Education and Science

Vol 42 (2017) No 190 401-422

Evaluation of Reading Skills Subtest Results of Countries Participating in the PISA 2009 Application via Profile Analysis *

Özkan Saatçioğlu [1], H. Deniz Gülleroğlu [2]

Abstract

The purpose of this research is to determine the characteristics possibly went unnoticed over general average scores in reading skills subtest of students in countries attending the Programme for International Student Assessment, PISA in 2009, and to reveal their strengths and weaknesses within the framework of item categories in question. The sample of this research constitutes of students in 15-years-old group participating in the PISA 2009 application. In the study, performances of countries participating in the PISA 2009 application within the framework of item format, text format and cognitive process level of the text in reading skills subtest have been analyzed using the profile analysis technique and intercountry comparisons are made. Obtained remarkable findings are discussed in accordance with the related literature. Within the scope of the research, it has been concluded that evaluating countries participating in the PISA application as per their average scores results in ruling out important country-specific information at item categories level, and that it is possible to employ profile analysis, as a new approach, with a view to reveal such information.

Keywords

PISA 2009
Reading skills
Rasch Model
Profile analysis
Intercountry comparison

Article Info

Received: 01.19.2016
Accepted: 03.20.2017
Online Published: 04.14.2017

DOI: 10.15390/EB.2017.6238

Introduction

Starting from the beginning of the twentieth century, psychometrists have developed various theories of measurement with a view to make more correct decisions by means of conducting more valid and reliable measurements and to explain implicit characteristics that underlie human behaviors in the most appropriate way (Crocker & Algina, 2008; Hambleton & Swaminathan, 1985; Lord, 1980). With these developed measurement theories, it is aimed to conduct measurement and evaluation applications more effectively, in this respect, studies have been conducted continuously on new models or approaches (Baker, 2001; De Gruijter & Van der Kamp, 2008; Verhelst, 2014).

Traditionally, Item Response Theory (IRT) models are established on the assumption of unidimensionality besides multidimensional model development studies (Hambleton & Swaminathan, 1985). The unidimensionality assumption which means that each item measures a single dimension,

* This study was produced from the master thesis of Özkan Saatçioğlu at the Institute of Educational Sciences at Ankara University.
* This study was orally presented at the II. International Eurasian Educational Research Congress (8-10 June 2015).

[1] Ankara University, Faculty of Educational Sciences, Dept. of Measurement and Evaluation, Turkey, saatcioglu@ankara.edu.tr
[2] Ankara University, Faculty of Educational Sciences, Dept. of Measurement and Evaluation, Turkey, denizgulleroglu@yahoo.com

indicates that all items effecting the test score measure the same latent trait or test items gather around a single latent trait and an individual's performance in the test is effected only that latent trait (Crocker & Algina, 2008; DeMars, 2010). Unidimensional IRT models require the unidimensionality assumption (Hambleton, Swaminathan, & Rogers, 1991). Many factors influencing the individual's response such as two or more abilities required to correctly answer items in the test, the problem of defining the psychological structure to be measured, cognitive skill, excitement, anxiety, motivation level, test response speed make it difficult to fully meet unidimensionality assumption in measurement and evaluation applications (De Ayala, 2009; Hambleton & Swaminathan, 1985; Osterlind, 2002). Thus, in recent years, it has become more debatable to make decisions about individuals taking the test by means of conducting measurements under unidimensionality assumption especially in achievement and ability tests (Embretson & Reise, 2000; Köse, 2010; Özer Özkan, 2012; Reckase, 2009; Walker & Beretvas, 2003).

The Rasch model is used which is a special version of one parameter logistic model among unidimensional IRT models in also PISA applications conducted by Organisation for Economic Co-operation and Development (OECD). It is considered as a limitation to use a unidimensional IRT model in PISA applications in which each item of the test has different characteristics in terms of item format, text format and the cognitive process level of the text (OECD, 2009, 2012). Therefore, it is stated that using unidimensional IRT model in test consist of items affected by two or more properties is not appropriate because of causing wrong decisions (Verhelst, 2001; Walker & Beretvas, 2003). Although unidimensional IRT models have difficulty in dealing with multidimensional data, this situation does not mean a criticism towards PISA practices, in contrast, it is stated that the Rasch model provides useful information even in cases where the unidimensionality assumption is met at a reasonable level (Köse, 2010; Özer Özkan, 2012; Verhelst, 2012; Yıldırım & Yıldırım, 2012). As a solution to the limitation brought up by the use of Rasch model, a unidimensional IRT model, in PISA applications (Yıldırım, Yıldırım, & Verhelst, 2014), either multidimensional IRT models (Reckase, 2009) or additional analyses in order to calculate deviations from unidimensionality are suggested (Verhelst, 2012). Along with the development of technology, computer software that can be used for multidimensional IRT models have been developed. However difficulty in interpreting obtained results or the fact that the software that show individual's possibility of correctly answering an item in multidimensional IRT are under development make it harder to use these models (Köse, 2010; Osterlind, 2002; Wiberg, 2012).

In the PISA application, it is possible to estimate difficulty parameters as an average value among answers of students from participant countries using the Rasch model, and performances of countries are listed in a proficiency scale based on students' total test score averages. As a result of estimating item difficulty parameters as an average value from answers by all the students participating in application, individual or group based differences may not be reflected on item parameters. In these applications, using item difficulty parameters estimated by combining all student responses results in missing strengths and weaknesses of countries and some features specific to countries within the general average (Verhelst, 2012). Verhelst (2012) recommends profile analysis as a complementary approach to PISA applications with a view to reveal group-based or individual systematic differences and some information possible for missing by Rasch model. With this technique, it becomes possible to determine individual or group-based differences that cannot be revealed by Rasch model, thus relative strengths and weaknesses of the relevant group can be detected (Yıldırım et al., 2014).

Beyond the average performances, determining countries' strengths and weaknesses in item categories level is important in PISA studies in order to reveal unique resemblances and differences among them (Verhelst, 2012; Yıldırım et al., 2014). For instance; questions containing visuals can be easy for students in a country which textbooks' mostly composed of visual elements in addition to descriptive texts and have central exams' include such questions, in comparison to the students who have never experienced this before (Yağmur, 2009; Yazıcı, 2006). However, because of the item

parameters of such questions will be predicted on the average of the data gathered from whole population, differences between students will not reflect on item parameters, so countries' strong and weak item categories cannot be determined (Yıldırım & Yıldırım, 2012). As a result, profile analysis can reveal some country-specific characteristics that are likely to be overlooked on average scores. It is thought that, this would enable countries to identify systematic differences in education and testing systems, and to develop educational reforms for this information.

During the process of analysis of test results in the PISA 2009 application, since item difficulty parameter is estimated by combining nearly 516 thousand students' responds from all 73 countries participating the application, and comparisons are made over total scores obtained from tests, it is thought that systematical differences among countries and some important country-specific features in the level of item categories may be overlooked. For this reason, the research problem of this study is to determine specific features of countries that are likely to be missed over the total score of the countries participating in the application regarding item categories discussed in defined dimensions within the framework of the evaluation of PISA 2009 reading skills subtest which is a major field.

Turkey has consistently participated in PISA applications and showed performance below the average of OECD since 2003 (OECD, 2009, 2012). A vast number of research have been done in order to determine factors affecting students' academic successes at mathematics, science and reading skills either in Turkey or in other participant countries (Anıl, 2009; Demir & Kılıç, 2010; Sun, Bradley, & Akers, 2012; Şengül, 2011; Yıldırım, 2012), make comparison between some countries in terms of several variables (Aydın, Erdağ, & Taş, 2011; Brozo, Shiel, & Topping, 2007; Cromley, 2009; Guo, 2014; İş Güzel, 2006), deliver validity of applications with making item bias studies (Ayan, 2011; Yıldırım & Berberoğlu, 2009; Le, 2009; Liu & Wilson, 2009; Lyons-Thomas, Sandilands, & Ercikan, 2014) and identify similarities and differences between countries (Acar, 2012; Kjærnsli & Lie, 2004; Soh, 2014; Zhang, Khan, & Tahirsylaj, 2015). On the occasion of PISA applications' framework of evaluation was developed to assess high level thinking skills of 21st Century; like problem solving, creativity, critical thinking, evaluation and effective communication skills, obtained results can be used to determine countries' future economic power and development level (Ministry of National Education [MNE], 2010, 2012; National Education Association [NEA], 2012; OECD, 2009, 2012). For this reason, deficiencies in the national educational programmes can be made up and important reforms can be put into practice with reference to the PISA results (Beaton, Postlethwaite, Ross, Spearritt, & Wolf, 1999; Schleicher, 2007).

Despite the fact that, reasons of Turkey's general failure in PISA applications were mentioned in many research in relevant literature (Acar & Öğretmen, 2012; Aydın, Sarıer, & Uysal, 2012; Balım, Evrekli, İnel, & Deniş, 2009), it can be seen that studies considering item categories are limited in number (Demir, 2010). In addition, in several studies; educational programmes in Turkey, questions in textbooks, activities and central exam systems are examined according to the international exams framework of evaluation (Aşıcı, Baysal, & Şahenk Erkan, 2012; Aydoğdu İskenderoğlu & Baki, 2011; Aydoğdu İskenderoğlu, Erkan, & Serbest, 2013; Coşkun, 2013; Delil & Yolcu Tetik, 2015; Güner, 2015; Savran, 2004). When these results from the studies considered, it is thought that contents and psychometric properties of items used in PISA applications, may affect students' performances too. In this way, analyzing items with classifying by common properties is seen important for determining Turkish students' performances in item categories level.

Because of the fact that high level of thinking skills gaining more importance day by day (OECD, 2009), it becomes a necessity to effectively evaluate PISA applications' results with secondary analyses. Consequently; within the scope of this research, it is aimed to make comparisons between participant countries by determining strengths and weaknesses regarding item format in reading skills that is one of the most influential factor on mathematics and science literacy (Bayat, Şekercioğlu, & Bakır, 2014; Cromley, 2009; Göktaş & Gürbüztürk, 2012), text format and item categories held within the dimensions of cognitive process level dimensions.

Pursuant to the main aim of the research, answers were sought for the questions below.

