



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Educational Sciences
Educational Measurement and Evaluation Program

DIFFERENTIAL ITEM FUNCTIONING ANALYSIS OF CHILE, COSTA RICA, AND
MEXICO'S PISA 2015 SCIENCE ITEMS

Cozby Janelle DZUL GARCIA

Master's Thesis

Ankara, 2019



With leadership, research, innovation, high quality education and change,

To the leading edge... Toward being the best...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Educational Sciences
Educational Measurement and Evaluation Program

DIFFERENTIAL ITEM FUNCTIONING ANALYSIS OF CHILE, COSTA RICA, AND
MEXICO'S PISA 2015 SCIENCE ITEMS

ŞİLİ, KOSTA RİKA VE MEKSİKA'NIN PISA 2015 FEN MADDELERİNİN ÜZERİNE
DEĞİŞEN MADDE FONKSİYONU ANALİZİ

Cozby Janelle DZUL GARCIA

Master's Thesis

Ankara, 2019

Acceptance and Approval

To the Graduate School of Educational Sciences,

This thesis, prepared by **Cozby Janelle DZUL GARCIA** and entitled “Differential Item Functioning Analysis of Chile, Costa Rica and Mexico’s PISA 2015 Science Items” has been approved as a thesis for the Degree of **Master** in the **Program of Educational Evaluation and Measurement** in the **Department of Educational Sciences** by the members of the Examining Committee.

Chair

Assist. Prof. Dr. Celal Deha
Doğan



Member (Supervisor)

Assoc. Prof. Dr. Burcu Atar



Member

Assist. Prof. Dr. Derya
Çobanoğlu Aktan



This is to certify that this thesis has been approved by the aforementioned examining committee members on 06/07/2018 in accordance with the relevant articles of the Rules and Regulations of Hacettepe University Graduate School of Educational Sciences, and was accepted as a **Master’s Thesis** in the **Program of Educational Evaluation and Measurement** by the Board of Directors of the Graduate School of Educational Sciences on/...../.....

Prof. Dr. Ali Ekber ŞAHİN
Director of Graduate School of Educational Sciences

Abstract

The purpose of this research study is to identify items with culture-related and/or gender-related differential item functioning (DIF), and examine the level of DIF manifested among a set of binary-scored science items from PISA 2015. The data pertaining to 16 dichotomous PISA 2015 science items from 2828 students from three Latin American countries was used to run the DIF analysis through Mantel-Haenszel (MH) and Logistic Regression (LR) procedures across culture and gender groups. Subsequently, the uniform DIF results provided by both techniques were analyzed and DIF effect size was calculated based on the corresponding classification indexes for MH and LR. When the variable of culture was examined, the findings from this investigation revealed 11 items with large DIF (C level): 4 for Chile and Costa Rica, 3 for Chile and Mexico, and 4 for Costa Rica and Mexico. It was found that Costa Rica was more favored than Chile; Chile was more favored than Mexico; yet Costa Rica and Mexico were equally favored. This study detected 3 items with C level DIF due to gender differences. It was found that for Chile the only item detected favored male students, whereas for Costa Rica 1 item favored males and 1 item favored females. After completing this investigation it was concluded that the analyzed set of 16 binary-scored PISA 2015 science items functioned more equally across genders than across cultures.

Keywords: differential item functioning, mantel-haenszel, logistic regression, PISA 2015, chile, costa rica, mexico

Öz

Bu tez çalışmasının amacı, PISA 2015 iki değerli ölçümlenmiş fen maddelerinin cinsiyet veya kültüre göre değişen madde fonksiyonu gösterip göstermediği belirlemek ve gösteren DMF düzeyini incelenmektir. Üç Latin Amerikan ülkesindeki 2828 öğrenciden alınan PISA 2015 iki değerli ölçümlenmiş fen maddelerine ait olan veri, DMF analizini kültür ve cinsiyet grupları arasındaki Mantel-Haenszel (MH) ve Lojistik Regresyon (LR) yöntemleri aracılığı ile işlemek için kullanıldı. Akabinde her iki tekniğin de sunduğu tek tip DMF sonuçları analiz edildi ve DMF etkisinin büyüklüğü, MH ve LR'ye ilişkin sınıflandırma indekslerine dayanarak hesaplandı. Kültür değişkeni incelendiğinde bu araştırmanın bulguları büyük bir DMF (C düzeyi) ile 11 madde ortaya çıkardı: Şili ve Kosta Rika öğrencileri için 4, Şili ve Meksika öğrencileri için 3 ve Kosta Rika ve Meksika öğrencileri için 4 madde. Kosta Rikalı öğrencilerin Şilili öğrencilerden daha avantajlı olduğu; Şilili öğrencilerin Meksikalı öğrencilerden daha avantajlı olduğu; fakat Kosta Rikalı ve Meksikalı öğrencilerin aynı ölçüde avantajlı olduğu keşfedildi. Bu çalışma, cinsiyet farklılıklarından dolayı C düzey DMF ile 3 madde saptadı. Şilili öğrenciler için saptanan tek maddenin sadece erkekler grubu lehine avantajlı olduğu, görülürken Kosta Rikalı öğrenciler için bir maddenin kızlar ve bir maddenin erkekler grubu lehine avantajlı olduğu keşfedildi. Bu araştırma tamamlandıktan sonra analiz edilen PISA 2015 16 iki değerli ölçümlenmiş fen maddelerinin, cinsiyetler arasında kültürler arasında olduğundan daha eşit bir biçimde işlev gördüğü sonucuna varıldı.

Anahtar sözcükler: değişen madde fonksiyonu, mantel-haenszel, lojistik regresyon, PISA 2015, şili, kosta rika, meksika

Dedication

To God, the Giver of all wisdom and who has faithfully sustained me throughout the writing of this research paper.

To my parents, Cozby GARCIA and José DZUL, for supporting this project with their prayers and advice.

To the students and educators that will benefit from this study.



Acknowledgements

I thank God for enabling me to complete this research project and my Master's program.

I thank my family, Cozby GARCIA, José DZUL, Jonatán DZUL, and Fabián DZUL, for giving me unconditional love, support, motivation, and inspiration all the way from Mexico and the United States.

I thank my thesis advisor, Assoc. Prof. Dr. Burcu ATAR, for patiently instructing me, providing valuable remarks, and being ever ready to offer much needed advice and encouragement.

I thank the members of the Examining Committee, Assist. Prof. Dr. Celal Deha DOĞAN and Assist. Prof. Dr. Derya ÇOBANOĞLU AKTAN, for their key observations that made this study better.

I thank my friend and partner in this academic journey, Judith ESPINOZA, for making the tough ride more enjoyable by being always there for me as we overcame challenges together and cheered each other up.

I thank my friends Galiya RAZYM, Harriet LUYANDO, and Marcella MORALES, for becoming my family in Turkey, sharing their timely advice, bringing joy into my life, and enriching my life with theirs.

I thank every HU Department of Educational Sciences' professor, assistant professor, and classmate that helped me while I wrote this paper and during my graduate studies.

Table of Contents

Abstract	ii
Öz	iii
Dedication	iv
Acknowledgements	v
List of Tables	viii
List of Figures	ix
Abbreviations	ix
Chapter 1 Introduction	1
Standardized Achievement Tests	1
Programme for International Student Assessment (PISA)	2
Scientific Literacy	3
Item Bias and Differential Item Functioning (DIF)	4
Types of DIF	5
DIF Detection	6
Research Purpose and Importance	7
Research Questions	8
Sub Research Questions	8
Assumptions	9
Limitations	9
Chapter 2 Review of Related Research Studies	10
DIF Analyses of Achievement Assessments	10
PISA Studies	12
Chapter 3 Methodology	16
Type of Research	16
Research Sample	16
Analyzed Region: Latin America	16

Analyzed Countries.....	17
Data Selection Process.....	17
PISA's Structure	18
Data Analysis	18
Mantel-Haenszel (MH) Method.....	18
Logistic Regression (LR) Method	22
Chapter 4 Findings and Interpretation.....	25
Interpretation Guidelines for MH and LR Methods.....	25
Findings and Interpretation Related to Sub Research Question 1.....	26
Findings and Interpretation Related to Sub Research Question 2.....	38
Overall Findings	50
Chapter 5 Results and Recommendations.....	51
Results Related to Sub Research Question 1	51
Results Related to Sub Research Question 2	52
Overall Results.....	53
Recommendations	54
References	56
APPENDIX-A: Probability of Correct Answer Graphs by Culture and Gender.....	63
APPENDIX B: SPSS Syntaxes for LR and Probability of Correct Answer Graphs.....	69
APPENDIX-C: Ethics Committee Approval	83
APPENDIX-D: Declaration of Ethical Conduct	84
APPENDIX-E: Thesis Originality Report	85
APPENDIX-F: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	86

List of Tables

Table 1 <i>Scientific Literacy Assessment Framework Aspects for PISA 2015</i>	4
Table 2 <i>Methods for DIF Detection</i>	7
Table 3 <i>Data for the j^{th} Matched Set of Members of R and F</i>	20
Table 4 <i>DIF Classification Index for MH</i>	22
Table 5 <i>DIF Classification Indexes for LR</i>	24
Table 6 <i>Culture-related DIF Classification for Chile-Costa Rica Based on MH</i>	36
Table 7 <i>Culture-related DIF Classification for Chile-Mexico Based on MH</i>	37
Table 8 <i>Culture-related DIF Classification for Costa Rica-Mexico Based on MH</i> .	28
Table 9 <i>Culture-related DIF Classification for All Country Pairs Based on MH</i>	29
Table 10 <i>Culture-related DIF Classification for Chile-Costa Rica Based on LR</i> .	30
Table 11 <i>Culture-related DIF Classification for Chile-Mexico Based on LR</i>	31
Table 12 <i>Culture-related DIF Classification for Costa Rica-Mexico Based on LR</i>	32
Table 13 <i>Culture-related DIF Classification for All Country Pairs Based on LR</i>	33
Table 14 <i>Summary of the Culture-related DIF Detection MH and LR Analyses</i> ...	34
Table 15 <i>Gender-related DIF Classification for Chile Based on MH</i>	38
Table 16 <i>Gender-related DIF Classification for Costa Rica Based on MH</i>	39
Table 17 <i>Gender-related DIF Classification for Mexico Based on MH</i>	40
Table 18 <i>Gender-related DIF Classification for All Countries Based on MH</i>	41
Table 19 <i>Gender-related DIF Classification for Chile Based on LR</i>	42
Table 20 <i>Gender-related DIF Classification for Costa Rica Based on LR</i>	43
Table 21 <i>Gender-related DIF Classification for Mexico Based on LR</i>	44
Table 22 <i>Gender-related DIF Classification for All Countries Based on LR</i>	45
Table 23 <i>Summary of the Gender-related DIF Detection MH and LR Analyses</i> ...	46

List of Figures

<i>Figure 1.</i> Example of an item displaying significant uniform DIF	6
<i>Figure 2.</i> Example of an item displaying significant non-uniform DIF	6
<i>Figure 3.</i> Probability of correct answer graph for item 10 (DS610Q01C) in the cases of Chile and Costa Rica	35
<i>Figure 4.</i> Probability of correct answer graph for item 13 (CS626Q01S) in the cases of Chile and Costa Rica	35
<i>Figure 5.</i> Probability of correct answer graph for item 1 (CS641Q01S) in the cases of Chile and Mexico	36
<i>Figure 6.</i> Probability of correct answer graph for item 7 (CS601Q01S) in the cases of Chile and Mexico	36
<i>Figure 7.</i> Probability of correct answer graph for item 2 (CS641Q02S) in the cases of Costa Rica and Mexico	37
<i>Figure 8.</i> Probability of correct answer graph for item 9 (CS601Q04S) in the cases of Costa Rica and Mexico	37
<i>Figure 9.</i> Probability of correct answer graph for item 1 (CS641Q01S) in the case of Chile	47
<i>Figure 10.</i> Probability of correct answer graph for item 7 (CS601Q01S) in the case of Chile	47
<i>Figure 11.</i> Probability of correct answer graph for item 1 (CS641Q01S) in the case of Costa Rica	48
<i>Figure 12.</i> Probability of correct answer graph for item 16 (DS626Q04C) in the case of Costa Rica	48
<i>Figure 13.</i> Probability of correct answer graph for item 8 (CS601Q02S) in the case of Mexico	49
<i>Figure 14.</i> Probability of correct answer graph for item 13 (CS626Q01S) in the case of Mexico	49
<i>Figure 15.</i> Graphs of culture-related DIF results from MH and LR methods	52
<i>Figure 16.</i> Graphs of gender-related DIF results from MH and LR methods	53

Abbreviations

DIF: Differential Item Functioning

ETS: Educational Testing Service

LR: Logistic Regression

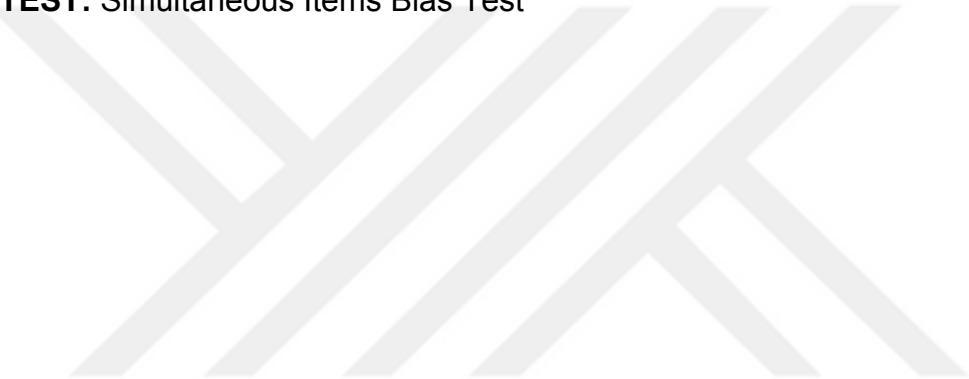
MC: Multiple Choice

MH: Mantel-Haenszel

OECD: Organisation for Economic Co-operation and Development

PISA: Programme for International Student Assessment

SIBTEST: Simultaneous Items Bias Test



Chapter 1

Introduction

Through the centuries, assessment has been essential to the fostering of progress in various domains. The field of education is no exception. Educational assessment advances student learning and development through the organized collection, analysis, and use of data about programs (Palomba & Banta, 1999 as cited in Hadjicharalambous, 2014). To accomplish such important goals, educational assessment relies on numerous assessment instruments and methods, testing being one of them.

A *test* is defined as the process of quantifying an individual's aptitude in a given subject (Brown, 2003 as cited in Hussain, Tadesse, & Sajid, 2015). Tests are usually applied with the aim of measuring the progress towards course objectives or educational programs; such tests are called *achievement tests* (Anastasi, 1976). Although not all achievement tests are standardized, and standardized exams are not the only way to assess academic achievement, they are often the preferred option.

Standardized Achievement Tests

A *standardized test* is a test that has clearly specified procedures for administration and scoring—including normative data (Cohen, Swerdlik, and Phillips, 1996). In other words, when across-the-board, all components of the test follow the same guidelines, conditions, and grading, a test is standardized (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], & Joint Committee on Standards for Educational and Psychological Testing [JCSEPT], 2014). Consequently, a standardized achievement test is one which was designed to assess the knowledge and skills students have achieved in reference to a specific instruction program. These tests have gone through a standardization process. It is said that such tests are designed to gage the outcome of comparative educational practices and commonly result in the final assessment of a student's understanding in a given subject (Anastasi, 1976).

When crafting standardized achievement tests, it is essential for test developers to create unidimensional tests and items. Unidimensionality occurs when one fundamental measurement construct, called dimension, causes a change in test taker answers (Yu, Popp, DiGangi, & Jannasch-Pennell, 2007). Unidimensionality is a critical requirement for measurement when the focus of research is to detect growth or change because, if the test answers are determined by more than one measurement construct (dimension), it is uncertain which construct accounts for the test answers variation (de Gruijter & van der Kamp, 2008).

Some of the strengths of standardized evaluations of learning are: (a) ease of administration and grading, (b) objectivity, (c) coverage of many topical areas, and (d) the ability to generate longitudinal data that make it possible to assess student performance over time (Black & Duhon, 2003; Walstad, 2001 as cited in Hadjicharalambous, 2014). It comes as no surprise that with time, these types of examinations reached wider population samples and now even international assessment instruments have been created.

Programme for International Student Assessment (PISA)

The Organisation for Economic Co-operation and Development (OECD), which is an organization committed to promoting policies that will improve the economic and social well-being of people around the world (OECD, 2016a), launched in 1997 an international assessment in response to the inquiry “what is important for citizens to know and be able to do?” and to the “need for cross-nationally comparable evidence on student performance” (OECD, 2016b, p. 10).

According to OECD (2016b) the Programme for International Student Assessment (PISA) has been applied since 2000 to OECD members and partners in order to determine if 15-year-old students at the end of their required education, are equipped with the necessary knowledge and skills to live in, and contribute to, modern society. PISA focuses on the three core subjects of science, reading, and mathematics, testing in detail one core subject in each year of assessment.

PISA is applied every three years to 15-year-old students from schools worldwide selected at random. PISA tests their skills and knowledge, thus evaluating education systems. PISA relies on a peculiar way of assessment: first,

on tests (cognitive assessment) crafted to evaluate how equipped students are, at the end of required education, to apply their knowledge to real-life situations and participate fully in society; second, on background questionnaires which provide context information and aid researchers during the results interpretation stage.

The data collected by PISA is very valuable since it allows participating economies to measure the achievement of students in their own countries in comparison with those elsewhere. PISA compares the performance of their students over time, assesses the effect of education policy decisions, establishes policy targets against objective goals which other education systems achieved, and learns from other countries' implemented policies and practices (OECD, 2014). As a result, PISA has gained acceptance over the years and more countries have joined in.

