European Commission

JRC SCIENTIFIC AND POLICY REPORTS

Differential Item Functioning (DIF): What Functions Differently for Immigrant Students in PISA 2009 Reading Items?

Patricia Dinis da Costa and Luísa Araújo

Report EUR 25565 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Patrícia Dinis da Costa
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 361, 21027 Ispra (VA), Italy
E-mail: patricia.costa@jrc.ec.europa.eu
Tel.: +39 0332 78 9173
Fax: +39 0332 78 5733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC75935

EUR 25565 EN

ISBN 978-92-79-26789-5

ISSN 1831-9424

doi:10.2788/60811

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

Differential Item Functioning (DIF): What Functions differently for Immigrant Students in PISA 2009 Reading Items?

INTRODUCTION

After the European Council of March 2008 and the subsequent Green Paper calling for the integration of immigrants through education, the European Commission has set as a priority the monitoring of the achievement gap between native learners and learners with a migrant background, using existing data and indicators. The main indicator of students´ performance across Europe is the Programme for International Student Assessment (PISA). Launched in 2000 by the Organization for Economic Co-operation and Development (OECD), PISA is implemented every three years and assesses students´ achievement in reading, mathematics and science. Reading has been the main area assessed in 2009 as well as in 2000, whereas Science was the focus in 2006 and Mathematics was the focus in 2003. The 2011 results will only be available at the end of 2012 and again they reflect a focus on Mathematics.

Although each assessment cycle presents a more complete picture of only one of the knowledge areas and the related ability of 15 year-olds to either enter the work force or continue further studies, it is possible to study trends in performance over time because PISA tests are reliable and valid measures that include anchor, or repeated, items that are used in all assessment rounds. Thus, after four PISA cycles, it was possible for OECD to conclude that across OECD "…the percentage of students with an immigrant background increased by an average of two percentage points between 2006 and 2009. The performance gap between students with and without an immigrant background remained broadly similar over the period. However, some countries noted large reductions in the performance advantage of students without an immigrant background. In Belgium, Switzerland and Germany, the gap narrowed by between 28 and 38 score points, due to improvements in reading proficiency among students with an immigrant background. However, the gap is still relatively wide in these countries (OECD, 2010, p. 20).

One factor specific to immigrants, both first and second generation[1], is that they often do not speak the language of the country of residence/language of test at home. However, in PISA 2003, 2006 and 2009 even after accounting for the language spoken at home differences between immigrant and native students are still significant in the majority of OECD countries with more than 3% of immigrant student population. The commonly adopted distinction between 1st and 2nd generation immigrants for the purpose of studying educational outcomes stems from consistent findings indicating that native-born children of immigrants – second generation children born in the country of assessment - tend to perform better than their counterparts that were not born in the country of assessment (1st generation). Furthermore, PISA data (OECD, 2010) shows that the average educational attainment of immigrant students is lower than that of natives and that within the immigrant category second generation students have higher achievement than first generation students (Liebig & Widmaier, 2009). Also, although first generation immigrants present a wider proportion of students at each end of PISA's benchmark distribution scale, "first-generation students – those who were born outside the country of assessment and who also have foreign-born parents – score, on average, 52 score points below students without an immigrant background" (OECD, 2010, p. 10).

Considering that 39 points in PISA is the equivalent to one year of schooling, it is clear that fifteen year-old immigrant students and first generation ones in particular, are far behind their native counterparts in terms of educational achievement. However, immigrant populations are quite diverse in terms of country of origin and of the cultural capital they bring with them to the host country. As Liebig and Widmaier (2009) refer, several empirical studies indicate that human capital is transmitted from generation to generation and factors such as cultural beliefs and attitudes, occupational status of the parents and their educational level may also explain differential achievement among sub-groups of the school population. In fact, an in-depth analysis of immigrant students´ achievement based on PISA 2003 data revels that in many countries the father's

[1] First generation immigrants include children of foreign born parents who were themselves born abroad. Second generation students are those born in the country of assessment that have foreign-born parents.

occupational level (ISCO) and parental educational explains some of the variance in student achievement (OECD, 2006).