Is there any significant difference between expected and observed profiles of participant countries' in terms of PISA 2009 reading skills subtest's categories within the dimensions of;

1. Item format,
2. Text format,
3. Cognitive process level of the text.

Method

Data Source

475,460 students representing approximately 26 million students from 65 countries participated in the PISA 2009 application. Then a second application was conducted and 40,498 students representing about 2 million students from 8 countries that are not members of the OECD were added to the research, and at the end of these two applications, participation by a total of 515,958 was ensured. Students with needs of special education are from various countries included in PISA 2009 application, and unlike other students, special booklets (Une Heure booklet-UH-) were developed for these students. Thus, 1,091 students in need of special education were included in these two applications, but these students' responses were not included in the item and ability parameter estimation processes (OECD, 2009, 2012). In this study, the population consists of all 15-year-old students attending formal education in 73 countries in 2009. The sample comprises 514,867 students of 15-year-old group participating in PISA 2009 application. Number of students by countries are given in Table 1.

Table 1. Countries and Number of Students Participating in PISA 2009 Application

Country	Number of Students	UH*	Country	Number of Students	UH*	Country	Number of Students	UH*
Albania	4596		India	4826		Peru	5985	
Arab Emir.	10867		Indonesia	5136		Poland	4917	
Argentina	4774		Ireland	3937		Portugal	6298	
Australia	14251		Israel	5761		Qatar	9078	
Austria	6590	110	Italy	30905		Romania	4776	
Azerbaijan	4691		Japan	6088		Russia	5308	
Belgium	8501	291	Jordan	6486		Serbia	5523	
Bulgaria	4507		Kazakhstan	5412		Shanghai	5115	
Brazil	20127		Kyrgyzstan	4986		Singapore	5283	
Canada	23207		Korea	4989		Slovakia	4555	30
Chile	5669		Latvia	4502		Slovenia	6155	162
Colombia	7921		Liechtenstein	329		Spain	25887	
Costa Rica	4578		Lithuania	4528		Sweden	4567	
Croatia	4994		Luxembourg	4622		Switzerland	11812	
Czech Rep.	6064	222	Macao Chi.	5952		Taipei Chi.	5831	
Denmark	5924		Malaysia	4999		Thailand	6225	
Estonia	4727		Malta	3453		Tobago	4778	
Finland	5810		Mauritius	4654		Tunisia	4955	
France	4298		Mexico	38250		Turkey	4996	
Georgia	4646		Moldova	5194		Unit. States	5233	
Germany	4979	179	Montenegro	4825		Unit. Kingdom	12179	
Greece	4969		Netherlands	4760	97	Uruguay	5957	
Hong Kong	4837		New Zeal.	4643		Venezuela	2901	
Hungary	4605		Norway	4660				
Iceland	3646		Panama	3969		TOTAL	515958	1091

* Number of students need special education (OECD, 2012)

As can be seen in Table 1, a total of 515,958 students seem to be participated in PISA 2009 application, 4996 of whom were from Turkey and 1091 of whom were in need of special education. Item and skill parameter estimations were conducted using data from 514,867 students by subtracting number of UH from total number of participants.

Data Retrieval

In this study, data obtained from reading skills subtest from countries participating in PISA 2009 application, item code numbers and item difficulty parameters published in the international technical report were used. Application data have been obtained through the internet from OECD official web site (www.pisa.oecd.org). In the report released by the OECD, information about item code numbers for text type and text structure were not given for PISA 2009 reading skills subtest. Although it is considered to conduct analysis for all dimensions determined within the framework of evaluation of reading skills, because necessary information is not provided in OECD technical report, analyses have been conducted only under three dimensions including item format, text format and text cognitive process level of the text. This situation constitutes the limitation of the research.

Data Analysis

Within the scope of the research, arrangements required for analysis of the data obtained from international database of PISA 2009 application have been made. As a result of this arrangement, a new database was established, and profile analysis technique developed by Norman Verhelst (2012) was applied on the prepared database as a secondary data analysis. In the process of data analysis, SPSS 17.0, MS Office Excel 2003 and PROFILEG software were used.

Profile analysis has been developed to support measurement models such as Rasch model or one-parameter logistic model in which total score concept is valid. With this analysis, items in a test can be divided into dimensions predetermined as per common traits and categories under these dimensions, and their profiles can be constructed on individual or group level from the point of item categories. Thus, it is indicated that this technique makes it possible to reveal systematic differences among countries which is difficult for the Rasch model to determine for item categories with determined dimensions (Verhelst, 2012; Yıldırım & Yıldırım, 2012).

Although it is possible to carry out profile analysis on each student level, since PISA project focuses on performance of countries participating in the application rather than that of students, the values that are calculated on individual basis can be combined in group level (for instance, students in the same city, region or country). Thus, it is possible to approach countries participating in PISA as a group, compare students' performances in determined item categories with the average performance of all students, and calculate students' differences from a certain country from average values (Verhelst, 2012; Yıldırım & Yıldırım, 2012).

Profile analysis technique is based on the calculation of probabilities towards how many items with certain commonality a student who has obtained certain score from the test can answer correctly by means of used item parameters acquired from the measurement model on condition that the used measurement model complies with the test data at an acceptable level. Afterwards, this expectation value –which is calculated based on item parameters obtained from the measuring model is compared with observed performance of students. For example, given that each item answered correctly is evaluated as one point in a 15-question test that consists of eight items with continuous text feature and seven items with non-continuous text feature. Supposing that the Rasch model complies with this test data at an acceptable level, and as a result of analyses, item difficulty parameters are estimated. With the profile analysis and using these difficulty parameters, it is possible to calculate the expectation value stating how many items a student obtained six points from the test may have answered correctly. Then,

the observed score of a student obtained six points from the test can be compared with the expectation values calculated based on difficulty parameters estimated using the Rasch model. Thus, it is possible to understand whether the relevant item group is easier or more difficult for students than estimated by the measurement model, and to determine in which item group, students' performance is higher or lower than expected (Verhelst, 2012).

It is necessary to have two or more participant groups and items in the test should consist of two or more categories within the framework of a certain dimension in order to conduct a profile analysis. A single student's score in each category is defined as observed score, and the observed profile is created by combining students' observed scores for all categories. The sum of scores in student's observed profile gives the total score obtained from the test. Then, for each category taken into consideration, students' expected scores are calculated using item parameters estimated with the Rasch model, and these expected scores are combined into the expected profile. Afterwards, deviation score is calculated by subtracting students' expected scores from their observed scores for each determined category. Calculated deviation scores, on the other hand, are combined for all categories, and deviation profile is acquired. Thus, the deviation profile is expressed as the difference between student's observed performance and his/her expected performance under measurement model. Deviation profiles can also be calculated at countries level by gathering students together. Finally, an index is being established by calculating excess percentages of countries in determined categories with a view to compare countries participating in PISA application. The percentages of the students who have demonstrated lower or high performances than expected by the measurement model can be calculated by these excess percentages (Verhelst, 2012; Yıldırım et al., 2014). In Equality 1, formula of excess percentage is given.

$$\text{Excess Percentage} = \frac{\text{Observed Profile} - \text{Expected Profile}}{\text{Total Number of People}} \times 100 \quad (1)$$

A disparity index is calculated based on Chi-square statistics with a view to determine whether there is a significant deviation as per the amount of deviation profile value calculated by subtracting the observed profile from the expected profile after calculating excess percentages for item categories for which each participant country shows lower or higher performance than anticipated by the measurement model (Büyüköztürk, Çokluk, & Köklü, 2011). In Equality 2, Chi-square formula is given.

$$D^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

With a view to determine whether calculated Chi-square value is meaningful in determined α level, obtained value is compared with the critical value in Chi-square distribution table (Baykul & Güzeller, 2013).

Profile Analysis in PISA 2009 Reading Skills Item Categories

Each student in the PISA 2009 application has received randomly one of 20 booklets that contain a total of four item sets consisting of reading skills, mathematics and science literacy fields. In this application, nine-item set for reading skills as the weighted area and three-item sets each for science and mathematics fields were used (OECD, 2012). Reading skills item sets in booklets are chosen from among 131 items used in PISA 2009 application.

The number of items in booklets changes so, the students who receive the different booklets give answers to the different items due to the use of incomplete block design in PISA applications. During the data analysis stage, answers given to items in booklets are coded with various numbers in

the database. When items at the end of the booklet are left empty, the code eight is given thinking that student could not reach these items. In addition, following the implementation, items that could not be answered by students because of literal error or those canceled by the national center are coded as seven. In accordance with these information, the database obtained from OECD was revised according to the profile analysis, and a secondary database was established. To ensure the requirements of the profile analysis within the scope of this research, students with more than five codes in seven and eight, those answering all items correctly or not answering any items correctly were excluded from the analysis, and a total of 461,250 students were included in profile analysis. For the PISA 2009 application, number of students included in profile analysis from each participant country and international acronyms for country names are given in Table 2.

Table 2. Number of Students Involved in Profile Analysis for PISA 2009 Application

Country	Acronym	Total	Country	Acronym	Total
Albania	ALB	4004	Lithuania	LTU	4259
Arab Emir.	ARE	9780	Luxembourg	LUX	4238
Argentina	ARG	3331	Macao China	MAC	5486
Australia	AUS	13142	Malaysia	MYS	4534
Austria	AUT	6125	Malta	MLT	3092
Azerbaijan	AZE	4235	Mauritius	MUS	4150
Belgium	BEL	7560	Mexico	MEX	32780
Bulgaria	BGR	3971	Moldova	MDA	4495
Brazil	BRA	17358	Montenegro	MNE	3996
Canada	CAN	21517	Netherlands	NLD	4469
Chile	CHL	4970	New Zealand	NZL	4244
Colombia	COL	6247	Norway	NOR	4338
Costa Rica	CRI	3957	Panama	PAN	3214
Croatia	HRV	4801	Peru	PER	4417
Czech Rep.	CZE	5540	Poland	POL	4589
Denmark	DNK	5556	Portugal	PRT	5783
Estonia	EST	4493	Qatar	QAT	8009
Finland	FIN	5492	Romania	ROU	4512
France	FRA	3834	Russia	RUS	4662
Georgia	GEO	3723	Serbia	SRB	5106
Germany	DEU	4498	Shanghai China	QCN	4840
Greece	GRC	4496	Singapore	SGP	4828
Hong Kong China	HKG	4533	Slovakia	SVK	4328
Hungary	HUN	4347	Slovenia	SVN	5764
Iceland	ISL	3354	Spain	ESP	23857
India	QHP	3228	Sweden	SWE	4175
Indonesia	IDN	4394	Switzerland	CHE	11137
Ireland	IRL	3632	Taipei China	TAP	5574
Israel	ISR	5023	Thailand	THA	5851
Italy	ITA	28520	Tobago	TTO	3735
Japan	JPN	5635	Tunisia	TUN	4102
Jordan	JOR	5881	Turkey	TUR	4725
Kazakhstan	KAZ	4577	United States	USA	4960
Kyrgyzstan	KGZ	3318	United Kingdom	GBR	11477
Korea	KOR	4801	Uruguay	URY	4763
Latvia	LVA	4255	Venezuela	QVE	2325
Liechtenstein	LIE	308	**TOTAL**		**461250**

Item parameters in the PISA 2009 application are estimated by combining all participant students' answers and published in OECD technical report. Within the scope of this research, since profile analysis is conducted using item parameters published in the technical report, all countries participating in the PISA 2009 application can be included in the analysis. This situation allows comparison of all countries at the same time and constitute strong side of the analysis.