Although PISA cannot find cause-and-effect relationships between policies/practices in the education field and student outcomes, it has proven to be useful for educators, policy makers and the interested public. It shows them how education systems are similar and different, and how that impacts the examinees (OECD, 2014). Given that PISA is considered a significantly influential international education and policy assessment tool, a thoughtful analysis of its data is useful to discern outcomes regarding the condition of the countries' education systems, the students' academic achievement, in order to later ask the appropriate 'why' questions about the revealed situation (Baird et al., 2011; Wiseman, 2010 as cited in Caro, Lenkeit, & Kyriakides, 2016).

The last PISA assessment was done in 2015 counting 540,000 15-year-old test takers from 72 participating countries. This was the first time that most of the examinees took the test on computer (OECD, 2016d). PISA 2015 encompassed the assessment of science, reading, mathematics, and collaborative problem solving areas with an optional assessment of financial literacy. However, its focus was on science literacy (OECD, 2016e).

Scientific Literacy. OECD (2016b) defines scientific literacy as “the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen” (p. 13). OECD (2016b) specifies that in PISA 2015, proficiency in science is assessed through questions related to specific contexts, knowledge,

competencies, and attitudes. Moreover, table 1 lists what each of these elements involves.

Table 1

Scientific Literacy Assessment Framework Aspects for PISA 2015

Aspect	Specifications
Contexts	-Personal -Local/National -Global
Knowledge	-Content knowledge -Procedural knowledge -Epistemic knowledge
Competencies	-To explain phenomena scientifically -To evaluate and design scientific inquiry -To interpret data and evidence scientifically
Attitudes	-Interest in science and technology -Valuing scientific approaches to enquiry, where appropriate -Perception and awareness of environmental issues

PISA 2015 emphasis on science proficiency serves as a reminder that, because we live in a world molded by science-based technology, science comprehension is useful outside the lab and needed for full participation in society (OECD, 2016e).

So far the present chapter has examined key introduction topics such as educational assessment instruments, standardized evaluations of learning, and PISA. The next part of this chapter will briefly explain when item bias and differential item functioning (DIF) happen, what are the types of DIF, and how to detect DIF.

Item Bias and Differential Item Functioning (DIF)

Three fundamental measurement qualities –validity, reliability and fairness (absence of bias)– guide educational assessment practices, being absence-of-bias evaluation a key procedure in this field (Tierney, 2016). When running

fairness analyses, it is important to discern between item bias and differential item functioning (DIF).

When a specific test item characteristic or testing situation irrelevant to the purpose of a test causes test takers from one group to be less likely to answer an item correctly than the ones from another group, item bias occurs (Zumbo, 1999). Therefore, it could be said that an unbiased item is one which allows for all individuals with the same ability level to have the same chance of getting the right answer no matter which group they belong to (Asil, 2005).

DIF in the words of Teresi et. al. (2008) is “the evaluation of conditional relationships between item response and group membership” (p. 540) while Holland & Wainer (1993) describe it as “a difference in item performance between two comparable groups of examinees, that is, groups that are matched with respect to the construct being measured by the test” (p. 35). Consequently, it is said that DIF occurs when examinees with equal ability level, participating in an examination where the same skill is being assessed –but belonging to different groups– have a different probability of answering correctly.

Now, DIF is required, but not sufficient condition for item bias; the fact that DIF is not manifested for an item means that no item bias is manifested either, however, the presence of DIF is not sufficient to affirm the presence of item bias and follow-up analyses would be necessary to determine if item bias is apparent or not (Zumbo, 1999). Throughout the years numerous techniques for identifying item bias have been used. Nevertheless, item bias procedures proved to be insufficient in various cases and researchers began to rely on a more descriptive method called Differential Item Functioning (DIF).

Types of DIF. It is important to point out that two types of DIF have been recognized: uniform and non-uniform. Uniform and non-uniform DIF vary in that non-uniform DIF shows a connection between group membership and ability levels, while uniform DIF shows no interaction between the two (Baştuğ, 2016). In other words, uniform DIF (Figure 1) happens when there is a greater probability of a correct response for one group uniformly over all ability levels. Non-uniform DIF (Figure 2) happens when the two groups display different probabilities of

answering an item correctly at all levels of ability (Rogers & Swaminathan, 1993; Camilli & Shepard, 1994 as cited in Le, 2006).

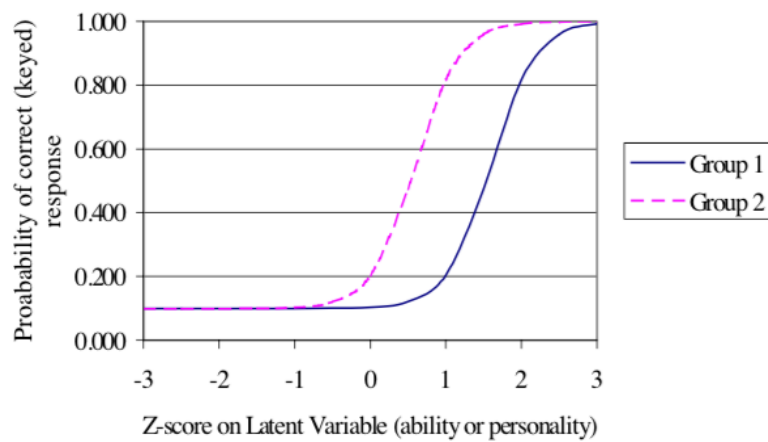


Figure 1. Example of an item displaying significant uniform DIF

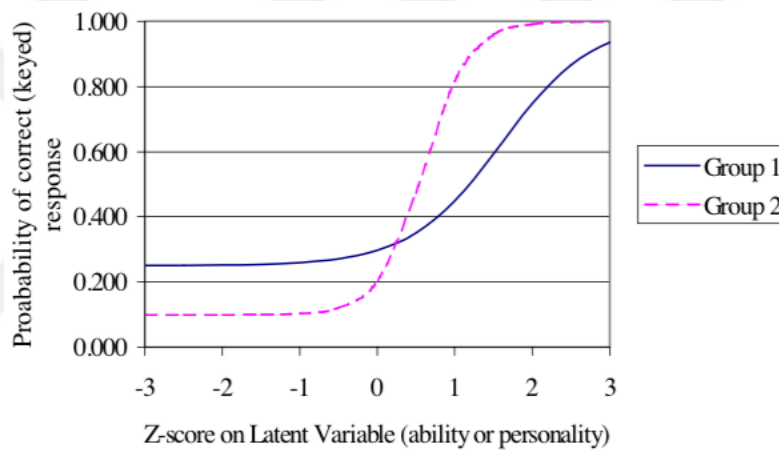


Figure 2. Example of an item displaying significant non-uniform DIF

DIF Detection. When assessing DIF, one must recall that “for DIF to be useful it is not sufficient to simply have a statistical test of DIF but one also needs a measure of the magnitude of the DIF effect” (Zumbo, 1991, p. 7). Teresi et al. (2008) says that the magnitude of DIF is the DIF degree, which is measured through the analysis of method-related parameters or statistics. Thus, it can be said that a comprehensive DIF analysis must include statistical DIF values and the corresponding DIF degree data.

Furthermore, as DIF analyses are implemented, the choice of procedures to use becomes more relevant. Atalay Kabasakal, Arsan, Gök, & Kelecioğlu (2014) affirm that DIF detection plays an important role in creating unbiased and reliable tests, hence finding the most accurate methods provides the most accurate results.

Hambleton, Merenda, and Spielberger (2005) supply brief yet very useful information about the most popular DIF methods in table 2.

Table 2

Methods for DIF Detection

Method	Sources	Appropriate for
Delta Plot	Angolff (1972, 1993)	Dichotomous data
Standardization	Dorans & Kulick (1986); Dorans & Holland (1993)	Dichotomous data
Mantel-Haenszel	Holland & Thayer (1988); Dorans & Holland (1993)	Dichotomous data
Logistic Regression	Swaminathan & Rogers (1990)	Dichotomous data Polytomous data Multivariate matching
Lord's Chi-Square	Lord (1980)	Dichotomous data
IRT Area	Raju (1988, 1990)	Dichotomous data Polytomous data
IRT Likelihood Ratio	Thissen et al. (1988) Thissen et al. (1993)	Dichotomous data Polytomous data
SIBTEST	Shealy & Stout (1993)	Dichotomous data

In order to accurately identify DIF several procedures or methods can be used, which are commonly divided into the categories of parametric and non-parametric. For this study two methods were implemented: Mantel-Haenszel (MH) –non-parametric– and Logistic Regression (LR) –parametric–. These two procedures have several advantages in common: easy calculation, a more accurate magnitude of DIF, suitability for dichotomous data analysis, feasibility with large data samples, and easy interpretation (Swaminathan & Rogers, 1990).

Research Purpose and Importance

It is the aim of this research study to identify items with culture-related and/or gender-related DIF among 16 binary-scored PISA 2015 science items from three Latin American countries –Chile, Costa Rica, and Mexico–.

All the countries considered for this study have Spanish as the official language and were ranked below the OECD average in their overall performance in PISA 2015. Although all three countries are considered part of the Latin American region of the world, sharing a somewhat common culture, the more specific geographical location –Mexico in North America, Costa Rica in Central America, and Chile in South America– causes cultural differences that may end up favoring one country more than the others.

In fact, these three Latin American countries –along with the rest of Latin America– are notorious for their usual weak test scores, lack of qualified educational assessment personnel, and poor management of large-scale examinations (Martínez-Rizo, 2016). Therefore, the present research will not only contribute to each of these countries, but to the Latin American region studies as well.

Additionally, it is necessary to point out that this study is unique given that no other DIF analysis of PISA has examined the three chosen countries and renders valuable insight due to the use of real data. Thus the current investigation will prove to be relevant for the educational assessment field's literature.

Research Questions

This study will provide an answer to the following questions:

1. Are there any DIF items present among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?
2. What level of DIF is manifested among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

Sub research questions. The next questions will also be answered through this study:

1. Are there any items that indicate culture-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?
2. Are there any items that show gender-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

Assumptions

The present investigation was developed on the following basis:

1. It is assumed that cultural differences of students from Chile, Costa Rica, and Mexico affected their science performance in PISA 2015 due to the adaptation items underwent as part of the standardization process.
2. It is assumed that gender differences of students from Chile, Costa Rica, and Mexico affected their science performance in PISA 2015 due to the natural features of each gender.

Limitations

The features listed below limited this investigation:

1. The DIF analysis was run with only the data pertaining to Chile, Costa Rica, and Mexico.
2. This study was limited to 16 selected dichotomous PISA 2015 science items.
3. The comparison of findings was done only pair-wise.
4. The estimation of DIF was limited to just uniform DIF.
5. This research had a limiting condition of no access to item content.
6. The chosen DIF detection methods did not involve any multi-group DIF methods.

Chapter 2

Review of Related Research Studies

This chapter presents a series of summaries from directly or indirectly related research studies that have shed some light on the main themes addressed by the current investigation. Since these studies pertain to two main themes, they have been organized in two subsections: DIF analyses of achievement assessments and PISA studies.

DIF Analyses of Achievement Assessments

Gnaldi and Bacci (2016) conducted a research study which tested unidimensionality and the absence of DIF in two national tests administered to lower secondary school students in Italy. They based their study on the data obtained by the National Institute for the Evaluation of the Education System (INVALSI) in June 2009 from a sample of 27,592 Italian students. The data corresponded to the Italian students' performance in the Italian Language Test – composed of a reading comprehension section and a grammar section– and the Mathematics Test. The results proved the presence of gender-related and geographical area-related DIF, especially in the grammar section of the Language Test, as well as the presence of multidimensionality in both sections of the Language Test but particularly in the grammar section.

García-Medina, Martínez-Rizo, and Cordero-Arroyo (2016) ran a DIF analysis using a substantially modified version –based on the Rasch model– of the MH procedure to determine if DIF because of gender or socioeconomic level differences was present in the items from the mathematics scale of the 2012 Examination of Educational and Quality Attainment in Mexico. The data from a national sample of 27,009 students in the third year of secondary school was examined. The findings revealed that none of the 100 math scale items presented DIF by gender. But 18 of those items did present severe DIF by socioeconomic level.

Latifi, Bulut, Gierl, Christie, and Jeeva (2016) carried out a study to assess fairness of a suite of multi-subject national Secondary School Certificate (SSC) tests administered at grade 10 in South Asian countries such as Bangladesh,

India, and Pakistan. They specifically tested for possible gender-related DIF in the multiple-choice (MC) items of the English, Mathematics, and Physics 2011 qualification examinations of 103 randomly chosen Pakistani schools. The results showed that examinees from both genders performed equally across all three subjects and did not show any major threat to the fairness and the validity of the SSC assessments.

Çepni (2011) evaluated the 2008 Fall application items from the Academic Staff and Postgraduate Education Entrance Examination Quantitative Ability tests for potential DIF and Differential Bundle Functioning (DBF) by gender and by department of completed/almost completed studies. Various methods were used for DIF analyses and SIBTEST was the only one used for DBF analyses. The gender-related DIF analyses flagged 7 items –3 in favor of males, 4 in favor of females– in the Quantitative Test 1 and 4 items –1 in favor of males, 3 in favor of females– in the Quantitative Test 2, whereas the department-related DIF analyses flagged 9 items –5 in favor of quantitative departments, 4 in favor of verbal departments– in the Quantitative Test 1 and 3 items –all in favor of quantitative departments– in the Quantitative Test 2. Additionally, according to the gender-related DBF analyses, operation items gave an advantage to females and word problem items to males in both tests, while the department-related DBF analyses detected that operation and geometry items resulted favorable to quantitative departments in both tests.

Doğan and Öğretmen (2008) examined the performance of three DIF detection techniques: Chi Square, Mantel-Haenszel, and Logistic Regression. Their sample consisted of 3345 Turkish students who entered the Secondary Education Institutions' Student Selection and Placement Test (SEISSPT) which was held country-wide in 2003. When the data from the science subtest's answers was analyzed, only the variable of gender was considered. The results yielded 15 DIF items by MH, 11 DIF items by Non-signed Chi Square, 0 DIF items by signed Chi Square, and 9 DIF items by LR. In conclusion, it can be said that all three DIF detection methods studied were similar in terms of the size of their corresponding Chi Square values, yet different in regard to the amount of items with DIF they yielded.

Asil (2005) performed a DIF analysis of the verbal section of Turkey's 2003 Student Selection Examination (SSE) through the MH procedure, aiming to identify any gender-related DIF, and examine the relationship between DIF values and item characteristics such as item difficulty and item discrimination index. He studied the data pertaining the Turkish and Social Science subtests from 2000 test takers. The results indicated that 4 C items were marked in the Turkish section and 1 C item in the Social Science section. While the females were found more favored than the males in the Turkish subtest, the Social Science subtest did not favor either group. Moreover, no significant relationship between item difficulty and DIF or between item discrimination and DIF was found in the case of the Turkish section of the test.

Bolt (2000) assessed the items from the mathematical sections of the Scholastic Assessment Test (SAT) by relying on a modified version of the Simultaneous Items Bias Test (SIBTEST) procedure in order to determine the presence of gender-related DIF, the effects of two DIF factors –item format and problem type– and the effect consistency across items. The test data corresponded to approximately 7000 test takers and included 60 items from three operational sections and 25 items from a variable pretest section, resulting in a total of 113,216 item responses. This research displayed a small yet consistent effect due to item format –favoring the male group in MC items– across items, while a larger yet less consistent effect of problem type was exhibited.

PISA Studies

Kartal and Kutlu (2017) researched the scientific literacy test of PISA 2015 in Turkey to determine the role of self-efficacy and motivation in science; to the success of both high and low performing students. Their sample corresponded to 5,287 students and they relied on regression models and a quantile regression analysis to carry out this study. It was found that science self-efficacy and motivation –both intrinsic and instrumental– towards science were significant predictors of science success.

Uzun and Gelbal (2017) ran a DIF analysis of PISA 2006 science items taking into consideration the performance of examinees from Australia, Canada, England, and Turkey, in respect to language and culture. MH, LR, along with

signed-unsigned area indexes were the chosen DIF detection methods in this case. It was shown that an increase in the linguistic and cultural differences between countries caused an increase in the amount of DIF items, and that the three different applied methods provided significantly varied number of DIF items. As part of the analysis, field specialists were consulted and identified various DIF sources: changes in word difficulty; changes in item format; differences in cultural relevance, translation, curriculum; linguistic variances; and changes in content across Turkish and English versions of the items.

Ercikan, Roth, and Asil (2015) compared five countries from three continents, and with varied rankings on the reading test of PISA 2009, to elaborate on the issues of international assessments diagnosis, by cautioning against limited and/or faulty interpretation of country rankings. For this study the performance of students from Canada, Shanghai-China, Germany, Turkey and the United States was analyzed and it showed that two factors –reading enjoyment and out of school enrichment activities– cannot entirely be responsible for higher performance scores. Furthermore, the study detected that reading levels have significant variations in different student sub-groups within a country, implementing one education system based on gender, language, and cultural group.

Demir and Köse (2014) performed MH, LR, and SIBTEST procedures to the mathematics items of PISA 2009 in order to detect if gender and culture DIF were evident. They considered the answers from the PISA 2009 mathematics fifth booklet in Turkey, Germany, Finland, and the United States for the culture variable and the data from 332 mathematics literacy sub-test takers for the gender variable. The findings revealed the presence of DIF caused by gender differences in 5 items –4 of them favoring the male group and 1 favoring the female group–, whereas DIF caused by culture differences was present in 16 items for Turkey and Germany's students, 14 items for Turkey and Finland's students, and 18 items for Turkey and United States' students.