In addition to the diversity found among the immigrant student population in PISA, the comparison between educational outcomes for immigrant students versus native ones in European countries is hindered by the small sample size of the immigrant population in many countries. For example, while in Luxemburg, Germany, France, Austria, Belgium, the Netherlands and Sweden there were more than 10 percent of migrant students in all PISA rounds in Denmark the immigrant student participation rate has been consistently below 10 percent. In Estonia and Slovenia – participating countries after 2006 – immigrant students are over 10 percent, but in countries like Greece, Portugal and Spain they are well below this percentage.

Purpose

The purpose of the present analysis is to use Differential Item Functioning (DIF) to identify differences in the performance of native and immigrant students in PISA 2009 that can be directly related to their responses to particular items. We analyzed 95 cognitive *Reading* items, administered to students in 29 European countries. Different items reflect a range of reading tasks in diverse text formats; answering questions related to interpreting short stories (continuous texts), tables and graphs (non-continuous texts) and travel brochures (multiple texts), and a range of aspects relating to reading difficulty with …"the easiest of the tasks, retrieve, requiring students to locate explicitly stated information according to a single criterion where there is little, if any, competing information in the text, or to identify the main theme of a familiar text (OECD, 2007, p. 285). In relation to all these dimensions of the PISA reading framework, students respond to different text types, from description, narration, and exposition to argumentation in different situations or for different purposes according to diverse texts and item formats (e.g. multiple choice and constructed response).

Since a large number of students, with a variety of cultural, ethnic, gender, socioeconomic status and/or curricular characteristics participate in PISA it is possible to investigate differences

among groups. These characteristics may influence the students' answers to some items. For instance, if an item is strongly related to a cultural aspect of a country, the item can be easier for students in this country than for students from another country that took the same test. When such an occurrence is found and there is a significant difference in the way items are answered by two or more distinct groups, such a phenomenon is called differential item functioning (DIF). DIF can occur when one group of individuals responds differently from another group on a given questionnaire item, even though both groups are equivalent. In general, one of the groups is fixed as the reference group (typically it comprises individuals whom the researcher suspects the test favors) and the other one is considered the focal group. For the focal group, item functioning is compared with that of the reference group.

Our analysis is intended to determine if DIF occurs between native students and immigrant students in PISA 2009 because we suspected that the test favours native students and because we wanted to investigate whether, if and when present, DIF presented itself differently in the various dimensions of the reading assessment framework. Applying this innovative method to this large scale survey can provide insights and further information on the cognitive processing of text by the group of immigrant students typically at a disadvantage in terms of school achievement.

The PISA Assessment Framework in Reading

With respect to *item format*, PISA reading items include a variety of items; the conventional *multiple choice* format and a *complex multiple choice*. The latter usually present response options like True/False with respect to alternative scenarios or statements. Three other types of items require students to write their answers to the questions. They are *short response, closed constructed response* and *open constructed response* items. In a typical short response question students may be asked to draw a route on a map or to fill in a missing word in a sentence. In a closed constructed response only a short answer with information from the text is called for and in an open constructed question students are asked to interpret the question by proving a written response in their own words.

The reading *aspect* in PISA relates to the difficulty level of the items and their corresponding proficiency levels with the *access and retrieve* aspect assessing the lowest benchmark proficiency levels (1 & 2), followed by the *Integrate and interpret* level (3 & 4) and with the *Reflect and evaluate* levels at the highest text processing level (5 & 6). These aspects call for a range of simple to complex processes of reading comprehension and are distributed by the different types of item format. As previously indicated the lowest proficiency levels require only a literal understanding of text whereas the middle ones require students to piece out information and use their own knowledge to answer the items and the top levels require an even more complex, critical and evaluative reading stance.

Regarding *text format* PISA considers the following formats: 1) *continuous texts* which are formed by sentences organized into paragraphs and include newspaper reports, essays, novels, short stories, reviews and letters, 2) *Non-continuous texts* where sentences are the smallest unit and are represented by lists, tables, graphs, diagrams advertisements, schedules, catalogues, indexes and forms, 3) *Mixed texts* which include both continuous and non-continuous texts and reflect a combination or variety of presentation, combining lists, paragraphs of prose and often graphics and 4) *Multiple Texts* which may include both continuous and non-continuous and present a unique situation in that they have been generated independently but should be processed simultaneously (e.g. two travel websites with information for tourists).