For the accuracy of profile analysis results, it is necessary to classify items in the test logically. Otherwise, findings obtained from analyses conducted with theoretically incorrect classifications may lead to misinterpretations (Yıldırım & Yıldırım, 2012). During the process of determining dimensions of item format, text format and the cognitive process level of the text dealt with in this research, the data given in PISA 2009 international technical report were taken into account. In Table 3, the dimensions, item categories and the number of items under these dimensions are given.

Table 3. Categories and the Number of Items Studied in Profile Analysis for PISA 2009 Application

Sizes	Categories	Number of Items
Item Format	Multiple Choice (MC)	62
	Open Ended *(OE)*	69
	Total	**131**
Text Format	Continuous Text *(CT)*	81
	Non-continuous Text *(NT)*	38
	Combined Text *(COT)*	12
	Total	**131**
Cognitive Process Level of the Text	Accessing and Retrieving Information *(ARI)*	31
	Integrating and Interpreting Information *(III)*	67
	Reflecting Their Own Thoughts and Evaluation *(RTOTE)*	33
	Total	**131**

Item Format. Multiple-choice items in PISA 2009 application, in which students are provided with options and that allow objective scoring are divided into two including; (1) *multiple choice* and (2) *complex multiple-choice* items. However, open-ended items in which students structure the answer and individual judiciary is required for scoring of responses are divided into three including (1) *open constructed response*, (2) *closed constructed response* and (3) *short response* items. Items in five categories as basis for item formats in PISA 2009 application can be collected in two categories including items for which students structure their responses (open-ended) and item with options given to students (multiple-choice) (MNE, 2010, 2012; Rodriguez, 2002). Besides in this research, since number of items for mixed multiple-choice, closed ended and short response item categories is lower in booklets, these items were combined into two categories including multiple-choice and open-ended items.

Text Format. The PISA 2009 application usually includes *continuous texts* in the form of one paragraph (articles, novels, short stories, summaries and Internet news); *non-continuous texts* consisting of different combination of lists, tables, graphs, maps, charts and shapes; *mixed texts* as single and compatible texts showing characteristics of continuous and non-continuous texts; *multiple texts* consisting of different texts without any certain relationships (MNE, 2010, 2012). Within the framework of evaluating the reading skills, although four different text formats have been defined including continuous, non-continuous, mixed and multiple texts, it can be seen that there are small number of items in mixed and multiple text categories. In accordance with information given in PISA 2009 technical report (OECD, 2012), since item features in these two categories are similar to each other, mixed and multiple text categories are discussed together by researchers considering related literature, and they are renamed as *combined text*.

Cognitive Process Level of the Text. The PISA 2009 application includes items categories at cognitive process levels such as *accessing and retrieving information*, that contains skills related to finding, differentiating and collecting information, *integrating and interpreting information* that requires

establishing cause-and-effect relationship between different parts of a text, conducting classification and sampling, realizing a close relationship, making inferences and understanding piece-whole relationships, *reflecting their own thoughts and evaluation* that requires individual to associate his/her own knowledge and experience with information in the text and to make judgment related to contents of the text. Although it is accepted that it is not possible for cognitive process level of the text to combine and interpret information without reaching information or reflect, evaluate and make comments about information without remembering it, items in PISA 2009 application have been developed to emphasize one of these cognitive process levels (MNE, 2012). Within the scope of this research, classifications related to cognitive process level of the text dimensions published in technical report by OECD (2012) are considered.

Findings

Observed profiles, expected profiles and excess percentages of countries participating in the application are calculated within the framework of multiple-choice and open-ended item categories for item format dimension, continuous text, non-continuous text and combined text categories for text format dimension, and accessing and retrieving information, integrating and interpreting information, reflecting their own thoughts and evaluation categories for the dimension of cognitive process level of the text in the PISA 2009 reading skills subtest. In Table 4, countries represented by their international acronym are listed as per calculated excess percentages, the countries with significant difference between observed and expected profiles at 0.01 level (deviation profile) are indicated in bold. Thus, within the framework of the objective set out for the research, item categories that countries' stronger and weaker aspects than measurement model estimated have been determined. In addition, prominent findings from some countries are given and intercountry comparisons have been made as shown in Figure 1, 2 and 3.

Table 4. Excess Percentages of Countries in Item Categories

Item Format			**Text Format**		**Cognitive Processes Level of the Text**		
MC	**OE**	**CT**	**NT**	**COT**	**ARI**	**III**	**RTOTE**
QAT (11.6)*	**HKG (14.6)***	**ALB (13.0)***	**SGP (9.3)***	**JPN (7.6)***	**HRV (8.4)***	**MNE(13.5)***	**TUN(14.7)***
SVN (10.8)*	**QCN (13.4)***	**JOR (12.9)***	**MLT (9.0)***	**KOR (7.0)***	**SVK (5.9)***	**AZE (10.9)***	**USA (8.7)***
MYS (10.6)*	**IRL (8.3)***	**GEO (11.8)***	**GBR (6.6)***	**SVN (5.8)***	**NLD (5.7)***	**GEO (9.9)***	**BRA (8.6)***
KAZ (9.7)*	**USA (7.4)***	**KGZ (10.8)***	**EST (5.3)***	**FRA (5.4)***	**MUS (4.8)***	**CZE (9.8)***	**CAN (7.0)***
KGZ (9.3)*	**EST (7.0)***	**QCN (10.5)***	**MUS (5.2)***	**GEO (5.2)***	**MAC (4.8)***	**KGZ (9.0)***	**TUR (5.5)***
ITA (7.3)*	**LTU (6.9)***	**PER (9.8)***	**NZL (5.1)***	**HKG (4.7)***	**DNK (4.7)***	**RUS (8.4)***	**ISR (5.1)***
QHP (7.1)*	**JPN (6.8)***	**CRI (9.4)***	LIE (5.0)	**ISL (4.7)***	**SWE (4.5)***	**ALB (8.3)***	**GRC (4.9)***
AZE (7.0)*	**LVA (6.4)***	**TUN (9.3)***	**AUS (4.9)***	**MAC (3.6)***	LIE (4.3)	**SVN (7.5)***	**URY (4.4)***
MNE (6.8)*	**CAN (6.1)***	**IDN (8.7)***	**NLD (3.6)***	**FIN (3.3)***	**BEL (4.1)***	**KAZ (7.2)***	**IDN (4.3)***
GEO (6.4)*	**QVE (5.8)***	**QAT (8.1)***	**BEL (3.5)***	**DNK (3.1)***	**THA (4.0)***	**BGR (5.9)***	**MLT (4.2)***
RUS (5.8)*	**TUR (5.4)***	**PAN (7.4)***	**CAN (2.7)***	LTU (1.8)	**MEX (4.0)***	**JOR (5.6)***	**MEX (4.1)***
HUN (5.2)*	**NZL (5.1)***	**MNE (7.1)***	**CHE (2.6)***	DEU (1.6)	**NOR (3.9)***	**QAT (4.8)***	**PRT (3.8)***
PER (5.0)*	**ROU (5.1)***	**KAZ (7.1)***	USA (2.5)	GRC (1.5)	**HUN (3.5)***	**ISL (4.6)***	**LVA (3.6)***
FRA (5.0)*	**TUN (4.7)***	**THA (5.6)***	FRA (2.5)	NLD (1.4)	**SRB (3.5)***	**MYS (4.4)***	**GBR (3.5)***
CZE (5.0)*	**GBR (4.6)***	**ISR (5.4)***	TAP (2.0)	HUN (1.1)	**RUS (3.2)***	**DEU (4.4)***	**QAT (3.4)***
MUS (5.0)*	**IDN (4.4)***	**ARE (5.4)***	AUT (1.3)	ITA (1.0)	**AUT (2.7)***	**ITA (4.1)***	**AUS (3.3)***
DNK (4.9)*	**SGP (4.2)***	**BRA (5.2)***	KOR (1.1)	CHE (0.6)	JPN (2.3)	**SRB (4.1)***	**HKG (3.2)***
JOR (4.4)*	**MDA (4.2)***	**ARG (5.1)***	MDA (0.4)	LUX (0.3)	NZL (2.3)	**POL (3.6)***	**NZL (3.2)***
DEU (4.3)*	**BRA (4.0)***	**BGR (5.0)***	SWE (0.4)	AZE (0.3)	SGP (2.2)	**TAP (3.4)***	**COL (3.2)***
LIE (4.3)	**COL (3.5)***	**URY (5.0)***	ROU (0.3)	NZL (0.02)	KAZ (2.1)	**FRA (3.4)***	IRL (3.0)
AUT (4.0)*	**TAP (3.5)***	**CHL (4.9)***	DEU (0.2)	AUS (-0.2)	**CHE (2.0)***	**CHL (3.3)***	**ARE (2.8)***
TTO (3.8)*	**HRV (3.5)***	**GRC (4.7)***	LUX (0.1)	CZE (-0.2)	CRI (2.0)	**SVK (3.3)***	PAN (2.7)
ISR (3.6)*	**AUS (2.9)***	**ESP (4.6)***	IRL (0.05)	AUT (-0.4)	SVN (1.5)	QHP (3.2)	QVE (2.6)
SVK (3.4)*	**MAC (2.9)***	**PRT (4.6)***	THA (0.001)	BEL (-0.6)	DEU (1.5)	LUX (2.6)	QCN (1.6)
CHL (2.8)*	**PRT (2.9)***	**ITA (4.6)***	ISL (-0.02)	PAN (-0.7)	HKG (1.2)	TTO (2.3)	JOR (1.6)
ARE (2.2)*	**POL (2.7)***	**SRB (4.4)***	FIN (-0.07)	SVK (-0.8)	LTU (1.1)	HUN (2.2)	SWE (1.3)
CHE (1.7)*	**BEL (2.2)***	**RUS (4.3)***	TTO (-0.2)	LIE (-1.0)	GBR (1.1)	PER (2.2)	NLD (1.2)
NOR (1.3)	**CRI (2.1)***	**COL (4.2)***	LVA (-0.3)	IRL (-1.1)	ISL (0.9)	QCN (2.2)	CRI (0.9)