Asil and Gelbal (2012) examined the cross-lingual and cross-cultural equivalence of the PISA 2006 Student Questionnaire across four countries located over the globe –Turkey, United States of America, Australia, and New Zealand–. The thorough study consisted of three stages: Confirmatory Factor Analysis (CFA) at the first stage, Multi-Group Confirmatory Factor Analyses (MG-CFA) based on

Mean and Covariance Structures (MACS) at the second stage, and identification of DIF sources by experts at the third and last stage. MG-CFA results revealed some DIF items between countries and it was concluded that when linguistic and cultural differences between countries increased, the number of DIF with items increased. Experts detected that the main sources of DIF were poor translation and differences in cultural relevance.

Dossey, McCrone, Turner & Lindquist (2008) compared the PISA 2003 results for Canada, Mexico, and the United States, paying special attention to mathematical literacy and cross-disciplinary problem-solving. This paper found significant differences in the overall assessment of mathematical literacy among the three studied North American nations –Canada was the top performer, next the US, and last Mexico–. In addition, corresponding statistical values showed significant gender differences, identifying the males of the three countries as the advantaged group. However, the identified gender differences are minor in comparison to the overall scope of performances.

Le (2006) evaluated the Field Trial Science items for 2006 PISA for potential uniform DIF by gender, country of test, and test language, implementing an analysis of DIF with an IRT approach. The data gathered included 210 science items and corresponded to approximately 84,000 students –about 51% were females and 49% were males– from almost all 50 participating countries –excluding a few countries with late data submission–. This study unveiled a lower DIF caused by gender differences than DIF caused by country of test or test language.

Yıldırım (2006) executed a DIF analysis through Restricted Factor Analysis, MH statistics, and Item Response Theory Likelihood Ratio methodologies in order to assess potential DIF items in the Turkish and English versions of the mathematics section of Trends in International Mathematics and Science Study (TIMSS) 1999 and PISA 2003. The results evidenced adaptation issues in both international assessments. Yet a group of fairly-functioning items across cultures was determined. In addition, the type of competency required to answer items played an important role in terms of Turkish students or American students being favored or not.

In the present study a DIF analysis of the PISA 2015 science items was conducted in order to test for possible uniform DIF caused by gender and by culture in the cases of Chile, Costa Rica, and Mexico. Cultural differences, as well as the gender factor were taken into consideration as MH and LR procedures were implemented to assess the fairness of the scientific literacy section of PISA 2015.



Chapter 3

Methodology

Type of Research

In view of the objectives of the present study, the fitting research design identified is the descriptive design. Kumar (1999) asserted that a descriptive research seeks to describe systematically a situation, problem, phenomenon, service or program, or provides information about specific conditions, or describes an issue. In other words, this study is intended to make known an existing situation.

This study will rely on two of the Differential Item Functioning (DIF) methods –Mantel-Haenszel and Logistic Regression– to accomplish an accurate detection of DIF items among the PISA 2015 dichotomous science items, by running the corresponding analyses according to the culture (country) and gender variables.

Research Sample

OECD (2016e) estimated that 540,000 students from 72 countries completed the 2015 PISA computer-based administration. Nevertheless, for the purpose of this analysis, the sample was composed of only the data from PISA 2015 pertaining to the science items of cluster number 12 from the countries of Chile, Costa Rica, and Mexico. At the end, the data pertaining to 16 multiple-choice (MC) dichotomous PISA 2015 science items from 2828 students from three Latin American countries was used to run the DIF analysis through Mantel-Haenszel (MH) and Logistic Regression (LR) procedures across culture and gender groups.

Analyzed Region: Latin America. The term Latin America refers to “20 countries: the 10 Spanish- and Portuguese-speaking countries of South America; the six Spanish-speaking countries of Central America; Mexico; and Cuba, the Dominican Republic, and Haiti in the Caribbean region” (Latin American and Caribbean Demographic Centre [CELADE], 2002 as cited in Brea, 2003, p. 5). The countries that make up this region of the world share similarities in language and culture, but at the same time are distinctive from each other.

Analyzed Countries. Out of the various countries that participated in PISA 2015, this research study focused on Chile, Costa Rica, and Mexico. These three countries share Spanish as their official language yet they are located in different parts of the American continent –Chile in South America, Costa Rica in Central America, and Mexico in North America– (Merriam-Webster, Inc., 1997; Grey House Publishing, Inc., 2003). Chile and Mexico are OECD members while Costa Rica is currently in the accession process (OECD, 2018), and in PISA 2015 the students from these Latin American countries performed below the OECD average in science, reading, and mathematics (OECD, 2016e).

Data Selection Process

For this study, part of the PISA 2015 database (secondary data) already collected and provided by OECD was used. This database was acquired from the OECD website.

After downloading the PISA 2015 database, it became evident that only two datasets were pertinent to the present study: Cognitive Student Questionnaire –including basic student and school information, as well as reading, mathematics, and science items responses– and Student Questionnaire –including gender, survey replies to questions about the student, his/her family, home conditions, school and classroom conditions–.

After identifying the two datasets that were needed, the next step was to examine the 'PISA 2015 Technical Report 2015 Annex A Item Pool Classification' file where detailed information about all items was disclosed. It was then determined that Science Cluster S12 had the largest number of MC dichotomous items (16) and it was selected for this analysis. One MC item –with item code CS637Q02S– was discarded because it was not a binary-scored item (e.g. correct/incorrect). A total of 16 items were kept for the sample, 12 MC binary-scored items and 4 open response yet binary-scored items.

Then the data cleansing process began and the Cognitive Student Questionnaire dataset underwent the removing of data corresponding to all the countries that were not going to be analyzed until only data from Chile (country ID 152), Costa Rica (country ID 188), and Mexico (country ID 484) remained. Following that, unnecessary variables were also deleted (from 2048 variables

down to 25). Seventeen variables –Country ID and 16 science items from the chosen cluster– were identified as the most essential ones to this research.

The intended DIF analysis called for the student's gender variable but the cognitive dataset did not have it. Therefore, that variable (ST004D01T) was pulled from the Student Questionnaire dataset. Separate datasets by countries were created (Chile, Costa Rica, Mexico) in order to subsequently run the corresponding gender-related DIF analyses.

Afterwards, the variable 'TOTAL' was calculated and added to all 3 datasets. This process revealed missing data –which occurs when a student's answer in PISA is labeled as valid skip, not reached, not applicable, invalid, or no response– and that data was discarded. Chile initially had data from 7053 students (missing data from 6099) and was reduced to 954. Costa Rica initially had data from 1348 students (missing data from 521) and was reduced to 827. Mexico initially had data from 1647 students (missing data from 600) and was reduced to 1047.

Three final datasets by countries consisted of 27 variables (25 variables previously kept plus gender and total). Separate datasets by country pairs were created (Chile-Costa Rica, Chile-Mexico, Costa Rica-Mexico) in order to later run the corresponding culture-related DIF analyses.

The final data included 16 science items with the following codes: CS641Q01S, CS641Q02S, CS641Q03S, CS641Q04S, DS637Q01C, DS637Q05C, CS601Q01S, CS601Q02S, CS601Q04S, DS610Q01C, CS610Q02S, CS610Q04S, CS626Q01S, CS626Q02S, CS626Q03S, DS626Q04C.

PISA's Structure. Yu, Popp, DiGangi, and Jannasch-Pennell (2007) point out that assessing the dimensionality of an evaluation instrument is necessary as part of the process of collecting evidence to backup the validity of interpretations based on test scores. The previous statement begs the question: are PISA 2015 science items unidimensional or multidimensional? According to OECD (2016f) a dimensionality assessment was done to new and existing science items and model-fit statistics demonstrated that a unidimensional model fits the data, concluding that both new and existing items make up a coherent and reliable

unidimensional scale. Based on this OECD report it was determined that the sample for the current study corresponded to data from a one-dimensional scale.

Data Analysis

Secondary data analysis was executed in order to fulfill the objectives of this study, given that the data to work with is secondary data. In other words, data which was gathered by someone other than the user was analyzed.

A proper data analysis for this study requires the use of SPSS 23.0, a reliable software employed for statistical analysis. Through the implementation of Mantel-Haenszel, and Logistic Regression procedures to the datasets in SPSS, the DIF analysis was done. In order to speed up and facilitate the Binary Logistic Regression analysis and the probability graphs making process on SPSS, a total of 8 syntax files –2 for LR and 6 for probability graphs– were created and used (see Appendix B).

Mantel-Haenszel (MH) Method. The MH procedure is considered a categorical non-parametric method and, because it is simple, practical, and successful in detecting uniform DIF, it is one of the most implemented DIF detection methods (Baştuğ, 2016). Nonetheless, this method could render misleading results in terms of non-uniform DIF or when more complex models are used (DeMars, 2009; Güler & Penfield, 2009; Millsap & Everson, 1993; Narayan & Swaminathan, 1996 as cited in Baştuğ, 2016).

Other advantages of the MH method may include “conceptually and computationally easy, consistent, stable in efficiency and robustness, powerful, and well understood based on its statistical properties” (Ramsay, 1993 as cited in Asil, 2005, p. 45).

Holland & Wainer (1993) indicate that the Mantel-Haenszel (MH) procedure allows the measurement of the average difference in difficulty between both analyzed groups –the reference one and the focal one– and a key advantage is that MH is able to detect which of the groups is favored.

The estimation of MH relies on the use of an ability matching variable, the selection of two groups to compare, and a *studied item* –the item that is being examined for presence of DIF in a given analysis– (Holland & Thayer, 1988 as

cited in Wainer & Braun, 1988). The most common measures for matching ability are: total test score, a cluster of questions, or the total score of a test section. Once the matching criterion has been chosen, the groups need to be determined. The *focal group* (F) is of primary interest, in other words, the main group. The *reference group* (R) functions as the standard against which the focal group will be compared. Only dichotomous item responses (1 or 0), obtained by accurately coding right and wrong item answers, should be included in this calculation.

Afterwards, all data for the studied item from both focal and reference groups is organized into several $2 \times 2 \times K$ tables, one such table for each matched set of members of the reference and focal groups. Asil (2005) defines K as the number of binary scored test items and states that the tables are necessary to later calculate the common odds ratio. The data for the j^{th} matched set performance on the studied item is shown in table 3.

Table 3

Data for the j^{th} Matched Set of Members of R and F

		Score on Studied Item		
		1	0	Total
Group	R	A_j	B_j	n_{rj}
	F	C_j	D_j	n_{fj}
	Total	T_{1j}	T_{0j}	T_j

The definitions of the entries in Table 3 are displayed below.

R: Reference group

F: Focal group

1: Correct answer

0: Incorrect answer

T_j : Total number of reference and focal group members

n_{rj} : Number of reference group members

n_{fj} : Number of focal group members

A_j : Number of R members that answered studied item correctly

B_j : Number of R members that answered studied item incorrectly

C_j : Number of F members that answered studied item correctly
 D_j : Number of F members that answered studied item incorrectly
 T_{1j} : Total number of R and F members that answered studied item correctly
 T_{0j} : Total number of R and F members that answered studied item incorrectly

It is important to remember that MH is an estimation of the constant odds ratio that produces a measure of effect size for an assessment of the amount of existing DIF (Gierl, Jodoin, and Ackerman, 2000). Once the tables have been formed, the subsequent step is to compute the MH chi-square test. This calculation is done as follows:

$$\chi^2_{MH} = \frac{\{\sum_j [A_j - E(A_j)] - 0.5\}^2}{\sum_j var(A_j)}$$

where,

$$E(A_j) = \frac{N_{rj} T_{1j}}{T_j}$$

and,

$$var(A_j) = \frac{N_{rj} N_{fj} T_{1j} T_{0j}}{T_j^2 (T_j - 1)}$$

The estimate of the constant odds ratio then is:

$$\alpha_{MH} = \frac{\sum_j A_j D_j / A_j}{\sum_j B_j C_j / T_j}$$

Alpha MH (α_{MH}) is the value reflecting the ratio of the odds that a reference group test taker will answer the studied item correctly compared to the odds for a matched focal group test taker. The rest of the constant odds ratio entries have been described already following the 2 x 2 x K table sample (see Table 3 and corresponding legends).

Afterwards, α_{MH} is transformed to delta MH (Δ_{MH}) to enhance the result interpretability by using this formula,

$$\Delta_{MH} = -(2.35) \ln (\alpha_{MH})$$

When testing for DIF items using MH, in the last stage it is necessary to label results according to the DIF classification index approved by the Educational Testing Service (ETS) researchers. Such index was proposed by Zwick & Ercikan in 1989 and is based on effect size measured by the Δ_{MH} value (Hidalgo & López-Pina, 2004). Zwick & Ercikan's criterion is displayed in table 4.

Table 4

DIF Classification Index for MH

DIF Category	Zwick & Ercikan's Criterion (1989)
No DIF	$\Delta_{MH} 0$
A (Negligible)	χ^2 test is not significant at .05 level or $\Delta_{MH} < 1$
B (Slight to Moderate)	χ^2 test is significant at .05 level and $1 \leq \Delta_{MH} < 1.5$
C (Moderate to Large)	χ^2 test is significant at .05 level and $\Delta_{MH} \geq 1.5$

Logistic Regression (LR) Method. The Logistic Regression (LR) method is a parametric approach –unlike MH– created by Swaminathan and Rogers in 1991, and is able to detect both uniform and non-uniform DIF (Baştuğ, 2016). Zumbo (1999, p. 22) describes the LR method as “based on statistical modeling of the probability of responding correctly to an item by group membership (i.e., reference group and focal group) and a criterion for conditioning variable”.

The LR method has proved to have several advantages such as, automatic categorization of a continuous criterion variable, modeling of both uniform and/or non-uniform DIF, and the possibility of generalizing a binary logistic regression model for ordinal item scores use (Zumbo, 1999).

Jodoin and Gierl (2001), Monahan, McHorney, Stump, and Perkins (2007) are some of the scholars that address the binary LR model in their papers. They indicate that the linear equation of the DIF identification LR model is expressed by:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 XG$$

The equation entries are explained below.

ln: natural logarithm

P: probability of responding correctly

X : measure of overall proficiency (usually total score)

G : dummy variable representing group membership (usually 1=R, 0=F)

XG : interaction term between total score and group membership

β_0 : intercept term

β_1 : weights for the ability term

β_2 : group difference term

β_3 : ability and group interaction term

Gierl, Jodoin, and Ackerman (2000) explain step by step in a clear yet thorough way the calculation of LR. In this case, the presence of DIF is determined by testing the improvement in model fit that occurs when a term for group membership and a term for the interaction between test score and group membership are successively added to the logistic regression model. Following this, a chi-square test is performed to assess the presence of uniform and non-uniform DIF on the studied item. This is done by successively testing each term included in the model. The general model for logistic regression looks like this:

$$P(u=1) = \frac{e^z}{1+e^z}$$

where u is the score on the studied item. Performance on the studied item is first conditioned on total test score. In this step, $z = \beta_0 + \beta_1 X$, where X is the test score (Model 1). This serves as the baseline model. The presence of uniform DIF is then tested by examining the improvement in chi-square model fit associated with adding a term for group membership (G) against the baseline model. That is, Model 2 (i.e., $z = \beta_0 + \beta_1 X + \beta_2 G$) is subtracted from Model 1.

The presence of non-uniform DIF is tested by examining the improvement in chi-square model fit associated with adding a term for group membership (G) and a term for the interaction between test score and group membership (XG) against model 2. In other words, Model 3 (i.e., $z = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 XG$) is subtracted from Model 2. It is worth pointing out that, non-uniform DIF can be evaluated with LR regardless of the results from the uniform DIF test because each model contains different terms.

Subsequently, in order to attempt to reduce the inflated Type I errors often associated with LR, $R^2\Delta$ is computed:

$$R^2\Delta = R^2_2 - R^2_1$$

β_2 : group difference term

β_3 : ability and group interaction term

As part of the LR procedure, two different criteria of DIF classification, using R^2 as a measure of effect size, need to be taken into consideration (Akin Arıkan, Uğurlu & Atar, 2016). There are two available criteria: one from Zumbo & Thomas (1997) and one from Jodoin & Gierl (2001). It is said that Jodoin & Gierl's proposed index is more in harmony with the guidelines approved by the ETS to assess DIF through the MH method. Therefore, this last criterion will be the one used for the labeling of LR results in this study.

Both DIF categorization systems using R^2 and the needed χ^2 test values are presented in table 5 (Gómez-Benito, Hidalgo & Padilla, 2009; Fu & Monfils, 2016).

Table 5

DIF Classification Indexes for LR

DIF Category	Zumbo & Thomas (1997)	Jodoin & Gierl (2001)	χ^2 test value
A (Negligible)	$\Delta R^2 < 0.13$	$\Delta R^2 < .035$	not significant at 0.50 alpha level
B (Slight to Moderate)	$0.13 \leq \Delta R^2 < 0.26$	$.035 \leq \Delta R^2 < .070$	significant at 0.50 alpha level
C (Moderate to Large)	$\Delta R^2 \geq 0.26$	$\Delta R^2 \geq .070$	significant at 0.50 alpha level

Chapter 4

Findings and Interpretation

Interpretation Guidelines for MH and LR Methods

Prior to the presentation of reached findings, it is worth underscoring that MH is a uniform DIF detection method (Atalay Kabasakal, Arsan, Gök, & Kelecioğlu, 2014). Thus it was only possible to calculate uniform DIF through the results rendered by this method.

Regarding the LR method, it is said to be a rich method that makes possible the modeling of uniform and/or non-uniform DIF, and grants the opportunity to incorporate additional explanatory variables in the models, besides ability level and group membership (Fu & Monfils, 2016). Since there is the option to perform only the analysis to detect uniform DIF, and this did not compromise the purpose of this research project, it was decided to only estimate uniform DIF. Therefore, all the LR results that will be shown later on in this chapter, express only uniform DIF.