Regarding *text types* the PISA framework specifies the typology as follows. First, "*Narration* is the type of text where the information refers to properties of objects in time. Narration typically answers questions relating to *when*, or *in what sequence*. Why characters in stories behave as they do is another important question that narration typically answers. *Exposition* is the type of text in which the information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analysed. The text provides an explanation of how the different elements interrelate in a meaningful whole and often answers questions about *how*"(OECD, 2009, p. 32). Further to this traditional text classification PISA also includes *argumentation*, of

which opinionative and persuasive texts are an example, and *description* and *instruction* types of texts. Descriptive texts are those that focus on describing objects, for example, while the instruction type explains how to solve a particular problem.

Lastly, the PISA assessment framework considers the reading *situation* by distinguishing among a variety of categories: 1) *personal* or reading for private use; 2) *public* or reading for public use; 3) *occupational* or reading for work and 4) *educational* or reading for education. The personal category relates to satisfying personal interests such as reading a novel or an article to satisfy curiosity while the public one relates to readings which subject matter address larger societal issues in public notices. An example of an occupational reading task "… is one that involves the accomplishment of some immediate task. It might include searching for a job, either in a print newspaper's classified advertisement section, or on line" (OECD, 2009, p.26). Educational reading, on the other hand, is specific to the purpose of acquiring information from print, typically from a textbook. As the PISA framework puts it: "Educational reading normally involves acquiring information as part of a larger learning task. The materials are often not chosen by the reader, but instead assigned by an instructor" (OECD, 2009, p.26).

According to the different dimensions of the PISA assessment framework, we focused on investigating whether differential item functioning could be found in reading according to the following categories: 1) *item format*, 2) aspect, 3) *text* format, 4) text *type*, and 5) *situation*. We limited our analysis to reading items because this was the focus of the latest PISA survey and, as such, the number of Mathematics items is much smaller, amounting to only 45 items. Also, due to the small number of immigrant students per country we had to run the analysis combining all sampled students from the 29 countries and without distinguishing between first and second generation background.

Methodology

Several methods have been developed to measure DIF and they can be classified in two major categories: classical or modern. Within the first category, it is possible to detect DIF using a Mantel-Haenszel based statistical procedure (Holland & Thayer, 1988) or the logistic regression method (Swaminathan & Rogers, 1990) while according to modern methods we can use item parameters from item response theory (IRT) models. The second category provides important advantages over classical methods. When the assumption of unidimensionality is met, invariance and information are the two main advantages of IRT over the classic approach. Invariance guarantees that item parameters are independent of the sample and examinee ability is independent of the items. IRT methodology is based on the simultaneously estimation for both item and examinee characteristics and it guarantees that the invariance property is achieved. Within the IRT framework the characteristics of each item are estimated or calibrated independent of an examinee's ability as measured by a latent construct. IRT methodology is used for estimating the probability that a particular examinee will respond correctly to an item. IRT identifies patterns of response and uses statistical models to predict the probability of answering an item correctly as a function of the examinees' achievement in answering other questions. The items characteristics are scaled and expressed on the same continuum as the examinee's ability, which is an added advantage for detecting DIF.

In PISA, the item parameters and the cognitive scales estimation are based on IRT techniques. Thus, in this work we assess DIF between native and immigrant students using a procedure based on IRT. In particular, we use the procedure based on the comparison of the metric parameters of the items (Holland & Wainer, 1993; Thissen, Steinberg & Wainer, 1993), in which we compare the Item Characteristic Curve (ICC) for both groups. An item is said to present DIF if the ICC is not the same for the compared groups, whose examinees have the same level of ability (Kim & Cohen, 1998; Mazor, Hambleton & Clauser, 1998). Typically IRT models can be distinguished by the number of parameters estimated for the items. The most common ones are discrimination,

difficulty and guessing item parameters. Following the approach used in PISA, in our study we analyze the existence of DIF in terms of the difficulty parameter. Moreover, and according to our purposes, we fixed the native students as the reference group and immigrants as a focal group and compared how each item functioned for the latter students compared to how the same item functioned for the reference group (Gamerman, Soares & Goncalves, 2010).