Table 4. Continued

Item Format			Text Format		Cognitive Processes Level of the Text		
MC	**OE**	**CT**	**NT**	**COT**	**ARI**	**III**	**RTOTE**
KOR (1.2)	**MEX (1.6)***	**SVK (4.1)***	MYS (-1.0)	TUR (-1.2)	TUR (0.7)	LTU (2.2)	KOR (0.8)
MLT (0.8)	SWE (1.6)	**HUN (3.9)***	IDN (-1.1)	EST (-1.4)	EST (0.7)	MDA (2.0)	SGP (0.6)
BGR (0.7)	ARG (1.3)	**TTO (3.9)***	QVE (-1.4)	QHP (-1.4)	MDA (0.1)	CHE (1.6)	ARG (0.6)
LUX (0.7)	ISL (1.1)	**NOR (3.8)***	JPN (-1.6)	HRV (-1.6)	FIN (0.1)	LVA (1.4)	ROU (0.5)
FIN (0.7)	GRC (1.1)	**HKG (3.5)***	ARE (-1.6)	**CAN (-1.7)***	MNE (0.03)	FIN (1.2)	ESP (0.5)
ALB (0.6)	THA (0.8)	QVE (3.5)	**MEX (-1.7)***	LVA (-1.8)	AZE (0.03)	AUT (1.2)	EST (0.5)
ESP (0.4)	URY (0.7)	**TUR (3.4)***	CZE (-1.7)	NOR (-1.8)	KOR (-0.01)	PAN (0.9)	THA (0.4)
PAN (0.3)	SRB (0.6)	HRV(2.5)	SRB (-1.8)	URY (-1.9)	MYS (-0.1)	EST (0.7)	CHL (0.05)
NLD (0.3)	NLD (-0.3)	AZE (2.5)	PRT (-1.8)	KAZ (-1.9)	TTO (-0.1)	ESP (0.6)	JPN (-0.4)
SRB (-0.6)	PAN (-0.3)	QHP (2.4)	DNK (-1.8)	POL (-2.1)	URY (-0.1)	ROU (0.4)	NOR (-0.7)
URY (-0.7)	ESP (-0.4)	**MEX (2.4)***	POL (-1.8)	MYS (-2.2)	CZE (-0.3)	MAC (0.2)	QHP (-1.1)
THA (-0.8)	ALB (-0.6)	MYS (2.3)	COL (-2.1)	ROU (-2.3)	AUS (-0.4)	MLT (-0.1)	FIN (-1.2)
GRC (-1.1)	FIN (-0.7)	LTU (2.0)	**BRA (-2.2)***	RUS (-2.3)	IRL (-0.5)	ARG (-0.2)	BEL (-1.4)
ISL (-1.1)	LUX (-0.7)	POL (1.9)	**SVN (-2.6)***	**GBR (-2.4)***	LUX (-0.5)	JPN (-0.2)	FRA (-1.9)
ARG (-1.3)	BGR (-0.7)	TAP (1.8)	**HRV (-2.8)***	CHL (-2.5)	TAP (-0.6)	QVE (-0.2)	POL (-2.3)
SWE (-1.6)	MLT (-0.8)	MAC (1.3)	QHP (-2.8)	JOR (-2.5)	BGR (-1.0)	GRC (-0.3)	PER (-2.4)
MEX (-1.6)*	KOR (-1.2)	IRL (0.8)	**MAC (-2.9)***	SWE (-2.5)	PER (-1.2)	ISR (-0.7)	**ITA (-2.9)***
CRI (-2.1)*	NOR (-1.3)	LVA (0.7)	**AZE (-3.1)***	**ESP (-2.5)***	**ESP (-1.3)***	DNK (-0.7)	TTO (-3.1)
BEL (-2.2)*	**CHE (-1.7)***	CZE (0.7)	**NOR (-3.3)***	SGP (-2.8)	ARG (-1.4)	ARE (-0.8)	**MDA(-**
POL (-2.7)*	**ARE (-2.2)***	MDA (0.5)	**TUR (-3.4)***	MDA(-2.9)	IDN (-1.8)	KOR (-1.1)	**TAP (-3.2)***
PRT (-2.9)*	**CHL (-2.8)***	USA (0.3)	**BGR (-3.5)***	**MEX (-2.9)***	**ITA (-1.9)***	NOR (-1.4)	**HRV (-**
MAC (-2.9)*	**SVK (-3.4)***	SWE (0.09)	**RUS (-4.1)***	QVE (-3.1)	PRT (-2.0)	BEL (-1.5)	**CHE (-3.5)***
AUS (-2.9)*	**ISR (-3.6)***	LUX (-0.1)	**SVK (-4.2)***	**QCN (-3.1)***	POL (-2.2)	MUS (-1.8)	**LUX (-3.5)***
HRV (-3.5)*	**TTO (-3.8)***	DNK (-0.2)	**ISR (-4.3)***	KGZ (-3.3)	**BRA (-2.2)***	IRL (-1.9)	**MUS (-**
TAP (-3.5)*	**AUT (-4.0)***	ROU (-0.4)	**LTU (-4.4)***	**MNE (-4.2)***	COL (-2.4)	LIE (-2.0)	**DNK (-**
COL (-3.5)*	LIE (-4.3)	**CAN (-1.4)***	**URY (-4.5)***	**ARG (-4.3)***	ROU (-2.7)	COL (-2.2)	**AUT (-4.0)***
BRA (-4.0)*	**DEU (-4.3)***	AUT (-1.7)	**ESP (-4.5)***	**COL (-4.3)***	FRA (-2.8)	CRI (-2.3)	**LTU (-4.1)***
MDA (-4.2)*	**JOR (-4.4)***	MUS (-1.8)	**CHL (-4.6)***	**ISR (-4.3)***	**ARE (-3.0)***	**PRT (-2.6)***	**HUN (-**
SGP (-4.2)*	**DNK (-4.9)***	DEU (-2.0)	**ARG (-4.7)***	**CRI (-4.6)***	**CHL (-3.2)***	**SGP (-2.9)***	LIE (-4.3)
IDN (-4.4)*	**MUS (-5.0)***	SVN (-2.1)	**HUN (-5.2)***	**QAT (-5.0)***	QHP (-3.3)	**AUS (-2.9)***	**ALB (-4.5)***
GBR (-4.6)*	**CZE (-5.0)***	JPN (-2.3)	**HKG (-5.5)***	**PER (-5.2)***	**MLT (-3.8)***	**CAN (-3.0)***	**GEO (-4.7)***
TUN (-4.7)*	**FRA (-5.0)***	**FIN (-2.8)***	**MNE (-5.5)***	**USA (-5.3)***	**USA (-3.9)***	**HRV (-3.4)***	**ISL (-5.0)***
ROU (-5.1)*	**PER (-5.0)***	**CHE (-2.9)***	**ITA (-5.9)***	**BGR (-5.5)***	**CAN (-4.0)***	**SWE (-3.5)***	**DEU (-5.2)***
NZL (-5.1)*	**HUN (-5.2)***	**BEL (-3.0)***	**GRC (-5.9)***	**PRT (-5.6)***	**QVE (-4.7)***	**GBR (-3.5)***	**MYS (-5.4)***
TUR (-5.4)*	**RUS (-5.8)***	LIE (-3.8)	**KAZ (-6.1)***	**MLT (-5.7)***	**LVA (-4.8)***	**IDN (-3.6)***	**MAC(-5.5)***
QVE (-5.8)*	**GEO (-6.4)***	**NLD (-3.9)***	**QAT (-6.3)***	**TAP (-6.5)***	**KGZ (-4.9)***	**URY (-4.1)***	**BGR (-6.1)***
CAN (-6.1)*	**MNE (-6.8)***	**ISL (-4.0)***	**TUN (-6.5)***	**TTO (-7.6)***	**ISR (-5.1)***	**NZL (-4.6)***	**SRB (-7.1)***
LVA (-6.4)*	**AZE (-7.0)***	**EST (-4.3)***	**CRI (-7.5)***	**ALB (-8.2)***	**ALB (-5.2)***	**HKG (-4.7)***	**SVK (-7.3)***
JPN (-6.8)*	**QHP (-7.1)***	**AUS (-4.6)***	**PAN (-7.7)***	**THA (-8.5)***	**PAN (-5.3)***	**NLD (-4.9)***	**KGZ (-7.5)***
LTU (-6.9)*	**ITA (-7.3)***	**NZL (-4.7)***	**PER (-8.7)***	**ARE (-8.6)***	**QCN (-6.4)***	**THA (-5.6)***	**SVN (-9.2)***
EST (-7.0)*	**KGZ (-9.3)***	**KOR (-5.1)***	**QCN (-8.8)***	**MUS (-8.7)***	**GRC (-6.4)***	**USA (-5.8)***	**KAZ (-9.2)***
USA (-7.4)*	**KAZ (-9.7)***	**GBR (-5.3)***	**KGZ (-9.1)***	**SRB (-9.3)***	**TUN (-7.9)***	**BRA (-6.7)***	**CZE (-9.7)***
IRL (-8.3)*	**MYS (-10.6)***	**FRA (-6.4)***	**ALB (-9.8)***	**TUN (-9.5)***	**JOR (-8.3)***	**TUR (-6.9)***	**RUS(-**
QCN (-13.4)*	**SVN (-10.8)***	**MLT (-6.7)***	**JOR (-11.8)***	**BRA(-10.3)***	**GEO (-8.9)***	**MEX (-7.5)***	**AZE(-14.0)***
HKG (-14.6)*	**QAT (-11.6)***	**SGP (-6.9)***	**GEO(-13.8)***	**IDN (-14.0)***	**QAT(-11.0)***	**TUN (-8.3)***	**MNE(-**

*p < 0.01

Note: MC: Multiple Choice, OE: Open-ended, CT: Continuous Text, NT: Non-continuous Text, COT: Combined Text, ARI: Accessing and Retrieving Information, III: Integrating and Interpreting Information, RTOTE: Reflecting Their Own Thoughts and Evaluation

In Table 4, countries participating in the PISA 2009 application are listed from high to low in order as per their excess percentage calculated in categories of multiple choice, open-ended, continuous text, non-continuous text, combined text, accessing and retrieving information, integrating and interpreting information, reflecting their own thoughts and evaluation. Looking at this table, one cannot comment about general success of countries in PISA application, but it is possible to make comparisons among countries based on their strong or weak aspects from the point of item categories to be discussed.