The MH method easily allows for the determination of which group is favored once the analysis has been carried out. A zero value for ΔMH would indicate that both the reference and focal groups odds for answering a test item correctly are the same, a ΔMH negative value (different to zero) expresses that the item is favoring the reference group, and a ΔMH positive value (different to zero) conveys that the focal group is favored (Holland & Thayer, 1988 as cited in Wright & Oshima, 2015).

Butakor (2015) states that items labeled as A indicate little or no DIF and are suitable for test construction. Items labeled as B evidence moderate DIF and may be used only if no “A” item is available to fill the test’s content required. Items labeled as C indicate large amounts of DIF and may be used exclusively if content experts consider them indispensable to meet test requirements (Clauser & Mazor, 1998 as cited in Butakor, 2015). Careful consideration and implementation of these guidelines are needed when proceeding to further stages of DIF analysis. Next, the findings and interpretation of MH and LR results pertaining to sub-research question 1 are presented.

Findings and Interpretation Related to Sub Research Question 1

Are there any items that indicate culture-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

After the MH method was completed when assessing if DIF caused by culture was present in the cases of Chile and Costa Rica, it was necessary to analyze obtained results, and determine magnitude, significance and direction of DIF.

Table 6

Culture-related DIF Classification for Chile-Costa Rica Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	0.229	0.936	0.155	A	Chile
CS641Q02S	2	20.447*	0.606	1.177	B	Chile
CS641Q03S	3	0.039	1.058	-0.134	A	Costa Rica
CS641Q04S	4	4.265*	0.766	0.627	A	Chile
DS637Q01C	5	2.231	1.195	-0.418	A	Costa Rica
DS637Q05C	6	0.003	1.014	-0.033	A	Costa Rica
CS601Q01S	7	4.108*	2.232	-1.887	C	Costa Rica
CS601Q02S	8	5.165*	1.344	-0.696	A	Costa Rica
CS601Q04S	9	36.145*	2.021	-1.654	C	Costa Rica
DS610Q01C	10	50.436*	0.375	2.303	C	Chile
CS610Q02S	11	0.793	0.817	0.475	A	Chile
CS610Q04S	12	0.158	1.055	-0.127	A	Costa Rica
CS626Q01S	13	45.399*	2.146	-1.795	C	Costa Rica
CS626Q02S	14	4.005*	0.796	0.538	A	Chile
CS626Q03S	15	0.616	1.103	-0.230	A	Costa Rica
DS626Q04C	16	14.162*	0.614	1.147	B	Chile

* Significant at 0.05 alpha level

Table 6 shows that 7 out of the 16 items did not present significant DIF. Consequently, it can be said that out of 16 items, there were 9 items with significant DIF in varied magnitudes –3 A items (4, 8, and 14), 2 B items (2 and

16), and 4 C items (7, 9, 10, and 13)–. In addition, when only the 9 items with significant DIF are considered, it becomes evident that Chilean students had an advantage in 5 items (2, 4, 10, 14, and 16), while Costa Rican students had an advantage in 4 items (7, 8, 9, and 13) according to MH results.

The following table displays MH results corresponding to Chile and Mexico.

Table 7

Culture-related DIF Classification for Chile-Mexico Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	56.949*	0.440	1.929	C	Chile
CS641Q02S	2	2.153	1.179	-0.385	A	Mexico
CS641Q03S	3	1.585	1.292	-0.602	A	Mexico
CS641Q04S	4	1.561	0.853	0.376	A	Chile
DS637Q01C	5	2.544	0.829	0.439	A	Chile
DS637Q05C	6	27.139*	1.849	-1.443	B	Mexico
CS601Q01S	7	11.691*	3.138	-2.688	C	Mexico
CS601Q02S	8	21.195*	1.756	-1.323	B	Mexico
CS601Q04S	9	0.012	0.981	0.045	A	Chile
DS610Q01C	10	0.212	1.064	-0.146	A	Mexico
CS610Q02S	11	14.211*	0.492	1.666	C	Chile
CS610Q04S	12	1.547	0.864	0.343	A	Chile
CS626Q01S	13	10.831*	1.418	-0.820	A	Mexico
CS626Q02S	14	1.513	0.872	0.322	A	Chile
CS626Q03S	15	0.021	0.979	0.052	A	Chile
DS626Q04C	16	1.323	0.869	0.331	A	Chile

* Significant at 0.05 alpha level

In table 7 it is observed that out of 16 items, 6 items exhibited significant DIF and 10 of them did not. Among these 6 items with significant DIF, 1 was labeled as an A item (13), 2 as B items (6 and 8), and 3 as C items (1, 7, and 11). Furthermore, based on the MH results it was detected that out of the 6 items exhibiting significant DIF, 2 items (1 and 11) favored test takers from Chile

whereas 4 items (6, 7, 8, and 13) favored test takers from Mexico. However, it is worth pointing out that out of the 3 flagged C items (1, 7, and 11), 2 items (1 and 11) resulted in favor of Chile while 1 item (7) is shown in favor of Mexico.

In the next table the DIF classification and the directions of DIF values in the cases of Costa Rica and Mexico are presented.

Table 8

Culture-related DIF Classification for Costa Rica-Mexico Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	46.088*	0.494	1.659	C	Costa Rica
CS641Q02S	2	41.576*	1.940	-1.558	C	Mexico
CS641Q03S	3	0.913	1.192	-0.414	A	Mexico
CS641Q04S	4	0.913	1.121	-0.268	A	Mexico
DS637Q01C	5	11.840*	0.668	0.949	A	Costa Rica
DS637Q05C	6	19.481*	1.662	-1.194	B	Mexico
CS601Q01S	7	0.782	1.382	-0.759	A	Mexico
CS601Q02S	8	9.027*	1.437	-0.851	A	Mexico
CS601Q04S	9	45.399*	0.458	1.835	C	Costa Rica
DS610Q01C	10	50.746*	2.720	-2.352	C	Mexico
CS610Q02S	11	11.374*	0.593	1.229	B	Costa Rica
CS610Q04S	12	3.362	0.808	0.501	A	Costa Rica
CS626Q01S	13	13.128*	0.688	0.879	A	Costa Rica
CS626Q02S	14	1.430	1.146	-0.320	A	Mexico
CS626Q03S	15	0.610	0.915	0.207	A	Costa Rica
DS626Q04C	16	9.850*	1.506	-0.961	A	Mexico

* Significant at 0.05 alpha level

Table 8 revealed 10 items with significant DIF degree and 6 without it. Out of these 10 items, 4 were categorized as A items (5, 8, 13, and 16), 2 as B items (6 and 11), and 4 as C items (1, 2, 9, and 10). Also, it was found that out of the 10 items displaying significant DIF, half (1, 5, 9, 11, and 13) were in favor of

examinees from Costa Rica, and the other half (2, 6, 8, 10, and 16) in favor of examinees from Mexico according to MH results.

The results from the MH procedure implemented to test for potential DIF by culture, corresponding to all country pairs, are supplied in the following table.

Table 9

Culture-related DIF Classification for All Country Pairs Based on MH

Item No.	Chile-Costa Rica			Chile-Mexico			Costa Rica-Mexico		
	Δ MH	DIF Level	Favored Group	Δ MH	DIF Level	Favored Group	Δ MH	DIF Level	Favored Group
1	0.155	A		1.929	C	<i>Chile^a</i>	1.659	C	<i>C. Rica</i>
2	1.177	B	<i>Chile</i>	-0.385	A		-1.558	C	<i>Mexico</i>
3	-0.134	A		-0.602	A		-0.414	A	
4	0.627	A	<i>Chile</i>	0.376	A		-0.268	A	
5	-0.418	A		0.439	A		0.949	A	<i>C. Rica</i>
6	-0.033	A		-1.443	B	<i>Mexico</i>	-1.194	B	<i>Mexico</i>
7	-1.887	C	<i>C. Rica</i>	-2.688	C	<i>Mexico</i>	-0.759	A	
8	-0.696	A	<i>C. Rica</i>	-1.323	B	<i>Mexico</i>	-0.851	A	<i>Mexico</i>
9	-1.654	C	<i>C. Rica</i>	0.045	A		1.835	C	<i>C. Rica</i>
10	2.303	C	<i>Chile</i>	-0.146	A		-2.352	C	<i>Mexico</i>
11	0.475	A		1.666	C	<i>Chile</i>	1.229	B	<i>C. Rica</i>
12	-0.127	A		0.343	A		0.501	A	
13	-1.795	C	<i>C. Rica</i>	-0.820	A	<i>Mexico</i>	0.879	A	<i>C. Rica</i>
14	0.538	A	<i>Chile</i>	0.322	A		-0.320	A	
15	-0.230	A		0.052	A		0.207	A	
16	1.147	B	<i>Chile</i>	0.331	A		-0.961	A	<i>Mexico</i>

a; Only in the case of significant DIF degree, the favored group was displayed.

Table 9 shows that according to MH, out of 16 items, 3 (3, 12, and 15) did not present a significant degree of culture-related DIF for any of the analyzed country pairs, and that 2 items were flagged in all three cases –item 8 twice as an A item and once as a B item, then item 13 twice as an A item and once as a C

item—. It also shows that nearly the same test items are flagged across country pairs.

Once all the culture-related DIF classifications for all country pairs based on MH results had been completed, the LR results in the cases of Chile and Costa Rica were examined to determine effect size and whether DIF was significant or not.

Table 10

Culture-related DIF Classification for Chile-Costa Rica Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	0.182	0.000	No DIF
CS641Q02S	2	22.594*	0.014	A
CS641Q03S	3	0.100	0.000	No DIF
CS641Q04S	4	4.615*	0.003	A
DS637Q01C	5	2.283	0.001	A
DS637Q05C	6	0.001	0.000	No DIF
CS601Q01S	7	3.428	0.008	A
CS601Q02S	8	5.592*	0.003	A
CS601Q04S	9	37.108*	0.022	A
DS610Q01C	10	52.585*	0.033	A
CS610Q02S	11	1.190	0.001	A
CS610Q04S	12	0.223	0.000	No DIF
CS626Q01S	13	47.176*	0.031	A
CS626Q02S	14	4.870*	0.003	A
CS626Q03S	15	0.511	0.000	No DIF
DS626Q04C	16	13.536*	0.007	A

* Significant at 0.05 alpha level

Table 10 identifies 5 no DIF items (1, 3, 6, 12, and 15) according to the LR analysis results, meaning that these items do not operate differently between the two analyzed countries (Chile and Costa Rica). Moreover, 8 out of 16 items (2, 4, 8, 9, 10, 13, 14, and 16) show a significant amount of DIF and all those 8 items are

identified as A items. Interestingly enough, this method does not render any items labeled with B or C level of DIF for Chile and Costa Rica.

The next table displays the DIF classification and the significance of DIF values according to the LR method in the cases of Chile and Mexico.

Table 11

Culture-related DIF Classification for Chile-Mexico Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	57.915*	0.033	A
CS641Q02S	2	1.739	0.001	A
CS641Q03S	3	1.731	0.002	A
CS641Q04S	4	1.473	0.001	A
DS637Q01C	5	2.911	0.002	A
DS637Q05C	6	28.177*	0.016	A
CS601Q01S	7	10.966*	0.021	A
CS601Q02S	8	21.946*	0.012	A
CS601Q04S	9	0.074	0.000	No DIF
DS610Q01C	10	0.34	0.001	A
CS610Q02S	11	15.306*	0.013	A
CS610Q04S	12	1.779	0.001	A
CS626Q01S	13	12.139*	0.007	A
CS626Q02S	14	1.697	0.001	A
CS626Q03S	15	0.096	0.000	No DIF
DS626Q04C	16	1.227	0.000	No DIF

* Significant at 0.05 alpha level

In table 11 it is evidenced that 6 items –1, 6, 7, 8, 11, and 13– exhibit a significant degree of DIF, yet are labeled with the slightest DIF magnitude (A) when the performance of students from Chile and Mexico was examined. Additionally, 3 of the 16 items present no DIF in this case. Similar to the findings in the previous table –table 9–, it is noticed that the LR procedure did not flag any DIF items with B level or C level in the cases of Chile and Mexico.

In the following table the LR results from the last country pair –Costa Rica and Mexico– are shown.

Table 12

Culture-related DIF Classification for Costa Rica-Mexico Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	47.455*	0.029	A
CS641Q02S	2	43.557*	0.027	A
CS641Q03S	3	0.974	0.001	A
CS641Q04S	4	0.962	0.001	A
DS637Q01C	5	12.408*	0.007	A
DS637Q05C	6	20.672*	0.014	A
CS601Q01S	7	1.251	0.003	A
CS601Q02S	8	9.103*	0.005	A
CS601Q04S	9	46.119*	0.026	A
DS610Q01C	10	54.426*	0.042	B
CS610Q02S	11	11.739*	0.010	A
CS610Q04S	12	3.562	0.002	A
CS626Q01S	13	12.951*	0.008	A
CS626Q02S	14	1.772	0.002	A
CS626Q03S	15	0.615	0.001	A
DS626Q04C	16	10.022*	0.006	A

* Significant at 0.05 alpha level

It is seen in table 12 that a very large number of items (10 items: 1, 2, 5, 6, 8, 9, 10, 11, 13, and 16) with significant DIF degree were detected. Nevertheless, the vast majority of these 10 DIF items are identified as A items. This country pair stands out as the one with the highest amount of significant DIF-labeled items and zero no-DIF items. It is important to point out that item 10 was the only one labeled as a B item when Costa Rican and Mexican students' performance was considered.

The table below provides the results from the LR procedure implemented to test for potential DIF by culture, corresponding to all country pairs.

Table 13

Culture-related DIF Classification for All Country Pairs Based on LR

Item No.	Chile-Costa Rica		Chile-Mexico		Costa Rica-Mexico	
	ΔR^2	DIF Level	ΔR^2	DIF Level	ΔR^2	DIF Level
1	0.000	No DIF	0.033	A	0.029	A
2	0.014	A	0.001	A	0.027	A
3	0.000	No DIF	0.002	A	0.001	A
4	0.003	A	0.001	A	0.001	A
5	0.001	A	0.002	A	0.007	A
6	0.000	No DIF	0.016	A	0.014	A
7	0.008	A	0.021	A	0.003	A
8	0.003	A	0.012	A	0.005	A
9	0.022	A	0.000	No DIF	0.026	A
10	0.033	A	0.001	A	0.042	B
11	0.001	A	0.013	A	0.010	A
12	0.000	No DIF	0.001	A	0.002	A
13	0.031	A	0.007	A	0.008	A
14	0.003	A	0.001	A	0.002	A
15	0.000	No DIF	0.000	No DIF	0.001	A
16	0.007	A	0.000	No DIF	0.006	A

Table 13 shows that based on LR results, 3 out of 16 items –3, 12, and 15– did not exhibit a significant degree of DIF by culture when the three country pairs were taken into account. The table also shows that items 8 and 13 were categorized as A items with significant DIF degree in all three cases. Moreover, 4 items –4, 5, 7, and 14– are only flagged by one of the country pairs as A items with significant DIF degree. This table makes possible to notice that only very few times the same items with significant DIF are identified across all country pairs.

The upcoming table sums up the results of both the MH and LR methods when 16 dichotomous science items from PISA 2015 were tested for potential culture-related DIF in the cases of Chile, Costa Rica, and Mexico.

Table 14

Summary of the Culture-related DIF Detection MH and LR Analyses

Item No.	Chile-Costa Rica				Chile-Mexico				Costa Rica-Mexico			
	Δ_{MH}	DIF Level	ΔR^2	DIF Level	Δ_{MH}	DIF Level	ΔR^2	DIF Level	Δ_{MH}	DIF Level	ΔR^2	DIF Level
1	0.155		0.000		1.929	C ^a	0.033	A	1.659	C	0.029	A
2	1.177	B	0.014	A	-0.385		0.001		-1.558	C	0.027	A
3	-0.134		0.000		-0.602		0.002		-0.414		0.001	
4	0.627	A	0.003	A	0.376		0.001		-0.268		0.001	
5	-0.418		0.001		0.439		0.002		0.949	A	0.007	A
6	-0.033		0.000		-1.443	B	0.016	A	-1.194	B	0.014	A
7	-1.887	C	0.008		-2.688	C	0.021	A	-0.759		0.003	
8	-0.696	A	0.003	A	-1.323	B	0.012	A	-0.851	A	0.005	A
9	-1.654	C	0.022	A	0.045		0.000		1.835	C	0.026	A
10	2.303	C	0.033	A	-0.146		0.001		-2.352	C	0.042	B
11	0.475		0.001		1.666	C	0.013	A	1.229	B	0.010	A
12	-0.127		0.000		0.343		0.001		0.501		0.002	
13	-1.795	C	0.031	A	-0.820	A	0.007	A	0.879	A	0.008	A
14	0.538	A	0.003	A	0.322		0.001		-0.320		0.002	
15	-0.230		0.000		0.052		0.000		0.207		0.001	
16	1.147	B	0.007	A	0.331		0.000		-0.961	A	0.006	A

a; Only in the case of significant DIF degree, the DIF level is specified.

Table 14 revealed consistency in the MH and LR results when it comes to DIF significance but not when the DIF magnitude (A, B or C) is considered. This summary table manifested that item 7 is the most inconsistent across methods.

As part of the DIF analysis, probability of accurate answer graphs were created to gain additional insight (see Appendix A). Figures 3 and 4 reveal that indeed, when Chile and Costa Rica were assessed, items 10 and 13 proved to

have a C level of uniform DIF. While item 10 shows Chile had an advantage over Costa Rica, item 13 favored Costa Rican students. It is important to highlight that according to Table 14, the MH procedure accurately flagged items 10 and 13 as C level DIF items in the cases of Chile and Costa Rica, but the LR method labeled the same items as A level DIF items.