Sample

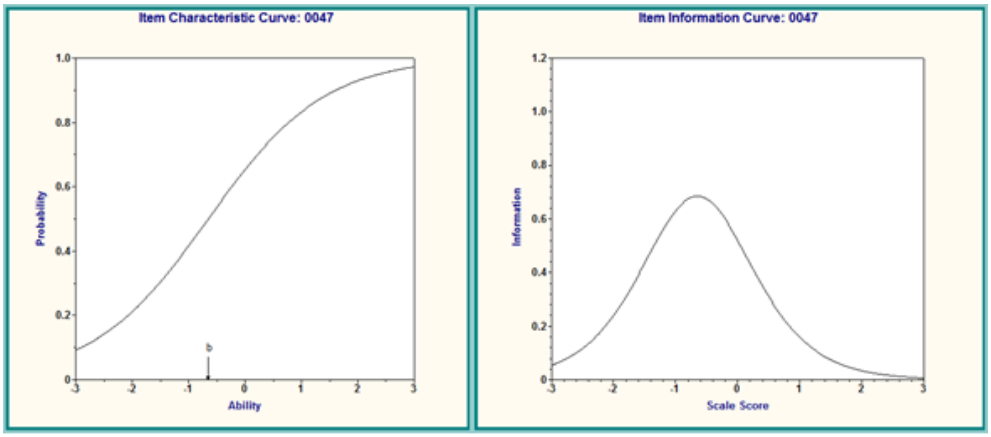
In our analysis we considered 25 EU Countries, plus 3 candidate countries and Norway. Specifically, the countries were: Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden, Turkey and the United Kingdom. The sample is composed of 197885 students, of which 182122 (92%) are natives and 15763 are immigrants (8%). PISA 2009 cognitive reading items are composed of 131 questions, but to guarantee that we have answers to all categories of each item, we took into account only the items that meet this requirement. In the end, we analyzed 95 reading items. In about half of our sample the validity criterion adopted in PISA for comparison purposes between immigrant students and non-immigrant ones is met. Thus, the percentage of students with a migrant background is at least 3%. For the remaining countries, this criterion is not met because these countries do not have a significant immigrant student population.

Analysis and Results

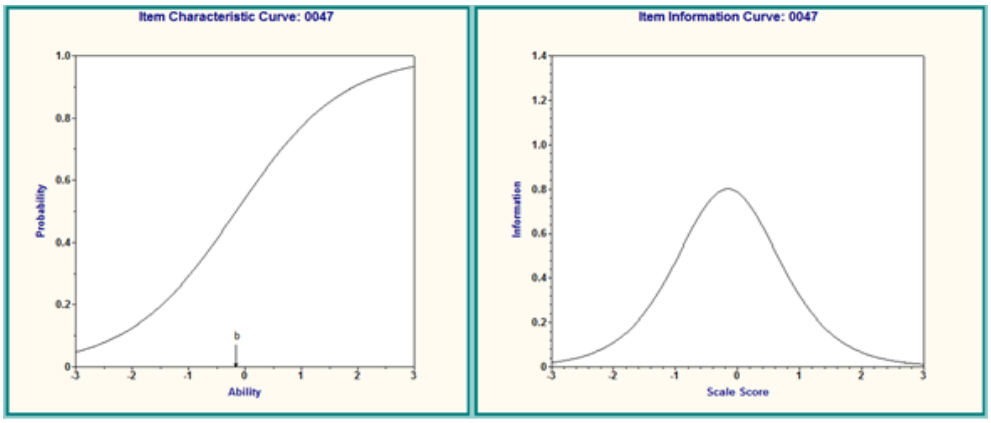
Initially, we estimated the item difficulty parameter for all items, considering separately native students and immigrants students, using the *Partial Credit Model* for polytomous items (Wright & Masters, 1982). In particular, to estimate the difficulty parameter we used the Marginal Maximum Likelihood (MML) (Baker & Kim, 2004) procedure adopted in the Parscale software (Muraki & Bock, 2003).

We found differences between native and immigrant students for some items in terms of item difficulty parameter estimates. Additionally, item characteristic curve (ICC) and the item information function (IIF) show different behavior between the two (Hambleton, Swaminathan & Rogers, 1991). The mean of the items´ difficulty parameter estimates for natives was -0.524 (SD=1.012) and for immigrants was -0.155 (SD=0.990). However, we verified that there are items that are easier for natives than for immigrants and other items that are more difficult for natives than for immigrant students. In Figures 1 and 2 we present examples of ICC and IIF obtained for natives and immigrants for two items. Figure 1 refers to the dichotomous item 47 and it shows that the value on the ability scale (difficulty parameter estimate) the 0.5 probability of students answering the item correctly is less for natives when compared with immigrants. This means that this item is easier for natives than for immigrants. Additionally, the IIF indicates that this item contributes with more information for native students with low and medium achievement whereas for immigrants the item provides more information only for students with a medium reading ability level.