For example, according to general results of reading skills subtest in PISA 2009 application given in Appendix 1, it can be seen that, although Qatar is ranked in the 68th place with an average of 372 points, students in Qatar has better performance in multiple-choice items than estimated by the measurement model. This situation shows that majority of items answered correctly by students participating in the application in Qatar have multiple choice characteristics in terms of item format.

In spite of the low performance of Qatar in reading skills subtest, answering the multiple choice items easier, presents strong part of the students of aforesaid country. On the other hand, in terms of item format, it can be said that Hong Kong and Shanghai-China have better performances than expected in open ended item categories. As well as Hong Kong and Shanghai-China have high general average scores in the reading skills subtest, as a result of operated profile analysis having a high performance in open ended items more than measurement model predicted and having lower performance in multiple choice questions than measurement model prediction propound that, these students performs better than expected in open ended items. In a similar manner, in terms of text format dimension; Albania performs better than expected and ranked as the first in continuous text, likewise Singapore in non-continuous text and Japan in combined text categories. As for cognitive level dimension, Croatia perform better than what measurement model predicted in accessing and retrieving knowledge category, while Montenegro; in integrating and interpreting knowledge and Tunisia; in reflect and evaluate categories. These obtained results put forth similarities and differences for consideration by showing characteristic features of countries.

Students participating to the application from Turkey, on the other hand, showed higher performance than estimated by the Rasch model in categories of open-ended, continuous text, accessing and retrieving information, reflecting their own thoughts and evaluation; showed lower performance than estimated in categories of multiple choice, non-continuous text, combined text and integrating and interpreting information. As a result of the Chi-square analyses, the difference between observed and expected profiles of students was not found significant in categories of combined text and accessing and retrieving information.

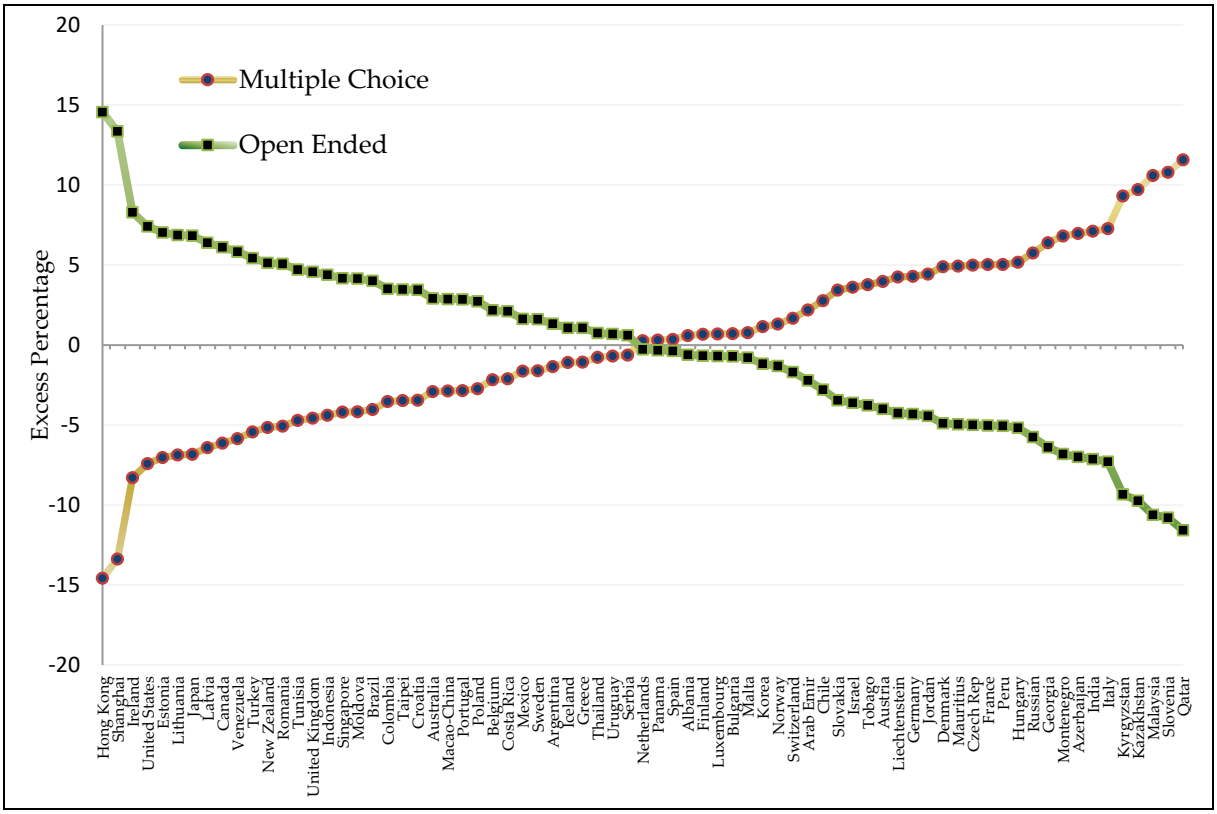

Figure 1. Countries Order with Respect to Their Excess Percentages in Item Format Dimension

As shown in Figure 1, Shanghai showed lower performance in multiple choice items than estimated by the measurement model, Kirghizstan showed significantly higher performance than estimated. In PISA 2009 reading skills subtest given in Appendix 1, Shanghai ranks in the first place and Kirghizstan ranks in the last place in the general ranking as per average scores of countries. This situation shows the fact that students participating the application from Kirghizstan respond to multiple-choice items correctly rather than open-ended items despite their overall failure.

Ranking in the 41st place with 464 points according to the general average reading skills subtest score, Turkey is ranked in the 11th place in open-ended item category and in the 63rd place in multiple choice item category, and presented a different performance from its general ranking. In this case; it can be said that there is a difference in favor of the open-ended item category between the expected profiles calculated by using the difficulty parameters of the items of the PISA 2009 reading skills subtest of the students participating in Turkey and the observed profiles obtained from the multiple-choice and open-ended item categories in the test. Accordingly, referring to the profile analysis results, it can be said that Turkish students performs better than what measurement model predicted in open ended items within reading skills subtest.

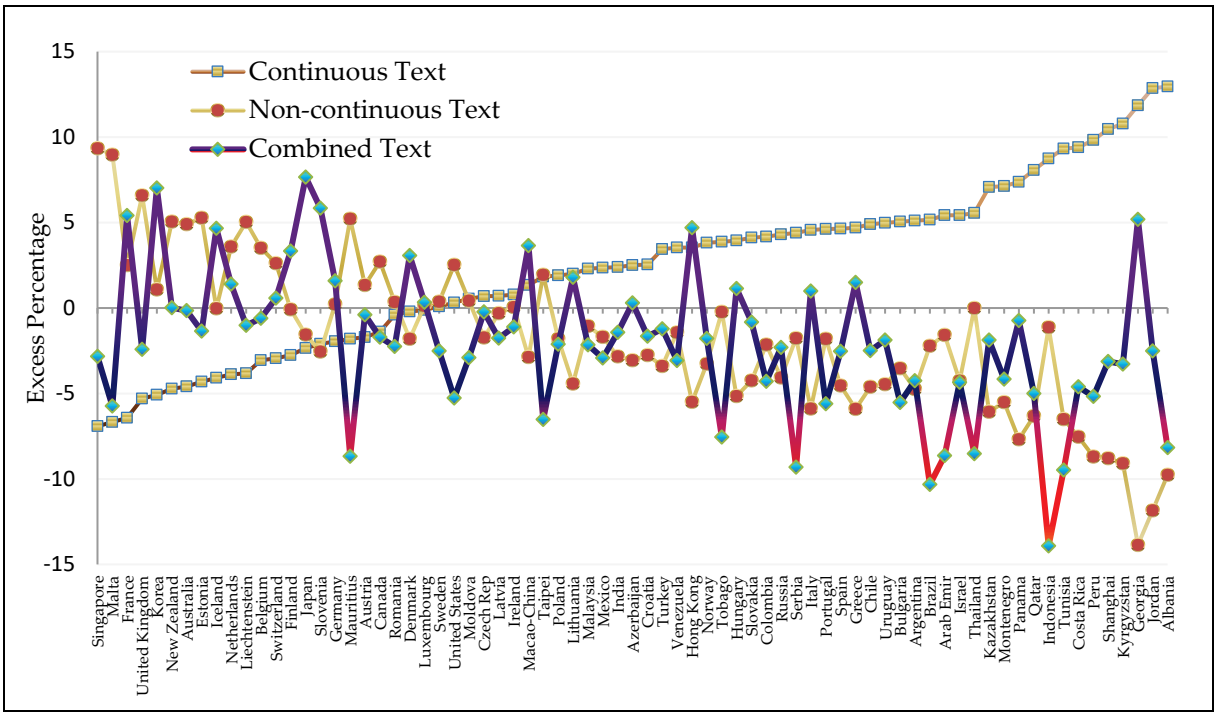

Figure 2. Countries Order with Respect to Their Excess Percentages in Text Format Dimension

In Figure 2, countries are listed increasingly according to their excess percentage in continuous text category of text format dimension. According to the general results of the reading skills subtest in PISA 2009 application given in Appendix 1, it can be seen that, although Albania is ranked in the 66th place with an average of 385 points, students in Albania show much better performance in items in the continuous text category than estimated by the measurement model. However, although Singapore shows better performance by ranking in the fifth place with average 526 points according to the general results, Singapore showed poorer performance in continuous text category than estimated by the model and is ranked in the last place among countries participating in the application. Singapore showed much higher performance in the non-continuous text category and is ranked in the first place. Thus, it can be clearly seen that students participating in the application in Singapore correctly answers items in non-continuous text category consisting of different lists, tables, shapes or graphs, on the other hand, they have difficulty in continuous texts consisting of paragraphs. In addition, analyzing the Figure 2, it can

be seen as a significant finding that many Asian countries are ranked near the top in items of non-continuous text and combined text types.