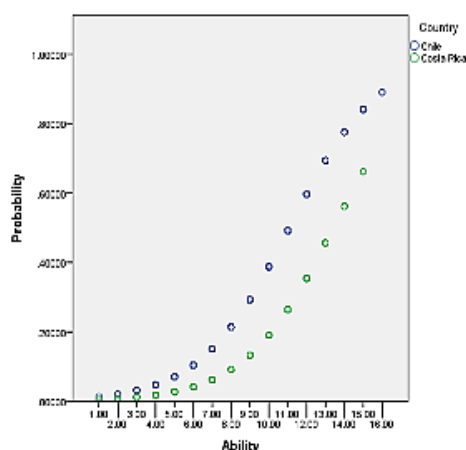


Figure 3. Probability of correct answer graph for item 10 (DS610Q01C) in the cases of Chile and Costa Rica

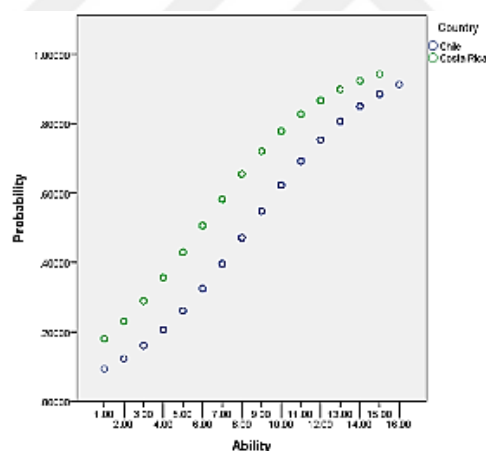


Figure 4. Probability of correct answer graph for item 13 (CS626Q01S) in the cases of Chile and Costa Rica

The following probability of accurate answer graphs –figures 5 and 6– confirm that in the case of Chile and Mexico, items 1 and 7 presented a C level of uniform DIF. Item 1 is shown to favor Chilean students over Mexican ones, whereas item 7 favored Mexican students. Interestingly enough and similar to the findings of the previous two graphs –figures 3 and 4–, the graphs below make evident the accuracy of the MH results for Chile and Mexico since items 1 and 7

were identified as C level DIF items, although the LR results identified those same items as A level DIF items (See Table 14).

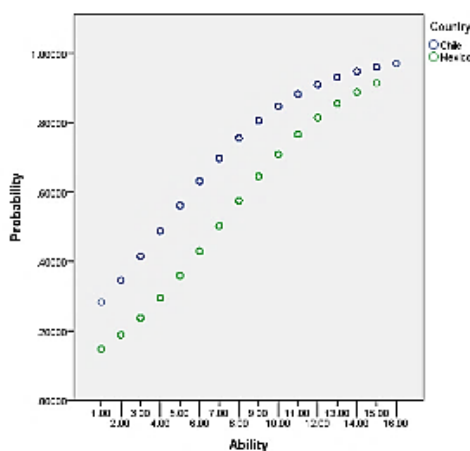


Figure 5. Probability of correct answer graph for item 1 (CS641Q01S) in the cases of Chile and Mexico

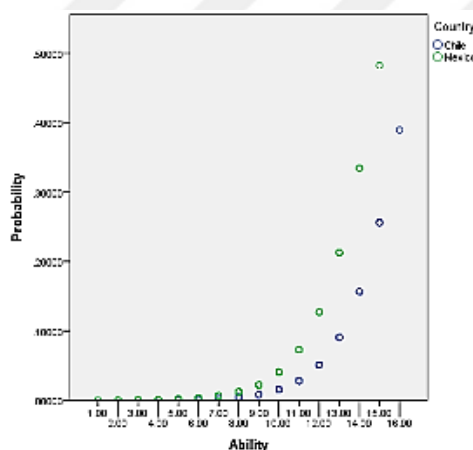


Figure 6. Probability of correct answer graph for item 7 (CS601Q01S) in the cases of Chile and Mexico

Additionally, in terms of item difficulty, it can be said that item 7 was the most difficult one from the set of 16 items for Chilean and Mexican students since in the graph it is observed that both curves remain adjacent to the ability axis when the ability level is low. It is then made evident that students with lower levels of ability from both analyzed countries were not able to perform well in this item.

Figures 7 and 8 –probability of accurate answer graphs– show that when the performance from Costa Rican and Mexican students was analyzed, items 2 and 9 exhibited a C level of uniform DIF. The reference group (Mexican test takers) had an advantage over their counterparts in item 2 while the focal group (Costa Rican test takers) was the advantaged group in item 9. Once again, the

graphs reveal the lack of consistency across DIF detection methods as the MH method rightly flagged items 2 and 9 with C level DIF while the LR method flagged these items with A level DIF (See Table 14). When all country pairs are considered, the graphs –figures 3, 4, 5, 6, 7, and 8– have confirmed a C level of DIF present among these items as the gap in between the graph curves is significant.

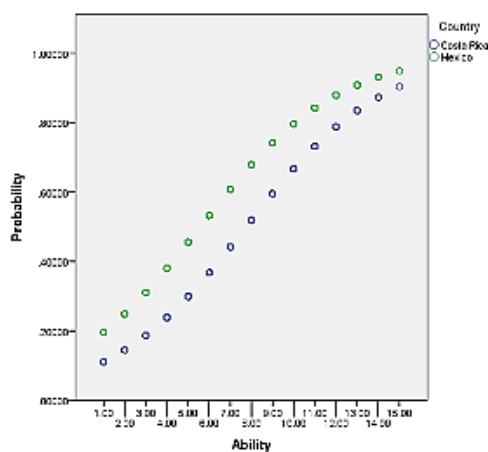


Figure 7. Probability of correct answer graph for item 2 (CS641Q02S) in the cases of Costa Rica and Mexico

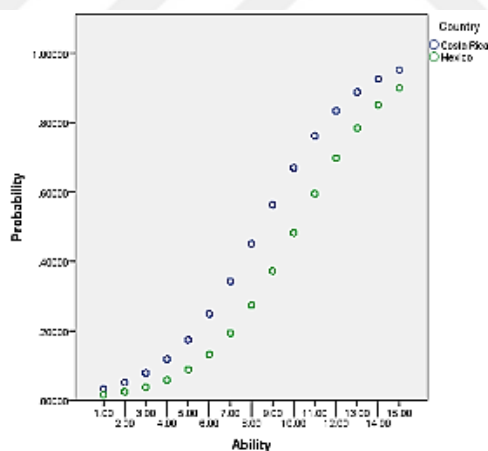


Figure 8. Probability of correct answer graph for item 9 (CS601Q04S) in the cases of Costa Rica and Mexico

After the variable of culture had been assessed through MH and LR procedures, the same analyses were carried out for the variable of gender with the corresponding data. Next, the findings and interpretation of MH and LR results pertaining to sub research question 2 are presented.

Findings and Interpretation Related to Sub Research Question 2

Are there any items that indicate gender-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

Following the completion of the MH analysis, executed to test for possible DIF caused by gender considering the students from Chile, the classification of DIF by magnitude, DIF degree significance, and DIF direction was done.

Table 15

Gender-related DIF Classification for Chile Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	26.750*	2.557	-2.207	C	Males
CS641Q02S	2	0.250	1.094	-0.212	A	Males
CS641Q03S	3	0.075	1.127	-0.282	A	Males
CS641Q04S	4	1.869	1.306	-0.627	A	Males
DS637Q01C	5	6.751*	0.668	0.949	A	Females
DS637Q05C	6	0.530	0.879	0.303	A	Females
CS601Q01S	7	1.124	1.745	-1.309	B	Males
CS601Q02S	8	1.249	0.811	0.491	A	Females
CS601Q04S	9	5.782*	0.689	0.877	A	Females
DS610Q01C	10	2.020	1.268	-0.557	A	Males
CS610Q02S	11	0.126	0.858	0.362	A	Females
CS610Q04S	12	0.000	0.987	0.031	A	Females
CS626Q01S	13	3.192	1.299	-0.616	A	Males
CS626Q02S	14	5.077*	0.709	0.808	A	Females
CS626Q03S	15	2.116	1.278	-0.576	A	Males
DS626Q04C	16	4.691*	0.698	0.846	A	Females

* Significant at 0.05 alpha level

Table 15 displays that out of 16 items, 11 items did not exhibit significant DIF. From the 5 items with significant DIF, 4 were labeled as A items (5, 9, 14 and 16) and 1 as C item (1). Furthermore, based on the MH results it was detected that

out of those 5 items, the majority (4 items: 5, 9, 14 and 16) favored female examinees from Chile whereas just 1 item (1) favored male examinees.

Also the MH results belonging to Costa Rica were subjected to analysis, and the resultant DIF classification with directions of DIF values is provided below.

Table 16

Gender-related DIF Classification for Costa Rica Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	37.263*	2.863	-2.472	C	Males
CS641Q02S	2	2.045	1.252	-0.529	A	Males
CS641Q03S	3	0.001	0.962	0.089	A	Females
CS641Q04S	4	0.625	1.158	-0.343	A	Males
DS637Q01C	5	1.663	0.798	0.531	A	Females
DS637Q05C	6	4.480*	0.680	0.907	A	Females
CS601Q01S	7	3.617	3.181	-2.719	C	Males
CS601Q02S	8	4.490*	0.671	0.938	A	Females
CS601Q04S	9	2.032	0.78	0.583	A	Females
DS610Q01C	10	0.529	0.823	0.456	A	Females
CS610Q02S	11	1.271	0.737	0.717	A	Females
CS610Q04S	12	0.783	1.174	-0.376	A	Males
CS626Q01S	13	7.642*	1.554	-1.036	B	Males
CS626Q02S	14	1.658	0.796	0.536	A	Females
CS626Q03S	15	0.088	0.941	0.143	A	Females
DS626Q04C	16	13.653*	0.452	1.868	C	Females

* Significant at 0.05 alpha level

Table 16 shows that 11 out of 16 items did not present significant DIF. Consequently, it can be said that out of 16 items, there were 5 items with significant DIF in varied magnitudes –2 A items (6 and 8), 1 B item (13), and 2 C items (1 and 16)–. In addition, when only the 5 items with significant DIF are considered, it becomes evident that Costa Rican female students had an

advantage in 3 items (6, 8, and 16), while the male students had an advantage in 2 items (1 and 13) according to MH results.

The following table displays MH results corresponding to Mexico.

Table 17

Gender-related DIF Classification for Mexico Based on MH

Item Code	Item No.	MH Statistics	α MH	Δ MH	Effect Size	Favored Group
CS641Q01S	1	9.992*	1.538	-1.011	B	Males
CS641Q02S	2	2.950	1.277	-0.576	A	Males
CS641Q03S	3	0.942	0.785	0.569	A	Females
CS641Q04S	4	1.589	1.224	-0.475	A	Males
DS637Q01C	5	2.532	0.757	0.653	A	Females
DS637Q05C	6	0.439	0.895	0.261	A	Females
CS601Q01S	7	3.510	2.437	-2.094	C	Males
CS601Q02S	8	6.113*	0.673	0.931	A	Females
CS601Q04S	9	2.804	0.748	0.684	A	Females
DS610Q01C	10	0.528	0.873	0.320	A	Females
CS610Q02S	11	0.082	1.069	-0.157	A	Males
CS610Q04S	12	0.004	1.022	-0.049	A	Males
CS626Q01S	13	9.380*	1.529	-0.996	A	Males
CS626Q02S	14	3.237	0.761	0.642	A	Females
CS626Q03S	15	0.001	0.986	0.035	A	Females
DS626Q04C	16	3.326	0.735	0.724	A	Females

* Significant at 0.05 alpha level

Table 17 identified 13 items that did not exhibit significant DIF degree. Out of the 3 items with significant degree of DIF, 2 were categorized as A items (8 and 13), and 1 as B item (1). Additionally, it was discovered that out of these 3 items, 2 (1 and 13) were in favor of male test takers from Mexico, and only 1 (8) in favor of the female Mexican test takers.

The results from the MH procedure implemented to test for potential DIF by gender corresponding to all three countries are supplied in the upcoming table.

Table 18

Gender-related DIF Classification for All Countries Based on MH

Item No.	Chile			Costa Rica			Mexico		
	Δ MH	DIF Level	Favored Group	Δ MH	DIF Level	Favored Group	Δ MH	DIF Level	Favored Group
1	-2.207	C	<i>Males^a</i>	-2.472	C	<i>Males</i>	-1.011	B	<i>Males</i>
2	-0.212	A		-0.529	A		-0.576	A	
3	-0.282	A		0.089	A		0.569	A	
4	-0.627	A		-0.343	A		-0.475	A	
5	0.949	A	<i>Females</i>	0.531	A		0.653	A	
6	0.303	A		0.907	A	<i>Females</i>	0.261	A	
7	-1.309	B		-2.719	C		-2.094	C	
8	0.491	A		0.938	A	<i>Females</i>	0.931	A	<i>Females</i>
9	0.877	A	<i>Females</i>	0.583	A		0.684	A	
10	-0.557	A		0.456	A		0.320	A	
11	0.362	A		0.717	A		-0.157	A	
12	0.031	A		-0.376	A		-0.049	A	
13	-0.616	A		-1.036	B	<i>Males</i>	-0.996	A	<i>Males</i>
14	0.808	A	<i>Females</i>	0.536	A		0.642	A	
15	-0.576	A		0.143	A		0.035	A	
16	0.846	A		1.868	C	<i>Females</i>	0.724	A	

a; Only in the case of significant DIF degree, the favored group was displayed.

Table 18 made evident that according to MH, out of 16 items, half (8 items: 2, 3, 4, 7, 10, 11, 12, and 15) did not present a significant degree of gender-related DIF for any of the analyzed countries. From the 8 items showing significant DIF degree, just 1 item (1) was flagged in all three countries –as a C item favoring males from Chile and Costa Rica, and a B item favoring Mexican males–. Also, it is evident that 4 items (5, 6, 9, and 14) from the PISA 2015 science section are flagged with gender-related DIF inconsistently, as it appears that those items functioned unequally only in one of the countries.

Following the analysis of gender-related DIF values based on MH, the counterpart method's results needed to be examined. After finishing the LR procedure to determine if DIF caused by gender was present in a portion of the PISA 2015 science section taken by students from Chile, the next step was to determine effect size and significance of DIF. Results are seen in Table 19.

Table 19

Gender-related DIF Classification for Chile Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	27.794*	0.039	B
CS641Q02S	2	0.317	0.000	No DIF
CS641Q03S	3	0.047	0.000	No DIF
CS641Q04S	4	1.969	0.002	A
DS637Q01C	5	6.787*	0.008	A
DS637Q05C	6	0.409	0.000	No DIF
CS601Q01S	7	2.004	0.007	A
CS601Q02S	8	1.090	0.001	A
CS601Q04S	9	6.062*	0.006	A
DS610Q01C	10	2.787	0.003	A
CS610Q02S	11	0.400	0.001	A
CS610Q04S	12	0.023	0.000	No DIF
CS626Q01S	13	3.549	0.004	A
CS626Q02S	14	4.392*	0.005	A
CS626Q03S	15	2.456	0.002	A
DS626Q04C	16	5.377*	0.005	A

* Significant at 0.05 alpha level

Table 19 shows that 5 items (1, 5, 9, 14, and 16) were classified with significant DIF degree and marked with two different DIF magnitudes –4 A items and 1 B item–. 4 out of the 16 items (2, 3, 6, and 12) were found without DIF due to a ΔR^2 value of zero. And it is worth pointing out that only 1 item (1) was marked with B level of gender-related DIF when Chilean examinees' performance was considered.

The following table displays the DIF classification and the DIF significance of gender-related DIF according to the LR method in the case of Costa Rican test takers.

Table 20

Gender-related DIF Classification for Costa Rica Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	39.653*	0.055	B
CS641Q02S	2	2.538	0.003	A
CS641Q03S	3	0.148	0.000	No DIF
CS641Q04S	4	0.513	0.000	No DIF
DS637Q01C	5	2.361	0.004	A
DS637Q05C	6	5.188*	0.008	A
CS601Q01S	7	3.914*	0.024	A
CS601Q02S	8	4.815*	0.005	A
CS601Q04S	9	2.657	0.003	A
DS610Q01C	10	0.665	0.001	A
CS610Q02S	11	2.055	0.005	A
CS610Q04S	12	0.808	0.001	A
CS626Q01S	13	8.844*	0.012	A
CS626Q02S	14	1.903	0.002	A
CS626Q03S	15	0.024	0.000	No DIF
DS626Q04C	16	14.617*	0.019	A

* Significant at 0.05 alpha level

Table 20 identifies 3 no DIF items –3, 4, and 15– according to the LR analysis results, meaning that these items function accurately both for the focal group (females) and reference group (males) from Costa Rica. Moreover, 6 out of 16 items –1, 6, 7, 8, 13, and 16– show a significant amount of DIF. From these 6 items, 5 display an A level of female/male-related DIF while just 1 displays a B-level (with a high ΔR^2 value of 0.055) of DIF.

The upcoming table provides the LR results from the Mexican students who answered the PISA 2015 science items taken into consideration in this investigation.

Table 21

Gender-related DIF Classification for Mexico Based on LR

Item Code	Item No.	Δ Chi-Square	ΔR^2	Effect Size
CS641Q01S	1	10.273*	0.012	A
CS641Q02S	2	3.029	0.004	A
CS641Q03S	3	0.970	0.002	A
CS641Q04S	4	1.632	0.001	A
DS637Q01C	5	2.665	0.003	A
DS637Q05C	6	0.589	0.001	A
CS601Q01S	7	4.935*	0.021	A
CS601Q02S	8	6.908*	0.007	A
CS601Q04S	9	2.638	0.002	A
DS610Q01C	10	0.766	0.001	A
CS610Q02S	11	0.080	0.000	No DIF
CS610Q04S	12	0.031	0.000	No DIF
CS626Q01S	13	9.032*	0.010	A
CS626Q02S	14	3.628	0.005	A
CS626Q03S	15	0.004	0.000	No DIF
DS626Q04C	16	3.923*	0.005	A

* Significant at 0.05 alpha level

In table 21 it is evidenced that 5 items –1, 7, 8, 13, and 16– exhibit significant DIF degree yet are flagged with the lowest DIF magnitude (A) when the performance of test takers from Chile and Mexico was examined considering the gender variable. Additionally, 3 of the 16 items –11, 12, and 15– present no DIF in this case.