Figure 1: Item 47 – ICC and IIF for natives

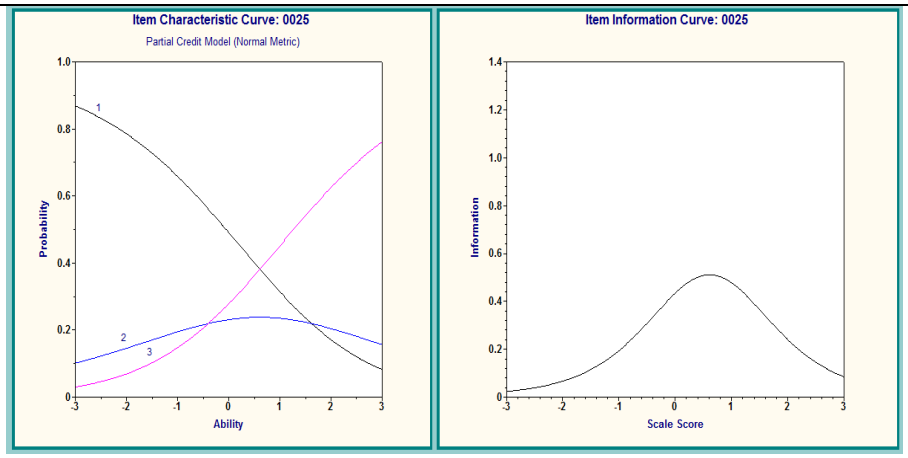


Item 47 – ICC and IIF for immigrants

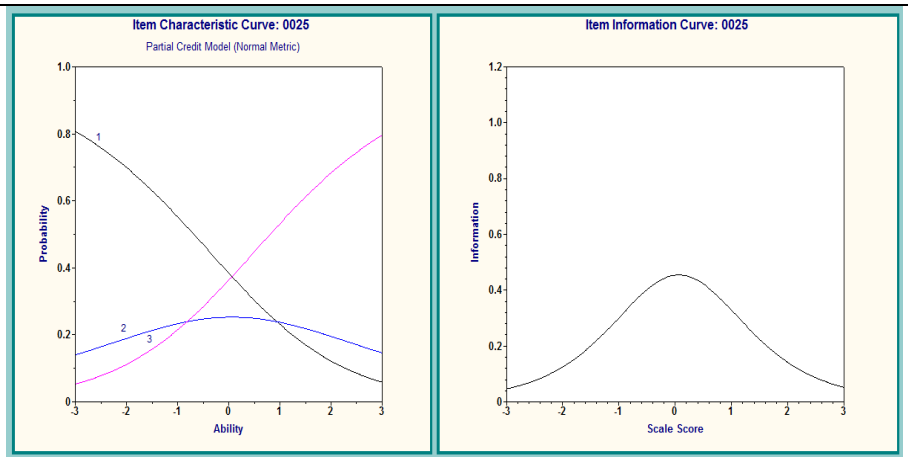


The second item has three categories of answer. The ICC presents 3 lines, the dark line represents the probability associated with the students giving the wrong answer to the item, the blue line refers to the probability of partially answering the item correctly and the pink line is the probability of students answering the item correctly. We can see that higher values are associated with native students' ability to answer the item correctly than it is the case for immigrants. The IIF shows that the curve for the immigrants is approximately centered in zero while the curve for the natives is shifted more to the right compared with the other group.

Figure 2: Item 25 – ICC and IIF for natives



Item 25 – ICC and IIF for immigrants



The initial procedure used for the goodness of fit test was based on the estimation of the item difficulty parameter for all the students (natives and immigrants). The value obtained for this statistic was 5384435.027. The next step was to estimate item difficulty parameter for natives and immigrants, considering natives as the reference group and immigrants as the focal group. The obtained goodness of fit test was 5384206.105. The difference between the goodness of fit values was 228.92 with 94 degrees of freedom (d.f.). This distribution follows a chi-square distribution with

a p-value of zero. This indicates that the results obtained justify the use of DIF analysis to identify how reading items might function differently for the two groups of students.