According to the general average score in PISA 2009 reading skills subtest, although Turkey is ranked in the 41st place, Turkish students have performed above expectations in continuous text categories but again Turkey could be only ranked in the 35th place among countries participating in the application. However, Turkish students had difficult in answering correctly the items in non-continuous and combined text categories, and they showed lower performance than expected.

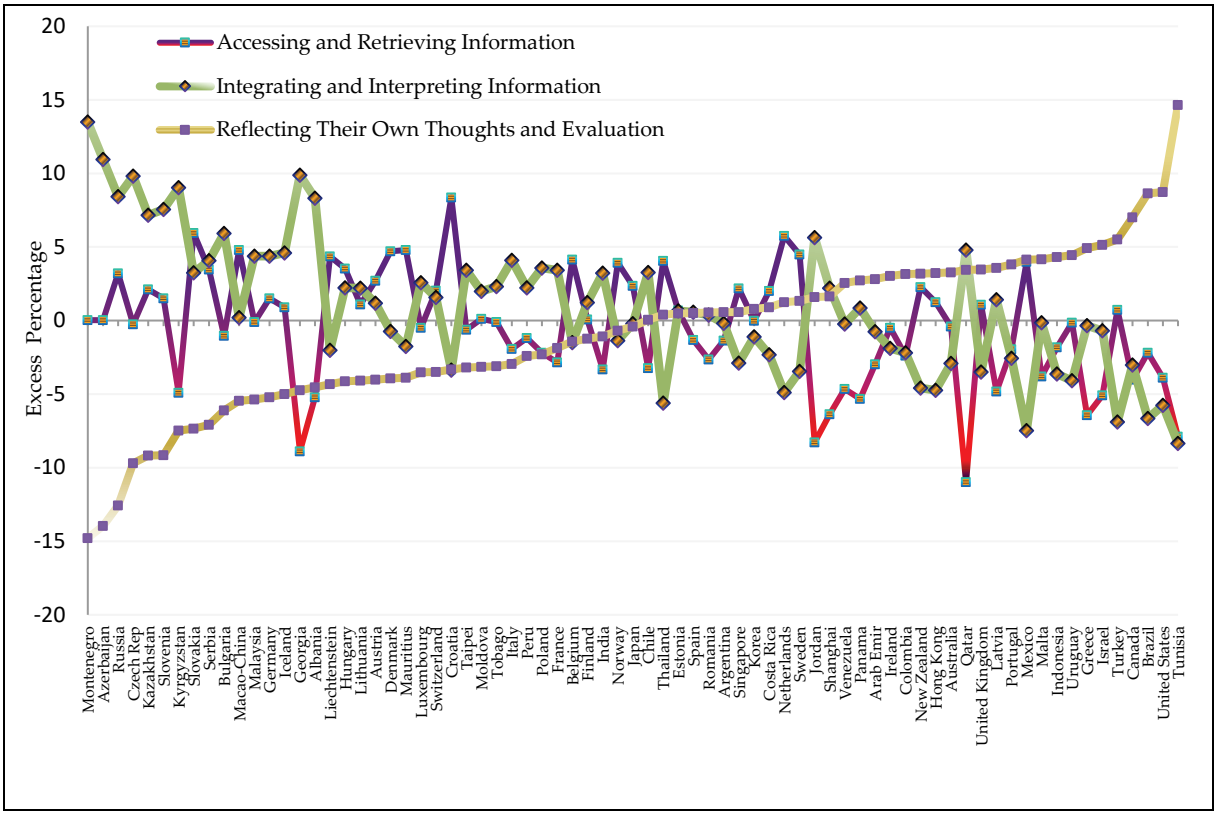

Figure 3. Countries Order with Respect to Their Excess Percentages in Mental Process Level of the Text

In Figure 3, countries are listed increasingly according to their excess percentages in reflecting their own thoughts and evaluation category in the dimension of cognitive process level of the text. According to the general results of reading skills subtest in PISA 2009 application given in Appendix 1, although Tunisia is ranked in the 61st place with an average 404 points and Montenegro is ranked in the 58th place with an average 408 points, Tunisian students in the items belonging to category of reflecting their own thoughts and evaluation and Montenegrin students in items belonging to category of integrating and interpreting information showed better performance than estimated by the measurement model and they were ranked in the first place in these fields. Turkey, on the other hand, was ranked in the 41st place according to general average score in PISA 2009 reading skills subtest. Another important finding is that Turkish students participating in the application are ranked in the fifth place by becoming more successful than expected for items in reflecting their own thoughts and evaluation. Despite general failure of the students participating in PISA 2009 application from Turkey, their higher performance in these items can relatively shows that the percentage of Turkish students with the ability to reasoning, relating and interpretation are higher than many other countries participated in the application. In the light of this information, with the aid of profile analysis technique, attention can be drawn to the deficiencies in the curriculas of the countries by determining item categories which participant countries have performances less than expected in as a weak side, and effective development for educational policy in line with global economic competition can be provided.

Discussion, Conclusion and Suggestions

PISA applications allow to determine students showing superior performance and reveal results suitable for interpreting human capital and competitive power of countries. Given that students showing superior performance will be able to contribute to the development of new information and technologies in various fields in the future (Education Reform Initiative [ERI], 2005, 2011), it can be suggested that, beyond comparing countries as per their general average scores, it is more important to determine performances of these countries at the level of item categories and determine their strong and weak points. PISA 2009 reading skills subtest results are evaluated with a new approach, the profile analysis technique, and at least some of the information possible for overlooking through average scores are revealed. In accordance with the obtained findings, it is considered that, by means of reanalyzing wide-scale test results developed with IRT models in which the concept of total test scores is valid using the profile analysis technique, more complete information can be provided to school administrators, teachers and parents regarding students' performances.

Strong and weak item categories of participant countries have been identified with using profile analysis in the PISA 2009 reading skills subtest item format, text format and cognitive process level of the text dimensions. According to the excess percentages obtained from country-level profile analyzes; the top 10 countries that performed above expectations in the multiple choice item category of the item format dimension, were below the OECD average ($\bar{X}$=493) according to the general average scores of the reading skills subtest. Among the top ten countries which performed above expectations in the open-ended item category; it is inferred that Lithuania, Turkey and Venezuela were below and the other seven countries were above the OECD average. With this finding, it can be said that countries whose strong side is multiple choice items in terms of item format are generally below competence level 3, while countries whose strong side is open ended items rank generally above the competence level 3. The fact that Turkey has shown better performance in open-ended items than measurement model expected, means that the correctly answered questions by the participating students in the sub-test of reading skills are generally in open-ended item category. Due to 18% of the open-ended items that students construct answers on their own in the test are closed-ended or short-response items, and mostly to be easy can be the reason for Turkish students' observed profiles are higher than the expected profiles in this category.

Similar to this research, in another study conducted by Verhelst (2012), strengths and weaknesses of 32 countries participating in the application were determined for item categories in dimensions including item format, text format and cognitive process level of the text for the PISA 2000 reading skills subtest. When findings from these two studies examined; it has been concluded that higher performance than estimated by the measurement model is achieved for open-ended category by United Kingdom, Mexico, Ireland, Latvia and Canada; for multiple-choice category by the Czech Republic, Switzerland, France, Italy, Germany and Hungary. The fact that the characteristic features of countries do not differentiate during the nine-year period can be explained by the invariance of the educational policies in these countries and the philosophy of education that originates in these policies (Klieme & Baumert, 2001; Şirin & Vatanartıran, 2014). In terms of item format, the Czech Republic and Slovakia have high and near excess percentages than expected in multiple choice items, which can be explained by the cultural similarities of these countries (Yıldırım et al., 2014).

People's Republic of China have participated as four different States including Hong Kong, Shanghai, Macao and Taiwan to the PISA 2009 application. According to the analysis results, it has been concluded that Chinese students correctly answer open-ended items, nevertheless, they have more difficulty than expected in multiple choice items. The reason for this situation is shown as the fact that a reform movement began in 1999 towards the development of a system in which all students in the Republic of China can show high-level performance across the country (Şirin & Vatanartıran, 2014). With this system, the content of the educational programmes has been amended so as to be employed in real life by students, arrangements towards student-centered educational approach were implemented, an understanding of education has been adopted that will provide learners to structure

knowledge rather than just memorization and contents of central exams across the country was amended towards the measurement of desired skills (OECD, 2011; Şirin & Vatanartıran, 2014). As a result, according to the findings obtained within the scope of this research, Chinese students participating in the application from four states are observed to be more successful than measurement model expected in open-ended items that they arrange themselves. Similarities between strengths and weaknesses of these four states are considered as a reflection of the curriculum in the country.

The nine countries except Shanghai China were below the OECD average in the top 10 countries which performed above expectations in the continuous text category of the text format dimension of the PISA 2009 reading skills subtest. In the non-continuous text category, eight countries except Malta and Mauritius were above the OECD average. Similarly, seven countries except Macau China, Slovenia and Georgia are above the OECD average in the combined text category. According to the obtained information, it is inferred that countries with higher than expected performance in the non-continuous and combined text categories are generally at the upper level of the competence, while countries with higher performance than expected in the continuous text category have generally reached the lower level of the competence. In a study conducted by Verhelst (2012), it has been concluded that higher performance than estimated by the measurement model is achieved for continuous text category by Brazil, Greece, Mexico, Portugal, Russia, Latvia, Poland, Italy and Hungary; for non-continuous text category by France, Australia, United Kingdom, New Zealand, Netherlands, Ireland, Belgium, Sweden, Canada and United States in the past nine years.

In a study done by Acar (2012), it was aimed to identify countries with similar characteristics to those of Turkey according to PISA 2009 achievements in mathematics, science and reading fields. According to the obtained results, it is seen that Japan, Korea, Hong Kong, Shanghai, Singapore and Finland were collected in the same cluster. As a result of the profile analysis carried out within the scope of this research, Singapore has achieved the high performance in the non-continuous text and the other countries in the combined text category more than expected level. In this way, Finland's clustering with Far Eastern countries is seen as an important finding. However, although Finland's cultural structure and educational policies differs from Scandinavian countries and Far Eastern countries (Ning, Van Damme, Gielen, Vanlaar, & den Noortgate, 2016; Simola, 2005; Soh, 2014; Zhang et al., 2015), it is thought that the similar results with Far Eastern countries (OECD, 2012) may cause to be clustered with them.