The table below provides the results from the LR procedure implemented to test for potential female/male-related DIF corresponding to all three countries.

Table 22

Gender-related DIF Classification for All Countries Based on LR

Item No.	Chile		Costa Rica		Mexico	
	ΔR^2	DIF Level	ΔR^2	DIF Level	ΔR^2	DIF Level
1	0.039	B	0.055	B	0.012	A
2	0.000	No DIF	0.003	A	0.004	A
3	0.000	No DIF	0.000	No DIF	0.002	A
4	0.002	A	0.000	No DIF	0.001	A
5	0.008	A	0.004	A	0.003	A
6	0.000	No DIF	0.008	A	0.001	A
7	0.007	A	0.024	A	0.021	A
8	0.001	A	0.005	A	0.007	A
9	0.006	A	0.003	A	0.002	A
10	0.003	A	0.001	A	0.001	A
11	0.001	A	0.005	A	0.000	No DIF
12	0.000	No DIF	0.001	A	0.000	No DIF
13	0.004	A	0.012	A	0.010	A
14	0.005	A	0.002	A	0.005	A
15	0.002	A	0.000	No DIF	0.000	No DIF
16	0.005	A	0.019	A	0.005	A

In table 22, 7 out of 16 items –2, 3, 4, 10, 11, 12, and 15– did not exhibit significant DIF degree by gender as examinees from the three countries were considered. Out of the 9 items with significant DIF degree, 2 items –1 and 16– were marked consistently across countries, showing that these items performed one-sidedly for females and males of Chile, Costa Rica, and Mexico. Also, out of the 9 items with significant DIF degree, 4 items –5, 6, 9, and 14– are singled out as items flagged with significant gender-related DIF by only one of the countries.

The next table reviews the results of MH and LR methods after 16 dichotomous PISA 2015 science items were tested for potential gender-related DIF in the cases of Chile, Costa Rica, and Mexico.

Table 23

Summary of the Gender-related DIF Detection MH and LR Analyses

Item No.	Chile				Costa Rica				Mexico			
	Δ_{MH}	DIF Level	ΔR^2	DIF Level	Δ_{MH}	DIF Level	ΔR^2	DIF Level	Δ_{MH}	DIF Level	ΔR^2	DIF Level
1	-2.207	C ^a	0.039	B	-2.472	C	0.055	B	-1.011	B	0.012	A
2	-0.212		0.000		-0.529		0.003		-0.576		0.004	
3	-0.282		0.000		0.089		0.000		0.569		0.002	
4	-0.627		0.002		-0.343		0.000		-0.475		0.001	
5	0.949	A	0.008	A	0.531		0.004		0.653		0.003	
6	0.303		0.000		0.907	A	0.008	A	0.261		0.001	
7	-1.309		0.007		-2.719		0.024	A	-2.094		0.021	A
8	0.491		0.001		0.938	A	0.005	A	0.931	A	0.007	A
9	0.877	A	0.006	A	0.583		0.003		0.684		0.002	
10	-0.557		0.003		0.456		0.001		0.320		0.001	
11	0.362		0.001		0.717		0.005		-0.157		0.000	
12	0.031		0.000		-0.376		0.001		-0.049		0.000	
13	-0.616		0.004		-1.036	B	0.012	A	-0.996	A	0.010	A
14	0.808	A	0.005	A	0.536		0.002		0.642		0.005	
15	-0.576		0.002		0.143		0.000		0.035		0.000	
16	0.846	A	0.005	A	1.868	C	0.019	A	0.724		0.005	A

a; Only in the case of significant DIF degree, the DIF level is specified.

Table 23 revealed an overall consistency in the MH and LR results if the DIF significance is taken into account but not when the DIF magnitude (A, B or C) is considered. This table manifested that item 7 was flagged inconsistently across DIF detection methods yet item 1 was flagged for the three countries by both analysis procedures.

In order to gain additional insight, probability of accurate answer graphs were created (see Appendix A). Figure 9 reveals that indeed, when Chile's examinees were assessed, item 1 proved to have a moderate to large level of uniform DIF and favored the Chilean males group. It is important to highlight that according to Table 23, there was inconsistency across methods for item 1 because the MH procedure flagged it as a C level DIF item, but the LR method labeled the same item as a B level DIF item.

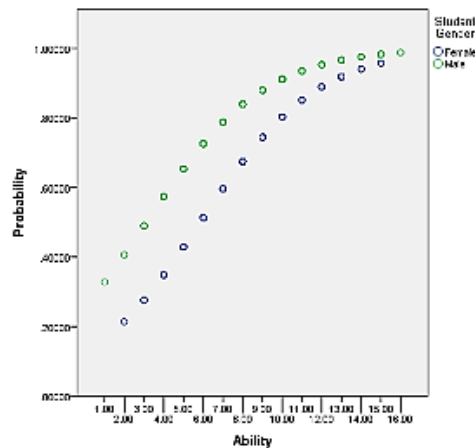


Figure 9. Probability of correct answer graph for item 1 (CS641Q01S) in the case of Chile

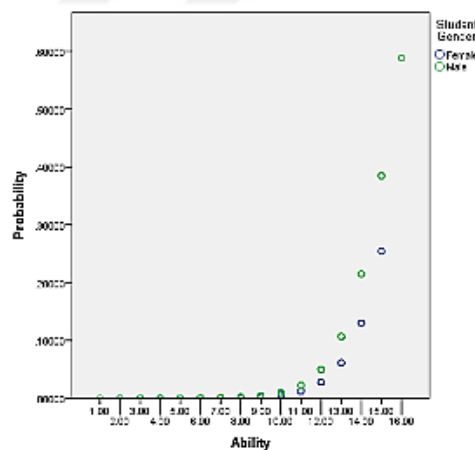


Figure 10. Probability of correct answer graph for item 7 (CS601Q01S) in the case of Chile

Furthermore, item 7 was found to be the most difficult one from the set of 16 items for Chilean students since in the graph it is observed that both curves remain adjacent to the ability axis when the ability level is low. It is then evident that students from both genders with lower levels of ability did not perform well in this item, and that males from Chile were slightly advantaged. An interesting fact is

that, out of the 16 assessed items, item 7 was consistently the most difficult one across variables –culture and gender– and DIF detection methods.

The following probability of accurate answer graphs –figures 11 and 12– confirm that in the case of Costa Rica, items 1 and 16 presented a moderate to large level of uniform DIF. Item 1 is shown to favor Costa Rican male students over female ones, whereas item 16 favored female students. The graphs below make evident the discrepancy between the MH and LR results for items 1 and 16, given that the MH method identified both as C level DIF items, while the LR method identified item 1 as a B level DIF item and item 16 as an A level DIF item (See Table 23). Additionally, item 1 seems to be rightly labeled as a C item by the MH procedure, however, item 16 appears to prove right the label given by the LR method –an A level DIF item–.

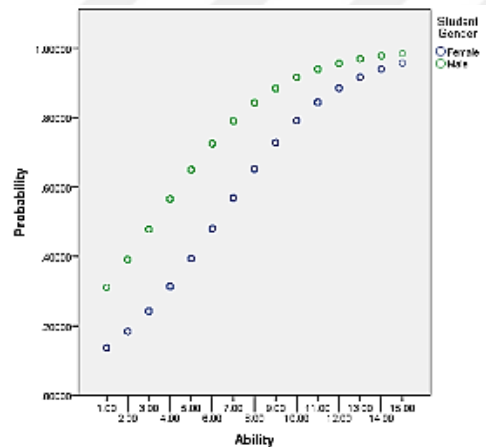


Figure 11. Probability of correct answer graph for item 1 (CS641Q01S) in the case of Costa Rica

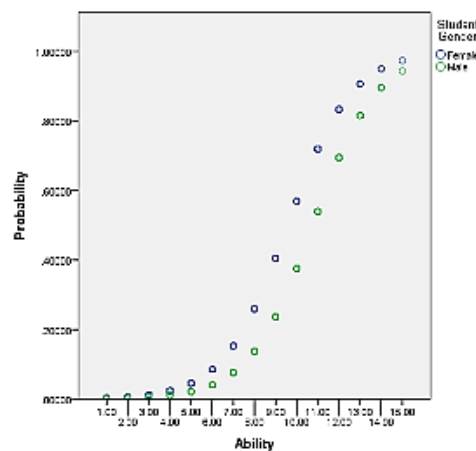


Figure 12. Probability of correct answer graph for item 16 (DS626Q04C) in the case of Costa Rica

Figures 13 and 14 –probability of accurate answer graphs– show that when the performance from Mexican female and male students was analyzed, items 8 and 13 exhibited a negligible level of uniform DIF. The focal group (Mexican female test takers) had an advantage over their counterparts in item 8 while the reference group (Mexican male test takers) was the advantaged group in item 13. This time, the graphs reveal consistency across DIF detection methods, as both MH and LR methods singled out items 8 and 13 as items with an A level of DIF (See Table 23). The graphs found below also confirm that the type of DIF found by both procedures is uniform DIF. When the graph curves are carefully observed, it is noticeable that the difference of probability of success between groups is mostly consistent across all levels of ability.

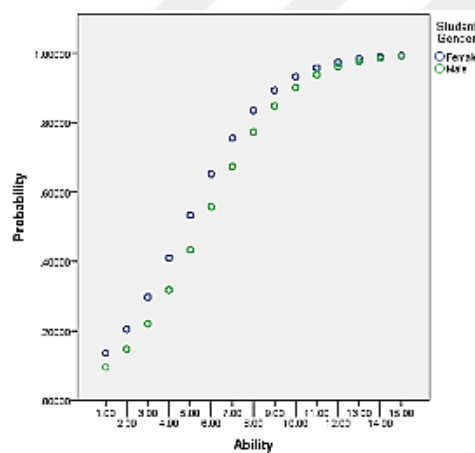


Figure 13. Probability of correct answer graph for item 8 (CS601Q02S) in the case of Mexico

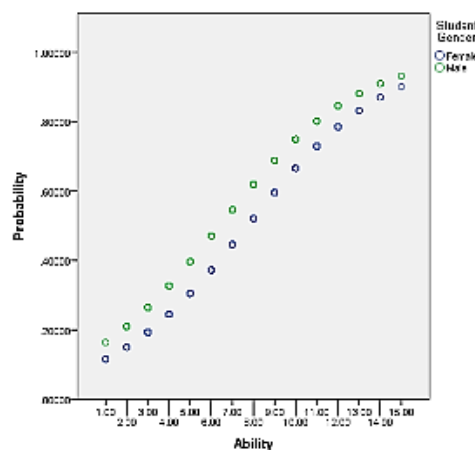


Figure 14. Probability of correct answer graph for item 13 (CS626Q01S) in the case of Mexico

Overall Findings

Reviewing all the findings as a whole, it is noticed that out of 16 items, this study flagged a total of 13 items as DIF items –with varying magnitudes of DIF– when the variable of culture was examined. However, a smaller amount of DIF items –9 out of 16– was labeled with gender-related DIF. These findings correspond with the ones from similar DIF PISA studies that have found test items functioning more equally across genders than cultures.

For example, Demir and Köse's 2014 study detected 5 DIF items caused by gender among the PISA 2009 mathematics items, whereas DIF caused by culture was present in 14-18 items, depending on the country pair –Turkey and Germany, Turkey and Finland, Turkey and United States– that was analyzed. A research done by Le (2006) to determine if DIF was present among PISA 2006 field trial science items, also revealed a lower DIF caused by female/male differences than DIF caused by cultural and linguistic differences –country of testing or test language–.

Fellow educational assessment researchers have performed DIF analyses with a focus on the cross-cultural and cross-lingual equivalence of a set of constructs. When they assessed the presence of DIF in PISA 2006 science items and 2006 student questionnaire caused by cultural and linguistic differences, they arrived to the conclusion that there is a relationship between the increase of cultural and linguistic differences across countries and the increase of DIF items detected (Uzun & Gelbal, 2017; Asil & Gelbal, 2012).

The literature review done for the present research identified several DIF analyses on PISA studies that provided useful insight, nevertheless, it is crucial to highlight that there were not any PISA studies analyzing DIF among Latin American countries. This points out the lack of literature in the field of educational assessment including data from this region of the world.

Chapter 5

Results and Recommendations

As it was specified earlier, it is the aim of this research study to determine if culture-related and gender-related DIF is present among a set of 16 binary-scored science items from PISA 2015 taken by students from Chile, Costa Rica, and Mexico.

The Mantel-Haenszel (MH) and Logistic Regression (LR) methods were implemented to identify Differential Item Functioning (DIF) and answer the two main questions conveyed in sub research questions 1 and 2.

Results Related to Sub Research Question 1

Are there any items that indicate culture-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

After completing the analysis to detect DIF caused by culture in the cases of Chile and Costa Rica, 9 DIF items were marked in the MH procedure –3 A items, 2 B items, 4 C items– whereas the LR procedure marked 8 DIF items –8 A items–. 5 of the 9 items exhibiting DIF favored Chilean examinees and 4 favored Costa Rican examinees.

Based on the DIF analysis done to find out if culture-related DIF was present when students from Chile and Mexico were considered, exactly the same set of 6 DIF items was found through MH –1 A item, 2 B items, 3 C items– and LR–6 A items–. Out of these 6 items with DIF, 4 items were to the advantage of Mexico and 2 to the advantage of Chile.

Following the DIF analysis carried out to assess if DIF by culture was present among the test takers from Costa Rica and Mexico, a consistent set of 10 items was labeled with DIF in both MH –4 A items, 2 B items, 4 C items– and LR – 9 A items, 1 B item–. From these 10 items, half (5) was in favor of the reference group (Mexico) while the other half (5) was in favor of the focal group (Costa Rica).

In order to better grasp which groups were favored in terms of DIF caused by cultural differences, figure 15 is shown below. It reveals that when examining the performance of Chile and Costa Rica, Chile was more favored than Costa Rica

–Chile had 5 DIF items, Costa Rica had 4 DIF items–. Then in the case of Chile and Mexico, Mexican students had a significant advantage over Chilean students– Mexico with 4 DIF items, Chile with 2–. And lastly, when students from Costa Rica and Mexico were analyzed, both groups were favored alike –Costa Rica and Mexico with 5 DIF items each–.

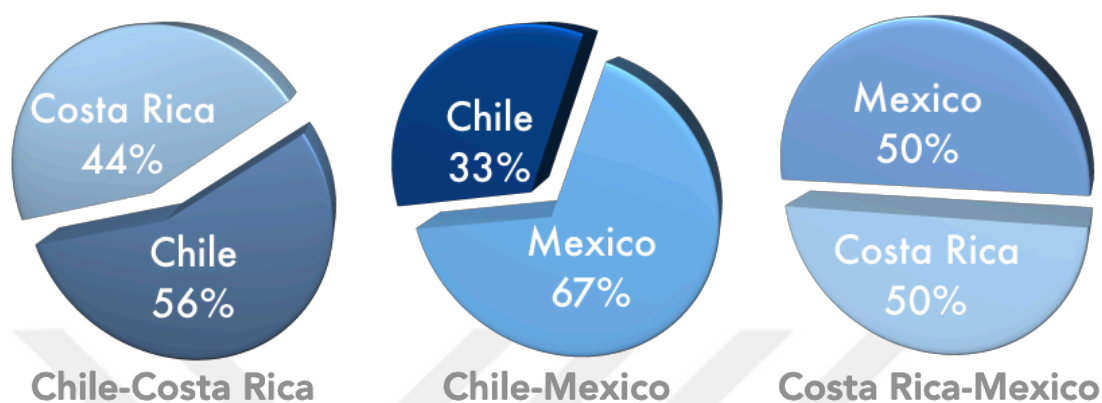


Figure 15. Graphs of culture-related DIF results from MH and LR methods

Results Related to Sub Research Question 2

Are there any items that indicate gender-related DIF among 16 dichotomous PISA 2015 science items from Chile, Costa Rica, and Mexico?

When the performance of the Chileans was considered and once the analysis to detect DIF caused by gender was finished, 5 DIF items were identified in the MH procedure –4 A items, 1 C item– whereas the LR procedure identified 5 DIF items –4 A items, 1 B item–. 4 of the 5 items showing DIF favored the focal group (female students) and just 1 favored the reference group (male students).

According to the DIF analysis ran to test for potential female/male-related DIF among the examinees from Costa Rica, 5 DIF items were found through MH – 2 A items, 1 B item, 2 C items– and 6 DIF's through LR –5 A items, 1 B item–. Out of these 6 items with DIF, 3 items were to the advantage of females from Costa Rica and 3 to the advantage of the males.

Regarding the DIF analysis carried out to assess if DIF by gender was present in the case of Mexico, it may be said that the MH technique categorized 3 items with DIF –2 A items, 1 B item– and the LR technique categorized 5 DIF's –5 A items–. From these 5 items, 3 items were in favor of the Mexican male test takers while 2 items were in favor of the Mexican female test takers.

Figure 16 is presented below in order to better demonstrate which groups were favored regarding DIF caused by the variable of gender. The graphs show that with respect to Chilean students' performance, the female ones were found considerably more favored than their male peers –Chilean females with 4 DIF items, Chilean males with just 1 DIF item–. Concerning students from Costa Rica, it was found that both genders were equally advantaged –females and males with 3 DIF items each–. When Mexican female and male test takers were assessed, the male group had more DIF items in their favor –males with 3 DIF items, females with 2 DIF items–.