Differential Item Functioning

The next step of our analysis was the detection of items with DIF. We considered native students as the reference group and immigrants as the focal group and we estimated the difficulty parameter of the items for both groups (Table 1). According to the IRT approach, more difficult items have higher values in this parameter. The table shows that 53% of the items were easier for immigrants than for natives, since the difficulty parameter is lower for the first group when compared with the second group.

In the subsequent step of the DIF analysis, we calculated the contrast using the difference between both groups and the correspondent standard error. The result obtained has a Chi-square distribution (only with one degree of freedom), which was used to test the statistical significance. In the previous table the difficulty parameter estimates in bold are easier for natives and in italics are easier for immigrants and are statistically significant for any level of significance. We found that 33 items (34.74% of the 95 items) present differential item functioning in difficulty, of which 16 were easier for natives (48.5%) and 17 items were easier for immigrants (51.5%). For example, for item 3 the difficulty is -0.46 for immigrants, a higher value than -0.62 observed for natives. This means that this item is easier for natives than for immigrants. For item 5, we found the parameter difficulty values -2.48 and -2.36 which means that this item is easier for immigrants than for natives.

Next, we present the distribution of the items with DIF in terms of item format, aspect, text format, text type and situation.

Item	Difficulty		Item	Difficulty		Item	Difficulty	
	Natives	Immigrants		Natives	Immigrants		Natives	Immigrants
1	-1.821	-1.829	**33**	*-0.441*	*-0.334*	65	0.266	0.219
2	-0.244	-0.296	34	-0.811	-0.768	66	-2.459	-2.492
3	*-0.62*	*-0.467*	**35**	**-1.087**	**-1.179**	67	-1.402	-1.376
4	-1.371	-1.359	36	-1.07	-1.042	68	-0.679	-0.704
5	**-2.369**	**-2.489**	37	-0.993	-0.953	69	-0.084	-0.076
6	-0.741	-0.755	38	-1.065	-1.079	**70**	*-1.948*	*-1.752*
7	-1.33	-1.295	39	0.121	0.09	71	-0.222	-0.202
8	**-0.486**	**-0.622**	**40**	0.83	1.035	72	2.008	1.914
9	**-1.969**	**-2.083**	41	0.02	-0.01	73	-0.836	-0.848
10	-1.593	-1.654	42	0.261	0.271	74	-0.57	-0.544
11	-1.055	-0.996	43	-1.142	-1.146	75	0.115	0.16
12	-0.765	-0.712	**44**	0.717	0.857	76	-1.589	-1.563
13	*-2.02*	*-1.902*	**45**	**-1.506**	**-1.643**	**77**	**-1.172**	**-1.297**
14	-0.576	-0.55	46	-0.011	0.043	**78**	**-0.625**	**-0.689**
15	**-1.412**	**-1.571**	47	-0.628	-0.588	79	-1.295	-1.344
16	0.316	0.343	48	-1.945	-1.89	**80**	0.923	1.087
17	0.566	0.566	49	0.186	0.148	81	-2.035	-2.093
18	0.228	0.261	**50**	**-2.192**	**-2.43**	**82**	*-0.859*	*-0.783*
19	*-1.902*	*-1.627*	**51**	**-1.303**	**-1.411**	83	1.229	1.237
20	**-1.491**	**-1.613**	52	0.282	0.265	**84**	**-2.979**	**-3.117**
21	*1.351*	*1.768*	53	-1.343	-1.4	85	-1.763	-1.839
22	**1.904**	**1.804**	54	-1.332	-1.359	**86**	**-1.544**	**-1.693**
23	-0.715	-0.738	55	-1.619	-1.573	87	-0.061	0.01
24	0.586	0.587	56	-0.894	-0.867	**88**	*-0.559*	*-0.468*
25	*0.053*	*0.233*	57	1.742	1.693	89	-0.671	-0.614
26	-2.095	-2.042	**58**	*-0.371*	*-0.206*	90	-0.824	-0.833
27	*-0.109*	*0.027*	59	-0.314	-0.364	**91**	*-1.547*	*-1.473*
28	-0.776	-0.816	60	1.954	1.942	92	-0.575	-0.548
29	-0.682	-0.685	61	-1.601	-1.652	93	0.055	0.027
30	**-1.502**	**-1.589**	62	-1.287	-1.304	94	2.048	2.013
31	**-1.093**	**-1.24**	63	0.287	0.29	**95**	**-1.568**	**-1.701**
32	*-0.540*	*-0.225*	64	0.565	0.513	**-**	-	-

Table 1 – Item difficulty parameter estimated for natives and immigrants considering the first group as the reference group. Items signaled in **bold** are easier for immigrants and items signaled in *italics* are easier for natives.