When Table 4 is examined, Turkey appears to have similar characteristics with Jordan, Bulgaria, Georgia, Kyrgyzstan, Kazakhstan, Qatar and Albania in terms of strong and weak item categories in the text format dimension. This finding shows that, Turkey are in the same cluster with border neighbors' and Middle East countries in terms of some variables; such as population, geographical location, cultural characteristics, economic development, gross national product per capita and budget allocated for education (Acar, 2012; Turanlı & Deniz, 2008; Yılmaz & Kaya, 2005). Similarly, it is seen that the neighboring European countries Netherlands and Belgium, and American neighbors USA and Canada, perform better than expected in non-continuous text category but perform less than expected in continuous text category. As a result, these similarities and differences between countries can be explained by educational policies, applied educational programmes, language and cultural factors (Kjærnsli & Lie, 2004; Lin & Shi, 2014; Zhang et al., 2015).

According to the findings of this research; it is reached to a conclusion that Turkish students participating in PISA 2009 application showed higher performance than estimated by the model in items within continuous text category. Turkish students' higher performance than expected in continuous texts in the form of paragraphs rather than non-continuous texts needed appropriate visual, is supported by many research findings. In a study conducted by Coşkun (2013), texts in primary education Turkish language textbook were analyzed within the framework of PISA 2009 reading skills, and non-continuous text type articles containing tables, maps, figures and pictures were not stated to be found. Similarly, it is found that such visual tools are not used in the questions of primary school social sciences textbook (Yazıcı, 2006). In another study conducted by Aşıcı et al. (2012), Turkish test

items in the Placement Test (PT) in Primary Education Level conducted in Turkey in 2009 and items from PISA 2009 application reading skills subtest was compared and it is determined that in the items of PT's Turkish test, continuous texts composed of proses rather than figures, tables or graphs were preferred. Although different text types are used in international exams such as PISA, in Turkey, the fact that texts in books published at primary education level consist of only story and poetry types (Yağmur, 2009) or that items in wide-scale exams do not show non-continuous text characteristics result in failure among students to become familiar to non-continuous texts (Savran, 2004). Thus, it is believed that Turkish students' lack of knowledge to interpret visuals causes them to have great difficulty in dealing with such items and this situation adversely affects results obtained by Turkey in PISA application.

In a study conducted by Erbaş, Alacacı, and Bulut (2012), school books in Turkey, Singapore and United States were analyzed and compared in terms of some criteria, and it has been concluded that school books in Singapore contain more visual elements compared to other school books in other countries. Among three countries discussed, it has been determined that Turkey is the country that uses the least visuals in school books. In this research, non-continuous text type containing visuals seems to be hard for Turkish students at most, and Singaporean students at least by comparison with other countries. Thus, it has been concluded that Turkish students read paragraphs in continuous text type throughout their learning lives and they become more familiar with such texts and Singaporean students, however, come across with non-continuous text related to the requirement for visual processing of the information, and these results are also supported in the literature.

Five out of the top 10 countries that performed above expected in the accessing and retrieving knowledge category of cognitive process level dimension of PISA 2009 reading skills subtest, were below the OECD average. In the integrating and interpreting knowledge category, 10 countries were below the OECD average. Similarly, eight countries, except Canada and the United States, were below the OECD average in the reflecting their own thoughts and evaluation category. It is recognized that it is not possible to precisely separate the cognitive process levels considered in PISA applications and it is known that the materials used are defined to emphasize one of these levels (MNE, 2012). This situation is thought to affect the performance of countries in the cognitive process level of the text dimension.

In a study conducted by Verhelst (2012), it has been concluded that higher performance than estimated by the measurement model is achieved for reflecting their own thoughts and evaluation category by Greece, Canada, United Kingdom and Ireland; for integrating and interpreting information by Finland, Germany, Switzerland, Poland and Czech Republic; and for accessing and retrieving information category by Germany, Switzerland, Liechtenstein, Finland, United Kingdom and Denmark. The fact that strong side of these countries do not change in terms of item categories reveals the determination of educational policies.

According to the findings of this research; it can be said that the USA and Canada perform above the expectations in the Reflecting Their Own Thoughts and Evaluation (RTOTE) category, while the Netherlands and Belgium in the Accessing and Retrieving Information (ARI) category as well. Similarly, it can be stated that Scandinavian countries; Denmark, Sweden and Norway, except Finland, outperform than expected in the Accessing and Retrieving Information (ARI) category. Considering that in many studies Finland differ from other Scandinavian countries (Zhang et al., 2015), these similarities and differences among countries can be explained by geographical location, educational policies, applied education programmes, language and cultural factors (Kjærnsli & Lie, 2004; Klieme & Baumert, 2001; Lin & Shi, 2014).

United States showed higher performance than expected in the category of reflecting their own thoughts and evaluation, and became in the second place in this category. It is an important finding that American students have a relative superiority in items in this category that requires the ability to pass judgment regarding the text content by relating information in the text with previous information. It is stated that American educational system motivates students towards research and practice rather than

providing rote learning and allows students to acquire more permanent learnings by realizing rich training experiences in libraries, museums or natural environments such as forests (Armstrong, 2009; Fogarty & Stoehr, 2008 as cited in Baş, 2013). Thus, it can be stated that critical thinking, producing solution and problem-solving skills of American students in a student-oriented environment and learning through experiencing developed more and they do not live any difficulties in items requiring high level of cognitive skills.

According to the findings of this research; it is possible to state that item categories discussed within the framework of PISA reading skills assessment for countries with educational systems developed on the basis of a certain educational philosophy have not changed as years pass, therefore, students' performances in two different applications are similar. In this context, educational programmes of Britain, Ireland and Canada which do not show difference between strength and weakness in the item category of the PISA 2000 and 2009 reading skills subtest, can be examined and compared.

In addition to evaluating PISA results based on general averages, examination at the item categories level with the help of profile analysis ensures that failing aspects of educational policies are accurately identified. For this reason, with taking measure by the Ministry of National Education (MNE) to eliminate deficiencies in the item categories that Turkish students are weak, content of the Turkish lesson curriculum and textbooks can be improved.

At central exams where the total test score is received, the strengths and weaknesses of the students in the item categories can be determined at the school, province and district level by the profile analysis technique using the item parameters estimated based on the Item Response Theory. Thus, school administrators, teachers and families can be provided with more comprehensive information about their performances.

Acknowledgement

We would like to thank Norman Verhelst and Hüseyin Hüsnü Yıldırım for their contributions in preparation of this study.

References

Acar, T. (2012). Türkiye'nin PISA 2009 sonuçlarına göre OECD'ye üye ve aday ülkeler arasındaki yeri. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(4), 2561-2572.

Acar, T., & Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA fen bilimleri performansının incelenmesi. *Eğitim ve Bilim, 37*(163), 178-189.

Anıl, D. (2009). Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim, 34*(152), 87-100.

Armstrong, T. (2009). *Multiple intelligences in the classroom* (3rd ed.). USA: Association for Supervision and Curriculum Development.

Aşıcı, M., Baysal, Z. N., & Şahenk Erkan, S. S. (2012). Türkiye'de yapılan 2009 PISA ve Seviye Belirleme Sınavı'ndaki (SBS) okuma becerileri sorularının karşılaştırılması. *Eğitim ve Öğretim Araştırmaları Dergisi, 1*(2), 210-217.

Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi* (Unpublished master's thesis). Hacettepe University, Institute of Social Sciences, Ankara.

Aydın, A., Erdağ, C., & Taş, N. (2011). 2003-2006 PISA okuma becerileri sonuçlarının karşılaştırmalı olarak değerlendirilmesi: En başarılı beş ülke ve Türkiye. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, *11*(2), 651-673.

Aydın, A., Sarıer, Y., & Uysal, Ş. (2012). Sosyoekonomik ve sosyokültürel değişkenler açısından PISA matematik sonuçlarının karşılaştırılması. *Eğitim ve Bilim, 37*(164), 20-30.

Aydoğdu İskenderoğlu, T., & Baki, A. (2011). İlköğretim 8. sınıf matematik ders kitabındaki soruların PISA matematik yeterlik düzeylerine göre sınıflandırılması. *Eğitim ve Bilim, 36*(161), 287-301.

Aydoğdu İskenderoğlu, T., Erkan, İ., & Serbest, A. (2013). 2008-2013 yılları arasındaki SBS matematik sorularının PISA matematik yeterlik düzeylerine göre sınıflandırılması. *Turkish Journal of Computer and Mathematics Education*, *4*(2), 147-168.

Baker, F. B. (2001). *The basics of item response theory*. United States: ERIC.

Balım, A. G., Evrekli, E., İnel, D., & Deniş, H. (2009). Türkiye'nin PISA 2006'daki durumu üzerine bir inceleme: Fen bilimleri yeterlilik düzeyinin bilgi ve iletişim teknolojilerinin kullanımına göre değerlendirilmesi. *E-Journal of New World Sciences Academy Education Sciences*, *4*(3), 1053-1066.

Baş, G. (2013). Amerika Birleşik Devletleri eğitim sistemi: Karşılaştırmalı bir çalışma. *Journal of Educational Sciences*, *1*, 63-93.

Bayat, N., Şekercioğlu, G., & Bakır, S. (2014). Okuduğunu anlama ve fen başarısı arasındaki ilişkinin belirlenmesi. *Eğitim ve Bilim, 39*(176), 457-466.

Baykul Y., & Güzeller, C. O. (2013). Sosyal bilimler için istatistik SPSS Uygulamalı. Ankara: Pegem Akademi.

Beaton, A. E., Postlethwaite, T. N., Ross, K. N., Spearritt, D., & Wolf, R. M. (1999). *The benefits and limitations of international educational achievement studies*. Paris: International Institute for Educational Planning.

Brozo, W. G., Shiel, G., & Topping, K. (2007). Engagement in reading: Lessons learned from three PISA countries. *International Reading Association, 51*(4), 304-315.

Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2011). *Sosyal bilimler için istatistik* (8th ed.). Ankara: Pegem Akademi.

Coşkun, Y. K. (2013). Türkçe ders kitaplarının PISA sınavı okuma ölçütleri açısından incelenmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, *26*, 22-43.

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory.* Ohio: Cengage Learning.

Cromley, J. G. (2009). Reading achievement and science proficiency: International comparisons from the Programme on International Student Assessment. *Reading Psychology, 30*, 89-118.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. United States: The Guilford Press.

De Gruijter, D. N. M., & Van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences.* United States: Chapman & Hall.