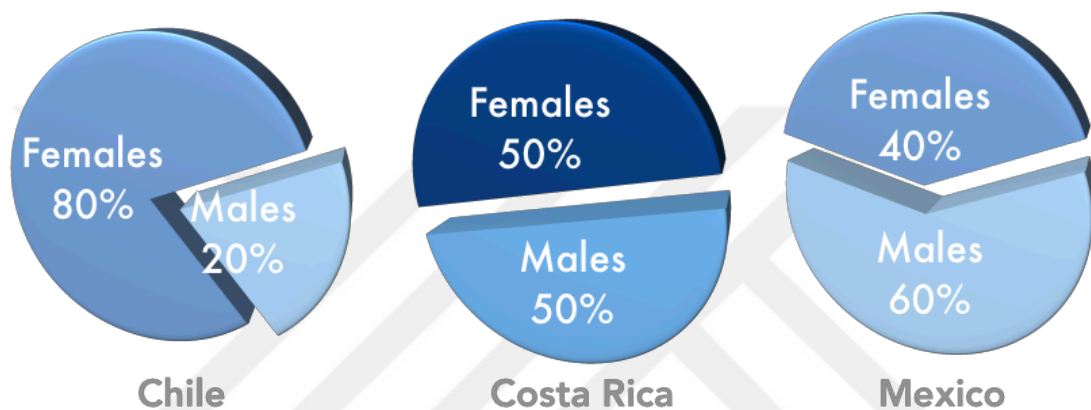


Figure 16. Graphs of gender-related DIF results from MH and LR methods

Overall Results

In general, it was identified that the two different techniques used for this DIF analysis yielded fairly consistent DIF estimates in terms of significance and direction of DIF. However, there were clear discrepancies in terms of magnitude of DIF which may be caused by variance in DIF categorization indexes yet, it is a regular part of research to compare performance of different methods.

It may be said that, when the variable of culture was examined, the MH procedure flagged more or the same number of DIF items in comparison to its counterpart. Interestingly enough, when the variable of gender was studied, the LR procedure classified more items as DIF items than MH. Furthermore, the MH analysis almost always yielded items in all DIF magnitudes (A, B, C), while LR mostly yielded A DIF items –a couple times B items but never a C item–.

It is worth underscoring that the culture-related DIF detection analysis revealed a large number of items marked with DIF as at least half and sometimes

more than half of the 16 items were labeled in the cases of two of the three country pairs. On the contrary, in DIF detection analysis implemented to identify DIF caused by gender differences, a much smaller amount of DIF items were found. It is concluded that the analyzed set of science items from PISA 2015 functioned more equally across genders than across cultures.

Recommendations

The two methods this research relied on were Mantel-Haenszel and Logistic Regression and, though non-uniform DIF detection is available in LR, that part of the method was not implemented. It is then recommended to include this part of DIF detection by running the third model in LR or by including other DIF methodologies that may identify non-uniform DIF presence. In addition, an analysis with multiple DIF methodologies could be carried out since they would compensate each other's disadvantages and this kind of analysis would offer a better understanding of DIF detection.

A set of 16 science items from PISA 2015 was investigated in this analysis yet a review of more science items could be conducted or a review including items from the three PISA core subjects as well. Moreover, this study was limited to only binary-scored items but it could be expanded to involve polytomous data.

The present study focused on DIF analysis, the first phase of item bias, and there is a chance to further it if the next phases of item bias are done. Also, the option of widening the research exists as the test level DIF factor could be added, and a comparison of item level DIF and test level DIF could be explored.

Item difficulty, item format and other features of items may be determining factors in DIF detection, it is therefore suggested a research related to the effects of varied DIF factors such as the ones already mentioned or others. Additionally, more than just the variables of gender and culture could be taken into consideration for group membership (reference and focal groups), given that other variables such as age, socio economic status, and school type could be used in other research projects.

It is also recommended that for future research both statistical and qualitative methods be performed for a more thorough examination of the

problematic items identified. Consequently, including a review of potential DIF sources could broaden this study, especially as the impact of cultural differences in international assessments' performance has been found significant.

Lastly, it can be said that there's room for improvement in terms of educational evaluation in all the examined countries. Their corresponding education authorities, test developers, and policy-makers are encouraged to work on developing the field of assessment, creating better evaluation instruments for examinees of all ages, and doing more researches about the performance of Spanish-speaking countries in order to learn from the findings. In fact, PISA personnel involved in the translation and adaptation processes of the test items could take thoughtful consideration of the results in this study and other studies, which have revealed a lack of consistent function of items across genders and cultures.

References

- Akın Arıkan, Ç., Uğurlu, S., & Atar, B. (2016). MIMIC, SIBTEST, Lojistik Regresyon ve Mantel-Haenszel Yöntemleriyle Gerçekleştirilen DMF ve Yanlılık Çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 34-52. DOI: 10.16986/HUJE.2015014226
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.
- Anastasi, A. (1976). *Psychological Testing* (4th ed). New York: Macmillan Publishing.
- Asil, M. (2005). *Differential Item Functioning (DIF) Analysis of the Verbal Section of the 2003 Student Selection Examination (SSE)*. (Master's Thesis, Ohio State University). Retrieved from <https://etd.ohiolink.edu/>
- Asil, M., & Gelbal, S. (2012). PISA Öğrenci Anketinin Kültürler Arası Eşdeğerliği. *Eğitim ve Bilim*, 37(166). Retrieved from <https://egitimvebilim.ted.org.tr/index.php/EB/article/view/1501>
- Atalay Kabasakal, K., Arsan, N., Gök, B., & Kelecioğlu, H. (2014). Comparing Performances (Type I Error and Power) of IRT Likelihood Ratio SIBTEST and Mantel-Haenszel Methods in the Determination of Differential Item Functioning. *Educational Sciences: Theory & Practice*, 14(6), 2186-2193. DOI: 10.12738/estp.2014.6.2165
- Baştuğ, Ö. Y. Ö. (2016). A Comparison of Four Differential Item Functioning Procedures in the Presence of Multidimensionality. *Educational Research and Reviews*, 11(13), 1251-1261. DOI: 10.5897/ERR2016.2803
- Bolt, D. M. (2000). A SIBTEST Approach to Testing DIF Hypotheses Using Experimentally Designed Test Items. *Journal of Educational Measurement*, 37(4), 307-327. Retrieved from: <https://www.jstor.org/stable/1435243>
- Brea, J. A. (2003). Population Dynamics in Latin America. *Population Bulletin*, 58(1), 1-36. Retrieved from <http://www.prb.org/Source/58.1PopulDynamicsLatinAmer.pdf>

- Butakor, P. K. (2015). A Comparison of DIF Detection and Effect Size Measures Among Mantel-Haenszel, SIBTEST, and Logistic Regression Using Science Test Data. *African Journal of Science and Research*, 4(2), 12-15. Retrieved from <http://ajsr.rstpublishers.com/>
- Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching Strategies and Differential Effectiveness Across Learning Contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, 30(41), 30-41. DOI:10.1016/j.stueduc.2016.03.005
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (3rd ed.). Mountain View, CA: Mayfield Publishing Company.
- Çepni, Z. (2011). *Differential Item Functioning (DIF) Analysis Using SIBTEST, Mantel Haenszel, Logistic Regression and Item Response Theory Methods*. (Doctoral Dissertation, Hacettepe University). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Demir, S., & Köse, İ. A. (2014). Mantel-Haenszel, SIBTEST ve Lojistik Regresyon Yöntemleri ile Değişen Madde Fonksiyonunu Analizi. *International Journal of Human Sciences*, 11(1), 700-714. DOI:10.14687/ijhs.v11i1.2798
- Doğan, N., & Öğretmen, T. (2008). Değişen Madde Fonksiyonunu Belirlemede Mantel-Haenszel, Ki-Kare ve Lojistik Regresyon Tekniklerinin Karşılaştırılması. *Eğitim ve Bilim*, 33(148), Retrieved from <https://egitimvebilim.ted.org.tr/index.php/EB/article/viewFile/679/130>
- Dossey, J., McCrone S., Turner, R., & Lindquist, M. (2008). PISA 2003-Mathematical Literacy and Learning in the Americas. *Canadian Journal of Science, Mathematics and Technology Education*, 8(2), 140-152. DOI: 10.1080/14926150802169289
- Ercikan, K., Roth, W-M., & Asil, M. (2015). Cautions About Inferences from International Assessments: The Case of PISA 2009. *Teachers College Record*, 117(010302), 1-28. Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17724>

- Fu, J., & Monfils, L. (2016, December). *LRDIF_ES: A SAS Macro for Logistic Regression Tests for Differential Item Functioning of Dichotomous and Polytomous Items*. ETS Research Memorandum Series, ETS RM-16-17. Retrieved from <https://www.ets.org/Media/Research/pdf/RM-16-17.pdf>
- García-Medina, A. M., Martínez-Rizo, F., & Cordero-Arroyo, G. (2016). Análisis del Funcionamiento Diferencial de los Ítems del Excale de Matemáticas para Tercero de Secundaria. *Revista Mexicana de Investigación Educativa*, 21(71), 1191-1220. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=5870210>
- Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000, April). *Performance of Mantel-Haenszel, Simultaneous Item Bias Test, and Logistic Regression When the Proportion of DIF Items is Large*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.385.5894>
- Gnaldi, M., & Bacci, S. (2016). Joint Assessment of the Latent Trait Dimensionality and Observed Differential Item Functioning of Students' National Tests. *Quality & Quantity: International Journal of Methodology*, 50(4), 1429-1447. DOI: 10.1007/s11135-015-0214-0
- Gómez-Benito, J., Hidalgo, M. D., & Padilla, J. L. (2009). Efficacy of Effect Size Measures in Logistic Regression: An Application for Detecting DIF. *Methodology*, 5(1), 18-25. DOI: 10.1027/1614-2241.5.1.18
- Grey House Publishing, Inc. (2003). *Nations of the World: A Political, Economic & Business Handbook* (3rd ed.). NY: Grey House Publishing, Inc.
- de Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). *Statistical Test Theory for the Behavioral Sciences*. Boca Raton, FL: Chapman & Hall / CRC.
- Hadjicharalambous, C. (2014). Commercial Versus Internally Developed Standardized Tests: Lessons from a Small Regional School. *Journal of Education for Business*, 89(1), 42-48. DOI: 10.1080/08832323.2012.740519

- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Hidalgo, M. D., & López-Pina, J. A. (2004). Differential Item Functioning Detection and Effect Size: A Comparison between Logistic Regression and Mantel-Haenszel Procedures. *Educational and Psychological Measurement*, 64(6), 903-915. DOI: 10.1177/0013164403261769
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential Item Functioning*. New York, NY: Routledge.
- Hussain, S., Tadesse, T., & Sajid, S. (2015). Norm-Referenced and Criterion-Referenced Test in EFL Classroom. *International Journal of Humanities and Social Science Invention*, 4(10), 24-30. Retrieved from <http://www.ijhssi.org/papers/v4%2810%29/Version-3/D04103024030.pdf>
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, 14(4), 329-349. DOI: 10.1207/S15324818AME1404_2
- Kartal, S. K., & Kutlu, Ö. (2017). Identifying the Relationships between Motivational Features of High and Low Performing Students and Science Literacy Achievement in PISA 2015 Turkey. *Journal of Education and Training Studies*, 5(12), 146-154. DOI: 10.11114/jets.v5i12.2816
- Kumar, R. (1999). *Research Methodology: A Step-by-step Guide for Beginners*. Malaysia: Sage Publications Ltd.
- Latifi, S., Bulut, O., Gierl, M., Christie, T., & Jeeva, S. (2016). Differential Performance on National Exams: Evaluating Item and Bundle Functioning Methods using English, Mathematics, and Science Assessments. *SAGE Publications*, 6(2), 1-14. DOI: 10.1177/2158244016653791
- Le, L. (2006, April). *Analysis of Differential Item Functioning*. Paper prepared for the annual meetings of the American Educational Research Association, San Francisco, CA. Retrieved from https://www.acer.org/files/analysis_of_dif.pdf

- Martínez-Rizo, F. (2016). Impacto de las pruebas en gran escala en contextos de débil tradición técnica: Experiencia de México y el Grupo Iberoamericano de PISA. *RELIEVE*, 22(1), 1-12. DOI: <http://dx.doi.org/10.7203/relieve.22.1.8244>
- Merriam-Webster, Inc. (1997). *Merriam Webster's Geographical Dictionary* (3rd ed.). USA: Merriam-Webster, Inc.
- Monahan, P., McHorney, C., Stump, T., & Perkins, A. (2007). Odds Ratio, Delta, ETS Classification, and Standardization Measures of DIF Magnitude for Binary Logistic Regression. *Journal of Educational and Behavioral Statistics*, 32(1), 92-109. DOI: 10.3102/1076998606298035
- OECD (2014). *PISA 2012 Results in Focus: What 15-Year-Olds Know and What They Can Do With What They Know*. Retrieved May 23, 2016 from <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>
- OECD (2016a). About the OECD. Retrieved from <http://www.oecd.org/about/>
- OECD (2016b). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264255425-en>
- OECD (2016c). *Skills in Ibero-America: Insights from PISA 2012*. Retrieved from <https://www.oecd.org/latin-america/Skills-IberoAmerica.pdf>
- OECD (2016d). "How Does PISA Assess Science Literacy?", *PISA In Focus*, No. 66, OECD Publishing, Paris. DOI: <https://doi.org/10.1787/5jl4nfnqt7l-en>
- OECD (2016e). "PISA 2015 Results In Focus", *PISA In Focus*, No. 67, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/aa9237e6-en>
- OECD (2016f). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264266490-en>
- OECD (2018). OECD and Costa Rica. Retrieved January 18, 2018 from <http://www.oecd.org/latin-america/countries/costarica/>
- Swaminathan, H., & Rogers, H. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational*

- Measurement*, 27(4), 361-370. Retrieved from <http://www.jstor.org/stable/1434855>
- Teresi, J. A., Ramirez, M., Lai, J. S., & Silver, S. (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly*, 50(4), 538-612. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2822361/pdf/nihms136945.pdf>
- Tierney, R. D. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory*. Singapore: Springer Reference. DOI 10.1007/978-981-287-532-7_400-1
- Uzun, N. B. & Gelbal S. (2017). PISA Fen Başarı Testinin Madde Yanlılığının Kültür ve Dil Açısından İncelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446. Retrieved from <https://dergipark.gov.tr/kefdergi/issue/31577/352755>
- Wainer, H., & Braun, H. I. (Eds.). (1988). *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Wright, K. D., & Oshima, T. C. (2015). An Effect Size Measure for Raju's Differential Functioning for Items and Tests. *Educational and Psychological Measurement*, 75(2), 338-358. DOI: 10.1177/0013164414532944
- Yıldırım, H. H. (2006). *The Differential Item Functioning (DIF) Analysis of Mathematics Items in the International Assessment Programs*. (Doctoral Dissertation, Middle East Technical University). Retrieved from <https://etd.lib.metu.edu.tr/upload/12607135/index>
- Yu, C. H., Popp, S. O., Digangi, S., & Jannasch-Pennell, A. (2007). Assessing unidimensionality: A comparison of Rasch modeling, Parallel analysis, and TETRAD. *Practical Assessment, Research and Evaluation*, 12(14), 1-19. Retrieved from <https://pdfs.semanticscholar.org/406d/212b3cbf51056b7c31ded5ac10824a6330f3.pdf>
- Zheng, Y., Gierl, M. J., & Cui, Y. (2007). *Using Real Data to Compare DIF Detection and Effect Size Measures Among Mantel-Haenszel, SIBTEST,*

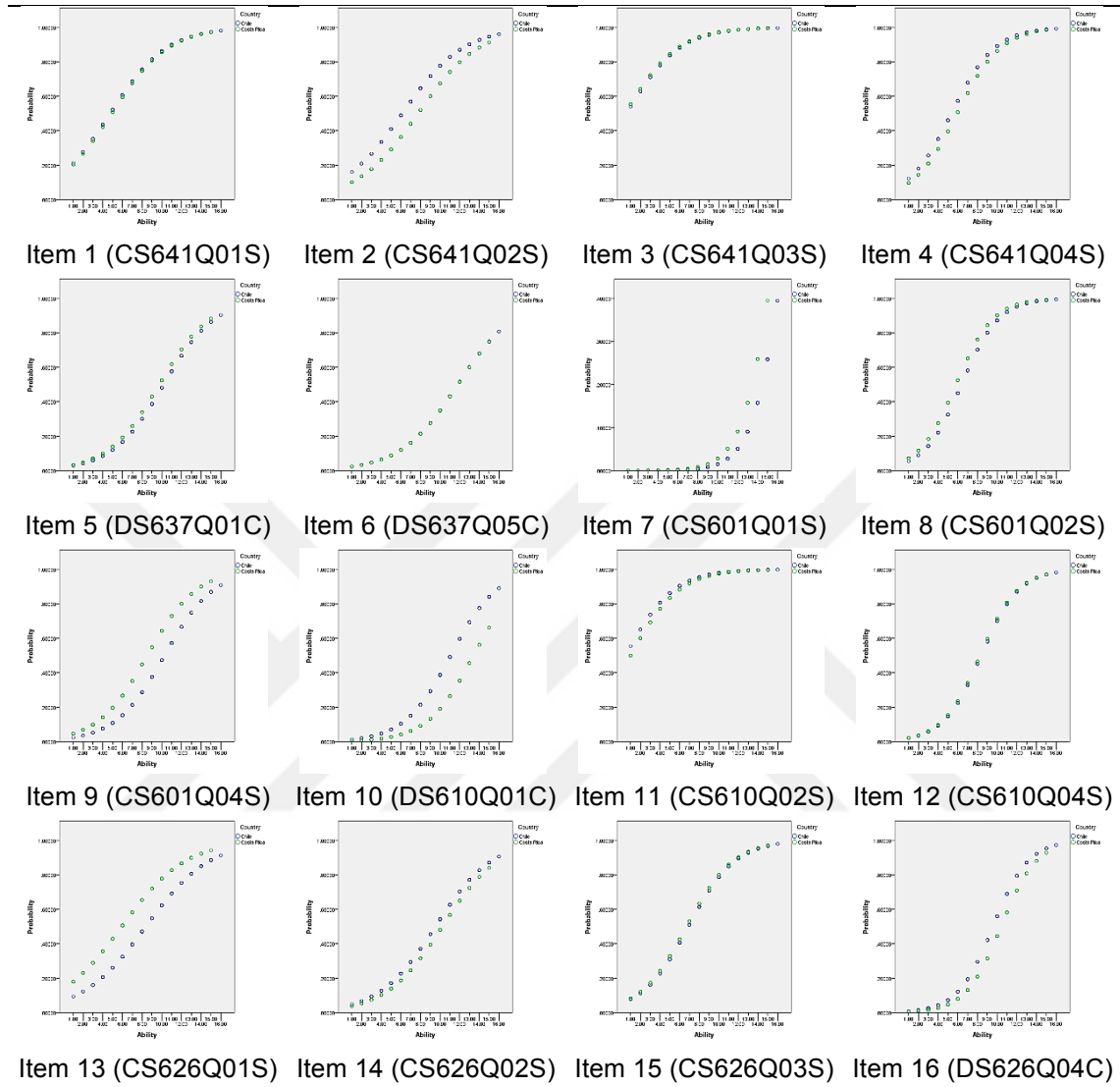
and Logistic Regression Procedures. Paper prepared for the Center for Research in Applied Measurement and Evaluation, Edmonton, Canada. Retrieved from <https://pdfs.semanticscholar.org/484d/e487ddba99a36fd5dda55658eb7e15bad9dd.pdf>