Results

Following the specifications in the PISA assessment framework, we present the distribution of the items in terms of their relative difficulty for the two groups of students in the following categorical order: 1) *Item format*, 2) *Aspect*, 3) *Text format*, 4) *Text type* and 5) *Situation*.

Item format	Number of items with DIF easier for	
	Natives	Immigrants
Closed Constructed Response	1	3
Complex Multiple Choice	1	2
Multiple Choice	8	6
Open Constructed Response	4	4
Short Response	2	2
Total	16	17

Table 2: Item format

In what refers to item format, we can verify that the higher numbers of items with DIF are multiple choice items (14). There are 8 multiple choice items easier for natives and 6 for immigrants. For immigrants, the item formats that are easier are closed constructed response and complex multiple choice. With respect to the aspect of text, natives perform better than immigrants in "Integrate and interpret" and "Reflect and evaluate" aspect items while the "Access and retrieve" items are easier for immigrants.

Aspect	Number of items with DIF easier for	
	Natives	Immigrants
Access and retrieve	5	8
Integrate and interpret	9	8
Reflect and evaluate	2	1
Total	16	17

Table 3: Aspect

As for the text format, Table 4 shows that there is a higher number of items that are easier for immigrants than for natives when the text formats are continuous and non-continuous. In contrast, mixed and multiple text format items are easier natives.

Text format	Number of items with DIF easier for	
	Natives	Immigrants
Continuous	9	11
Mixed	3	1
Multiple	1	0
Non-continuous	3	5
Total	16	17

Table 4: Text format

Differences were also found in text type with native students performing better than immigrants in the description text type items. The opposite occurs for the exposition, instruction and narration text type items.

Text type	Number of items with DIF easier for	
	Natives	Immigrants
Argumentation	1	1
Description	6	2
Exposition	5	6
Instruction	2	4
Narration	2	4
Total	16	17

Table 5: Text type

In what refers to the situation of the reading items we found that immigrant students perform better in items related to the educational and occupational situations, whereas native students exhibit better performance in items related to the personal and public situation domains.

Situation	Number of items with DIF easier for	
	Natives	Immigrants
Educational	3	6
Occupational	3	4
Personal	5	4
Public	5	3
Total	16	17

Table 6: Situation

Conclusions and Future Analysis

We found some *Reading* items where immigrants perform better than natives. Additionally, the results show that some *Reading* items present differential functioning between natives and immigrants students and that this can depend of the classification of the items in terms of *question format, text format, aspect, text type* and *situation*. We believe that the main differences found between the performance of immigrant and native students according to *Situation* are the most relevant ones. Immigrant students perform better than native students in educational situations or in contexts where reading serves the purpose of learning or acquiring information and in situations linked to occupational reading or reading that involves accomplishing a task such as looking for a job in a newspaper or to following directions in the workplace. Conversely, native students perform better in personal and public situations that imply reading for recreational purposes and to attend public events such as a concert. This suggests that schooling, and specifically the school curriculum, matters for it reflects the *situation* of "reading to learn" that typically occurs in school. Furthermore, immigrant students perform better in the exposition and instruction types of text, which again are text types likely found in textbooks used in school. As explained in the PISA framework "… a chapter in a textbook might include some definitions, some directions on how to solve particular problems" (OECD, 2009, p. 32). The descriptive text type favours native students vis à vis immigrant students whereas the later perform better in narration. Again, perhaps school is exerting an influence because fictional texts predominate in textbooks.

For future analysis, it would be interesting to undertake a longitudinal analysis of all PISA surveys, from 2000 to 2011 to identify patterns for item format, aspect, text format, text type and situation. In order to do this we would have to obtain the undisclosed linking/anchor items from OECD and extend our analysis to other subjects - Mathematics and Science. This data availability would also allow us to do a country-specific longitudinal analysis with basis on the data pertaining to all used items in all assessments. In addition, subsequent analysis could focus on an in-depth study of other covariates possibly associated with these results. For example, by using data from the Students' Questionnaire related to reading habits and attitudes.