Delil, A., & Yolcu Tetik, B. (2015). 8. sınıf merkezi sınavlardaki matematik sorularının TIMSS-2015 bilişsel alanlarına göre analizi. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 13*(4), 165-184.

DeMars, C. (2010). *Item response theory-Understanding statistics measurement*. United States: Oxford University Press, Inc.

Demir, E. (2010). *Uluslararası Öğrenci Değerlendirme Programı (PISA) bilişsel alan testlerinde yer alan soru tiplerine göre Türkiye'de öğrenci başarıları* (Unpublished master's thesis). Hacettepe University, Institute of Social Sciences, Ankara.

Demir, İ., & Kılıç, S. (2010). Öğrencilerin matematik başarısını etkileyen faktörlerin PISA 2003 kullanılarak incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *38*, 44-54.

Education Reform Initiative. (2005). *Yeni öğretim programlarını inceleme ve değerlendirme raporu*. İstanbul: İstanbul Politikalar Merkezi.

Education Reform Initiative. (2011). *Eğitim izleme raporu 2010*. İstanbul: İstanbul Politikalar Merkezi.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.

Erbaş, A. K., Alacacı, C., & Bulut, M. (2012). Türk, Singapur ve Amerikan ders kitaplarının bir karşılaştırılması. *Kuram ve Uygulamada Eğitim Bilimleri, 12*(3), 2311-2330.

Göktaş, Ö., & Gürbüztürk, O. (2012). Okuduğunu anlama becerisinin ilköğretim ikinci kademe matematik dersindeki akademik başarıya etkisi. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi, 2*(4), 52-66.

Guo, Y. (2014). Cross-cultural comparison of the school factors affecting students' achievement in mathematical literacy: Based on the multilevel analysis of PISA 2012. *China Examinations*, *10*.

Güner, N. (2015). 6.-8. sınıf matematik ders kitaplarındaki geometri, veri ve olasılık sorularının TIMSS bilişsel düzeylerine göre sınıflandırılması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *37*(1), 77-90.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and applications* (2nd ed.). Boston: Academic Publishers Group.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory* (1th ed.). London: Sage Publication.

İş Güzel, Ç. (2006). *A cross-cultural comparison of the impact of human and physical resource allocations on students' mathematical literacy skills in the programme for international student assessment (PISA) 2003* (Unpublished doctoral dissertation). Middle East Technical University, Graduate School of Natural and Applied Science, Ankara.

Kjærnsli, M., & Lie, S. (2004). PISA and scientific literacy: Similarities and differences between the nordic countries. *Scandinavian Journal of Educational Research, 48*(3), 271-286.

Klieme, E., & Baumert, J. (2001). Identifying national cultures of mathematics aducation: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, *16*(3), 385-402.

Köse, İ. A. (2010). *Madde Tepki Kuramı'na dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* (Unpublished doctoral dissertation). Ankara University, Institute of Educational Sciences, Ankara.

Le, L. T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing*, *9*(2), 122-133.

Lin, E., & Shi, Q. (2014). Exploring individual and school-related factors and environmental literacy: Comparing U.S. and Canada using PISA 2006. *International Journal of Science and Mathematics Education*, *12*(1), 73-97.

Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education, 22*(2), 164-184.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Unites States: Lawrence Erlbaum Associates, Inc.

Lyons-Thomas, J., Sandilands, D., & Ercikan, K. (2014). Gender differantial item functioning in mathematics in four international jurisdictions. *Education and Science, 39*(172), 20-32.

Ministry of National Education. (2010). *PISA 2009 Öğrenci Başarılarını Değerlendirme Programı ulusal ön raporu*. Ankara: Eğitimi Araştırma ve Geliştirme Dairesi.

Ministry of National Education. (2012). *PISA Türkiye.* Ankara: Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü.

National Education Association (2012). *Preparing 21st century students for a global society, an educator's guide to the "Four Cs"*. NEA Publishing. Retrieved from http://www.nea.org/tools/52217.htm

Ning, B., Van Damme, J., Gielen, S., Vanlaar, G., & den Noortgate, W. V. (2016). What makes the difference in reading achievement? Comparisons between Finland and Shanghai. *Scandinavian Journal of Educational Research, 60*(5), 515-537.

OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD Publications.

OECD. (2011). *Lessons from PISA for the United States, strong performance and successful reformers in education.* Paris: OECD Publications.

OECD. (2012). *PISA 2009 technical report*. Paris: OECD Publications.

Osterlind, S. J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance and other formats.* Boston: Kluwer Academic Publishers.

Özer Özkan, Y. (2012). *Öğrenci Başarılarının Belirlenmesi Sınavından (ÖBBS) Klasik Test Kuramı, tek boyutlu ve çok boyutlu Madde Tepki Kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması* (Unpublished doctoral dissertation). Ankara University, Institute of Educational Sciences, Ankara.

Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York: Springer.

Rodriguez, M. C. (2002). Choosing an item format. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy and implementation* (pp. 213-232). New Jersey: Lawrence Erlbaum Associates, Inc.

Savran, N. Z. (2004). PISA projesinin Türk eğitim sistemi açısından değerlendirilmesi. *Türk Eğitim Bilimleri Dergisi, 2*(4), 397-414.

Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess?. *Journal of Educational Change*, *8*(4), 349-357.

Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, *41*(4), 455-470.

Soh, K. (2014). Finland and Singapore in PISA 2009: Similarities and Differences in Achievements and school management. *Compare: A Journal of Comparative and International Education*, *44*(3), 455-471.

Sun, L., Bradley, K. D., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Education*, *34(*14), 2107-2125.

Şengül, A. (2011). *Türk öğrencilerin PISA 2009 okuma becerilerini açıklayan değişkenlerin CHAID analizi ile belirlenmesi* (Unpublished master's thesis). Ankara University, Institute of Educational Sciences, Ankara.

Şirin, S. R., & Vatanartıran, S. (2014). *PISA 2012 değerlendirmesi: Türkiye için veriye dayalı eğitim reformu önerileri.* İstanbul: TÜSİAD-T.

Turanlı, M., & Deniz, Ö. (2008). Ülkelerin çok boyutlu ölçekleme analizi ile karşılaştırılması. *Marmara Üniversitesi Sosyal Bilimler Enstitüsü Öneri Dergisi*, *8*(29), 187-192.

Verhelst, N. D. (2001). Testing the unidimensionality assumption of the Rasch model. *Methods of Psychological Research Online*, *6*(3), 231-271.

Verhelst, N. D. (2012). Profile analysis: A closer look at the PISA 2000 reading data. *Scandinavian Journal of Educational Research, 56*(3), 315-332.

Verhelst, N. D. (2014). Test theory: Some basic notions. *Eğitim ve Bilim, 39*(172), 3-19.

Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement, 40*(3), 255-275.

Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional Item Response Theory?. *Educational Research and Evaluation, 18*(4), 307-320.

Yağmur, K. (2009). *Türkçe ders kitapları üst düzey bilişsel becerilerin gelişimini ne oranda sağlamaktadır?*. Ankara: Cito Eğitim Kuram ve Uygulama.

Yazıcı, K. (2006). Sosyal bilgilerde kullanılan görsel araçlar: Haritalar, küreler, resimler, tablolar ve grafikler. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 15, 651-662.

Yıldırım, H. H. ve Yıldırım, S. (2012, September 19-21). *Eski köye yeni adet: Profil analizi*. Paper presented at III. Ulusal Eğitimde Ölçme ve Değerlendirme Kongresi, Abant İzzet Baysal University, Bolu.

Yıldırım, H. H., & Berberoğlu, G. (2009). Judgmental and statistical DIF analyses of the PISA-2003 matematics literacy items. *International Journal of Testing*, *9*(2), 108-121.

Yıldırım, H. H., Yıldırım, S., & Verhelst, N. D. (2014). Profile analysis as a generalized Differential Item Functioning analysis method. *Eğitim ve Bilim, 39*(172), 49-64.

Yıldırım, Ö. (2012). *Okuduğunu anlama başarısıyla ilişkili faktörlerin Aşamalı Doğrusal Modellemeyle belirlenmesi (PISA 2009 Hollanda, Kore ve Türkiye Karşılaştırılması)* (Unpublished doctoral dissertation). Ankara University, Institute of Educational Sciences, Ankara.

Yılmaz, Ö., & Kaya, V. (2005). Genişleme sürecindeki Avrupa birliği: Ekonomik performansa dayalı kümeleme analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *5*(1), 361-376.

Zhang, L., Khan, G., & Tahirsylaj, A. (2015). Student performance, school differentiation, and world cultures: Evidence from PISA 2009. *International Journal of Educational Development*, *42*, 43-53.

Appendix 1

General average scores for countries participating in the PISA 2009 application reading skills subtest and their overall rankings are given in the table below.

Raw No	Countries	Scores	Raw No	Countries	Scores
1	Shanghai China	556	39	Austria	470
2	Korea	539	40	Lithuania	468
3	Finland	536	41	Turkey	464
4	Hong Kong China	533	42	Arab Emirates	459
5	Singapore	526	43	Russia	459
6	Canada	524	44	Chile	449
7	New Zealand	521	45	Costa Rica	443
8	Japan	520	46	Malta	442
9	Australia	515	47	Serbia	442
10	Netherlands	508	48	Bulgaria	429
11	Belgium	506	49	Uruguay	426
12	Norway	503	50	Mexico	425
13	Estonia	501	51	Romania	424
14	Switzerland	501	52	Venezuela	422
15	Poland	500	53	Thailand	421
16	Iceland	500	54	Tobago	416
17	United States	500	55	Malaysia	414
18	Liechtenstein	499	56	Colombia	413
19	Sweden	497	57	Brazil	412
20	Germany	497	58	Montenegro	408
21	Ireland	496	59	Mauritius	407
22	France	496	60	Jordan	405
23	Taipei China	495	61	Tunisia	404
24	Denmark	495	62	Indonesia	402
25	United Kingdom	494	63	Argentina	398
26	Hungary	494	64	Kazakhstan	390
27	Portugal	489	65	Moldova	388
28	Macao China	487	66	Albania	385
29	Italy	486	67	Georgia	374
30	Latvia	484	68	Qatar	372
31	Greece	483	69	Panama	371
32	Slovenia	483	70	Peru	370
33	Spain	481	71	Azerbaijan	362
34	Czech Republic	478	72	India	327
35	Slovakia	477	73	Kyrgyzstan	314
36	Croatia	476	OECD Average		493
37	Israel	474			
38	Luxemburg	472			

(OECD, 2012)