).

So in Latvia student achievement is higher and item difficulty lower for items which are meant to measure higher cognitive domains. And there is no difference in gender gap depending on item's cognitive domain – girls outperform boys equally in easy as well as in difficult items (Mullis et al., 2007).

The most difficult item for Latvian boys was the above mentioned item about making a wormery which was also the most difficult item for all Latvian students (difficulty estimate for girls 2.01, for boys 2.82).

The most complicated item for girls was a one point short answer question (difficulty estimate for girls 2.29, for boys 2.34):

Explain why putting the onion and potato on the surface of the soil is important to the wormery project.

There is quite a difference in the difficulty estimates of boys and girls in the "boys' item" and in the "girls' item". Is there a reason to think about the differential item functioning?

Differential Item Functioning

As mentioned earlier, differential item functioning appears where respondents from different groups, let us say boys and girls, have the same ability but different probability to solve an item correctly.

Out of 125 PIRLS 2006 analyzed 63 did not show any DIF, the other 62 item split in 30 favoring girls and 32 favoring boys. So in terms of DIF the set of PIRLS 2006 items was perfectly balanced for Latvian students (Fig.1).

A DIF estimate in this case were calculated for every item and every gender (150 altogether) was represented by a value which has to be added to the overall item difficulty estimate

thereby getting an item difficulty for a particular gender. The absolute values of the DIF estimates of particular item for both genders are equal, but the arithmetical sighs are opposite (manifesting as horizontal symmetry in Fig.1).

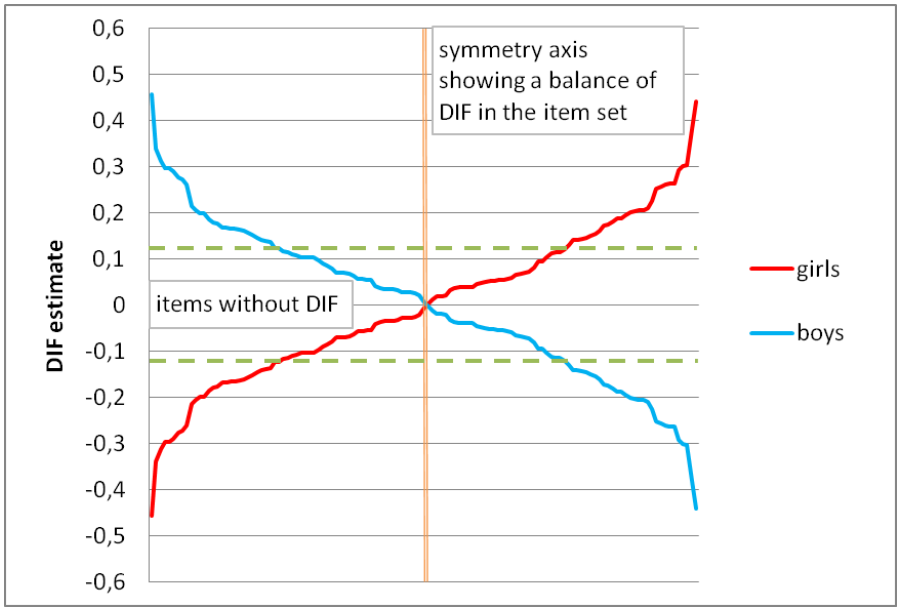

Fig.1: DIF estimates for each item per gender

An item with no evidence of DIF at all is a one point item:

How do pill bugs find the food?

A They walk down the passage.

B They sense food with their antennae.

C They follow the scent trail.

D They see the food in the dark.

The distance between item difficulties of boys and girls with the same abilities was about 0.01 (0.06).

DIF and Other Item Characteristic

The Pearson's correlation between item difficulty and DIF (represented as distance between item difficulties of boys and girls with the same abilities) was negative (-0.30) and statistically significant. So the easier the item the bigger the DIF. The higher the item difficulty the lower differential item functioning between sexes.

The item length in words did not have any significant correlation neither with DIF neither with item difficulty as such.

T-test was used to check if there is more evidence of DIF in literary or informative passages and no statistically significant difference was found. The same test was done for multiple choice and free answer question types and no difference in DIF by item format was found.

There was a statistically significant negative correlation (-0.18) of DIF and levels of comprehension. This result is in line with the above mentioned negative correlation between DIF and the item difficulty.

After checking that the DIF values in each of four types of comprehension processes are following the normal distribution, the t-test was run to check if there is a difference in DIF for the items meant to address the lowest and the highest of comprehension processes. And a significant difference (α=0.1) was found. Items from a comprehension process for examination and evaluation of content, language, and textual elements had smaller DIF than items from a comprehension process for focusing on and retrieving explicitly stated information.

Conclusions and Implications

PIRLS 2006 test was absolutely appropriate for measuring reading literacy abilities of Latvian students, both – boys and girls. Half of the items did not show any differential item functioning another half was equally split in items favoring girls and favoring boys.

The question length was neither related to item difficulty neither to differential item functioning.

There is no difference in differential item between literary and informational texts and between multiple choice and free answer questions.

The higher the item difficulty the lower the differential item functioning which means that if abilities of both genders are in the same advanced level, items to measure them work more similar for boys and girls than in the case of low abilities.

The achievement gap between boys and girls in Latvia is caused by the differences in their abilities and not by a differential item functioning.

References

Gurian, M. (2001). Boys and Girls Learn Differently!: A Guide for Teachers and Parents. San Francisco: Jossey-Bass.

Lafontaine, D., Monseur, C. (2009). Gender Gap in Comparative Studies of Reading Comprehension:

to what extent do the test characteristics make a difference? *European Educational Research Journal, 8*(1), 69-79.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy P. (2007). PIRLS 2006 International Report. IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Schwippert, K., Bos, W., Lankes, E. M. (2004). Lesen Mädchen anders? Vertiefende Analysen zu Geschlechtsdifferenzen auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU. *Zeitschrift für Erziehungswissenschaft, 7*(2), 219-234.

Willingham, W. W., Cole, N. S. (1997). Gender and Fair Assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

Zumbo, B. D. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly, 4*(2), 223–233.