We need to keep in mind that a test like PISA should avoid the inclusion of items that present DIF. Test construction should be such that one is not expected to find items with a large DIF. When DIF is found, wrong conclusions can be made about students' knowledge and abilities. On the other hand, it is impossible to completely eliminate DIF in cognitive tests, especially in large-scales tests such as PISA. The growing interest in DIF arose precisely from the desire to create test items that are not affected by the cultural and ethnic characteristics of the individuals taking the tests. Nevertheless, the presence of DIF in certain items may be very useful to study social and cultural differences that are not easily noticed. In educational assessment, DIF items can help us detect content that is treated differently among groups and can help us interpret why this may be the case. More importantly, it can help us decide what we could change and/or which equity measures could be implemented in schools to ensure educational achievement for all students. In this sense, the challenge in explaining DIF is to find patterns in the items that present DIF.

In this report, we provide a descriptive look at the different patterns in items that presented DIF in PISA 2009 reading items. Previous studies on DIF have mainly focused on gender differences and on identifying differences in performance related to curriculum content. One limitation inherent to the sampled immigrant population in some of the countries included in this analysis is that its proportion relative to native students is very low. Immigrant populations can differ substantially, even in countries where there are more than 3% of immigrant students, due to the ethnic and socio-

economic characteristics of the country of origin and level of integration in the host country, among other factors. Nonetheless, this analysis identifies differences that can be further explored at the country level and with national assessments of educational progress.

References

- Baker, F.B. & Kim, S.H. (2004). Item Response Theory - Parameter Estimation Techniques - 2nd ed. New York: Marcel Dekker Inc.
- Gamerman, S.; Soares, T. & Goncalves, F. (2010). Bayesian analysis in item response theory applied to a large-scale educational assessment. In Anthony O'Hagan & Mike West. Oxford Handbook of Applied Bayesian Analysis. Oxford: University Press, 624-652.
- Hambleton, R.K.; Swaminathan, H. & Rogers, H.J. (1991). Fundamentals of Item Response Theory. North Caroline: Sage Publications.
- Holland, P. W. & Wainer, H. (1993). Differential item functioning. New Jersey: Lawrence Erlbaum.
- Holland P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braum (Orgs.). Test validity. Hillsdale, NJ: Lawrence Erlbaum.
- Kim, S. & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. Applied Psychological Measurement, 22, 345-355.
- Liebig, T. and S. Widmaier (2009), "Children of Immigrants in the Labour Markets of EU and OECD Countries: An Overview", *OECD Social, Employment and Migration Working Papers*, No. 97, OECD Publishing.

  doi: 10.1787/220823724345
- Mazor, K. H., Hambleton, R. K. & Clauser, B. E. (1998). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. Applied Psychological Measurement, 22, 357-367.
- Muraki, E. e Bock, R.D. (2003). PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks (Version 4.1) [Computer software]. Chicago: Scientific Software.
- OECD, 2010. PISA 2009 results: What students know and can do. Paris: OECD Publications.
- OECD, 2009. PISA 2009 Assessment Framework – Key Competences in Reading, Mathematics and Science. Paris: OECD Publications.
- OECD, 2006. Where immigrant students succeed - A comparative review of performance and engagement in PISA 2003. Paris: OECD Publications.
- OECD, 2007. Science competencies for tomorrow's world. Paris: OECD Publications.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27, 361-370.
- Thissen, D., Steinberg, L. & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Orgs.). Differential item functioning (pp. 67-114). New Jersey: Lawrence Erlbaum.
- Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago, IL: MESA Press.

European Commission
EUR 25565 EN– Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **Differential Item Functioning (DIF): What Functions Differently for Immigrant Students in PISA 2009 Reading Items?**

Authors: Patrícia Dinis da Costa and Luísa Araújo

Luxembourg: Publications Office of the European Union

2012 – 24 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-26789-5

doi:10.2788/60811

Abstract

This analysis focuses on identifying items that present Differential Item Functioning (DIF) between native and immigrant students in the PISA 2009 reading assessment. The findings indicate that thirty-five percent of the 95 items or questions included in the analysis present DIF. This is an important finding because it can help us detect content that is treated differently between different groups of students. In particular, we found that immigrant students perform better than native ones in about half of the items with DIF and that they tend to perform better in questions that relate to situations and types of reading typically done in school settings. Policy measures aimed at ensuring educational achievement for all students can be informed by studies of this kind.

LB-NA-25565-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

ISBN 978-92-79-26789-5

Publications Office