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

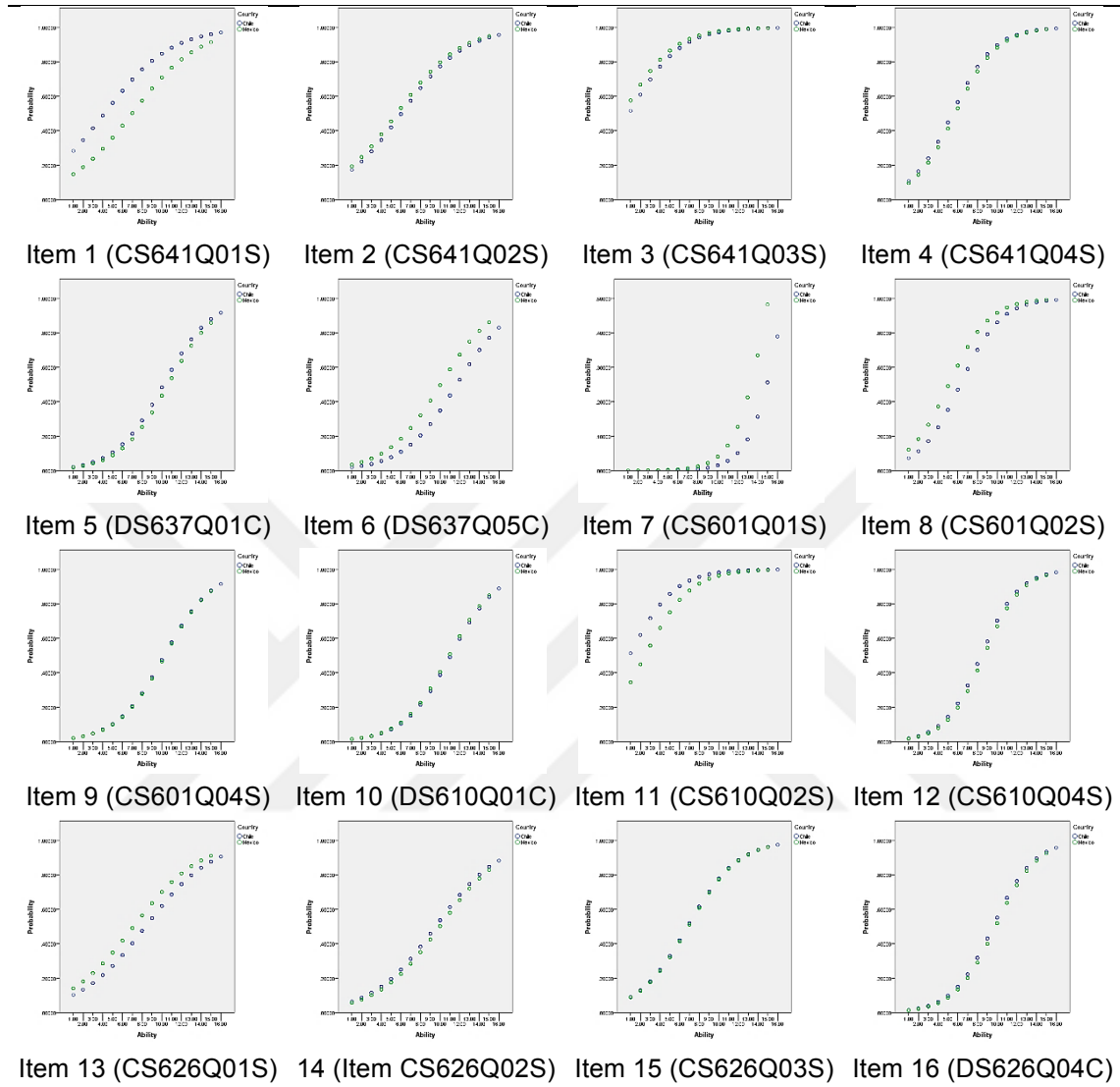
Zwick, R. (2012, May). *A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement*. ETS Research Report Series, ETS RR-12-08. Retrieved from <https://www.ets.org/Media/Research/pdf/RR-12-08.pdf>

APPENDIX-A: Probability of Correct Answer Graphs by Culture and Gender

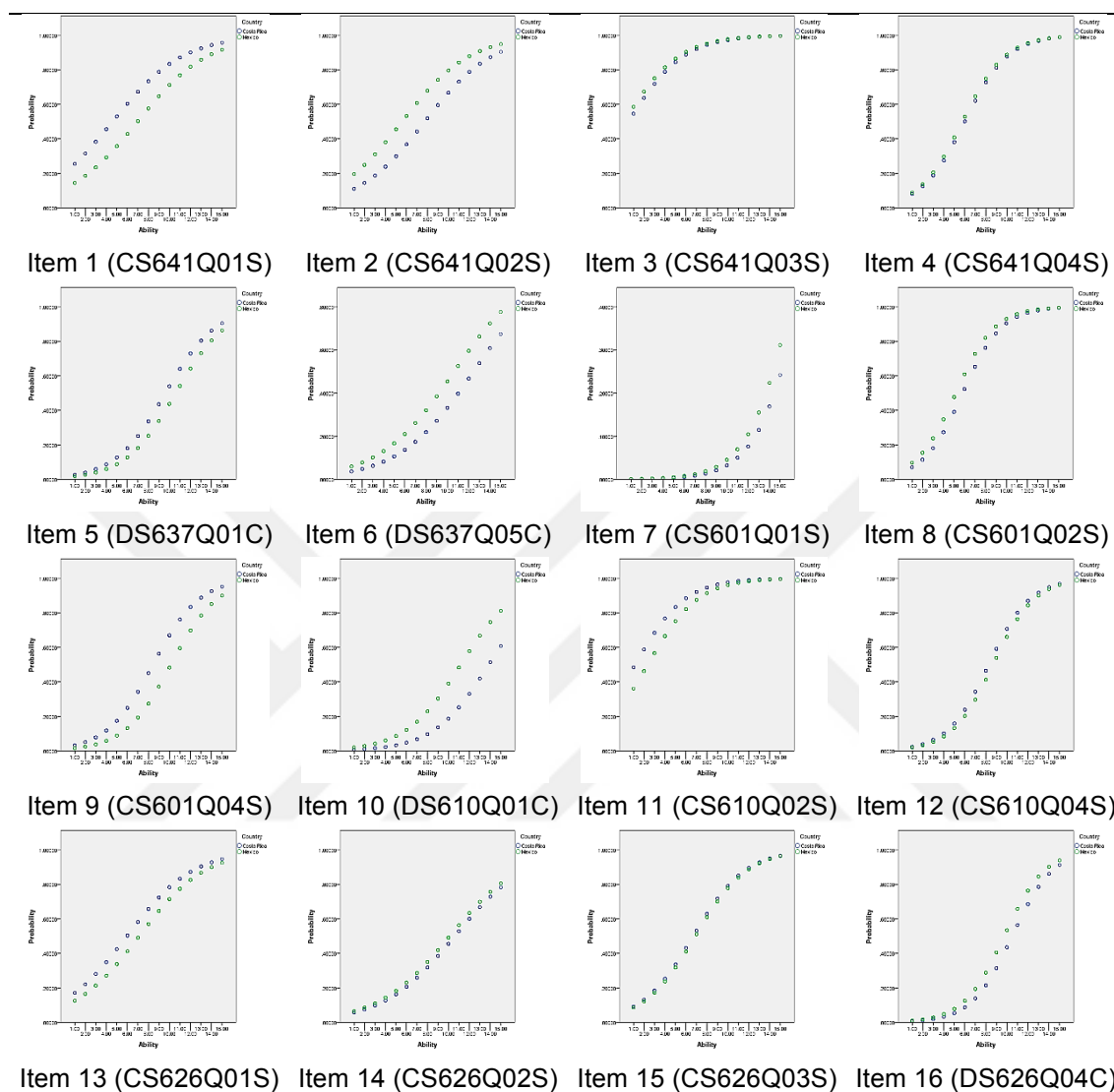
Probability of Correct Answer by Culture for Chile and Costa Rica



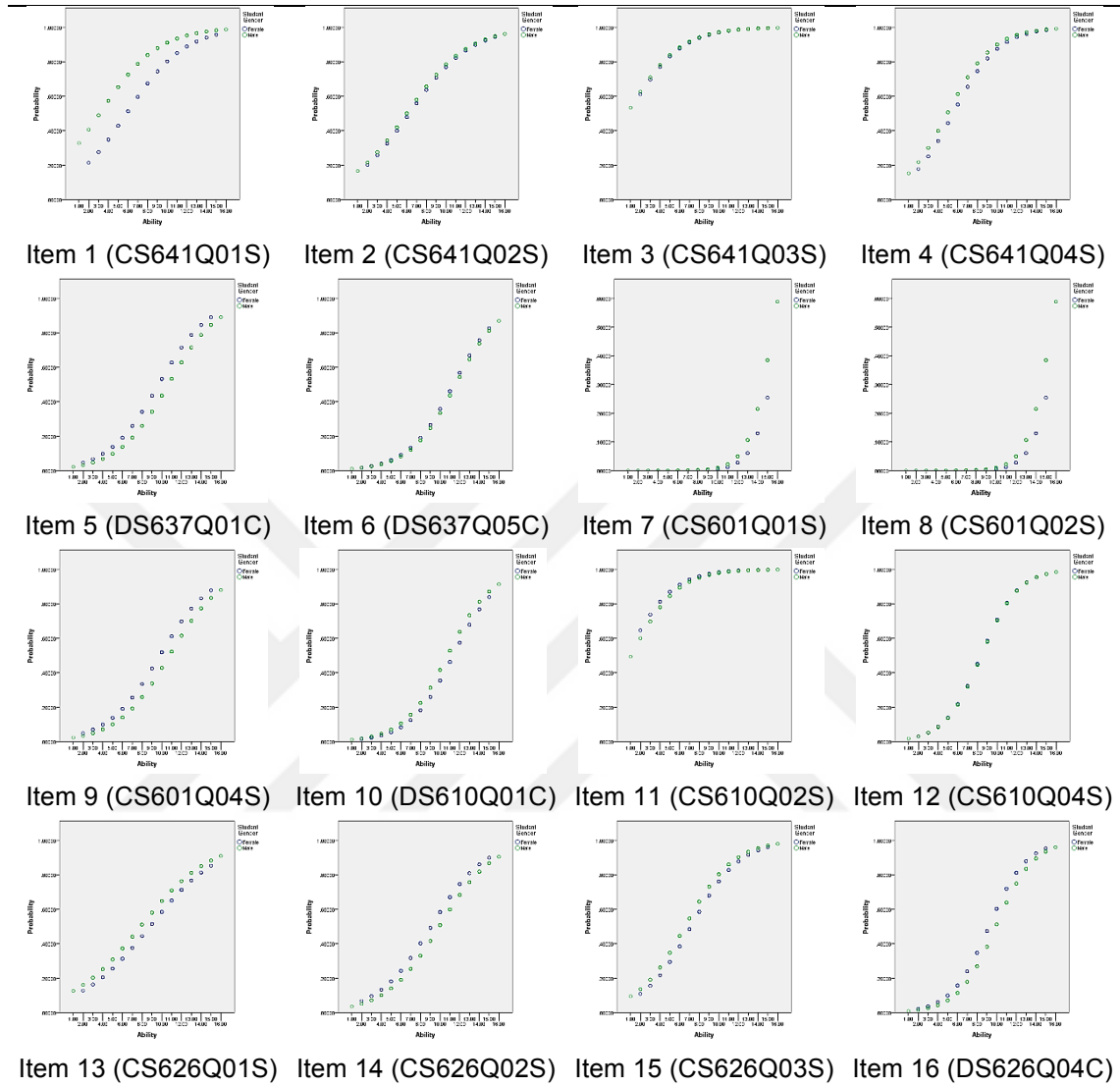
Probability of Correct Answer by Culture for Chile and Mexico



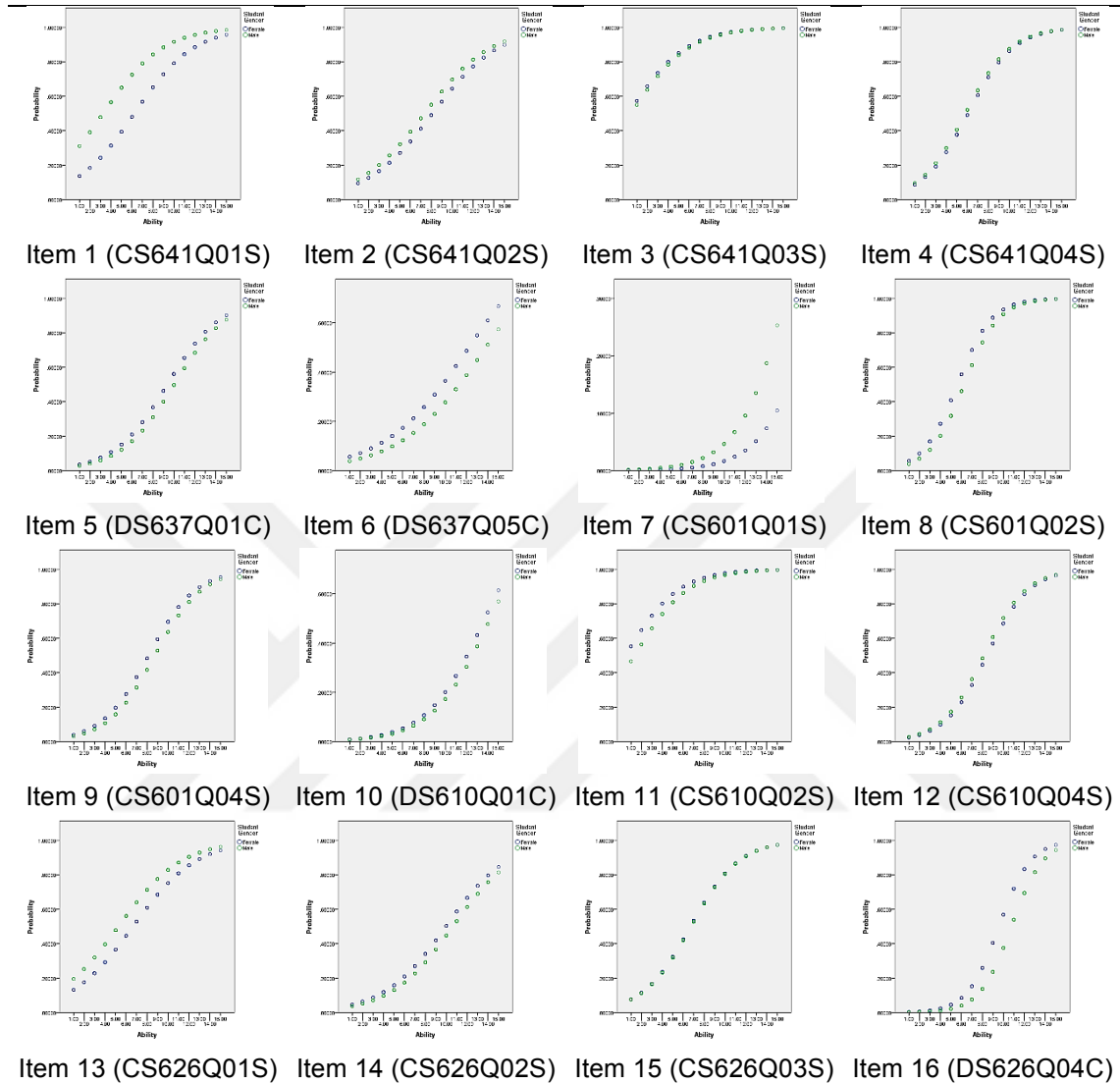
Probability of Correct Answer by Culture for Costa Rica and Mexico



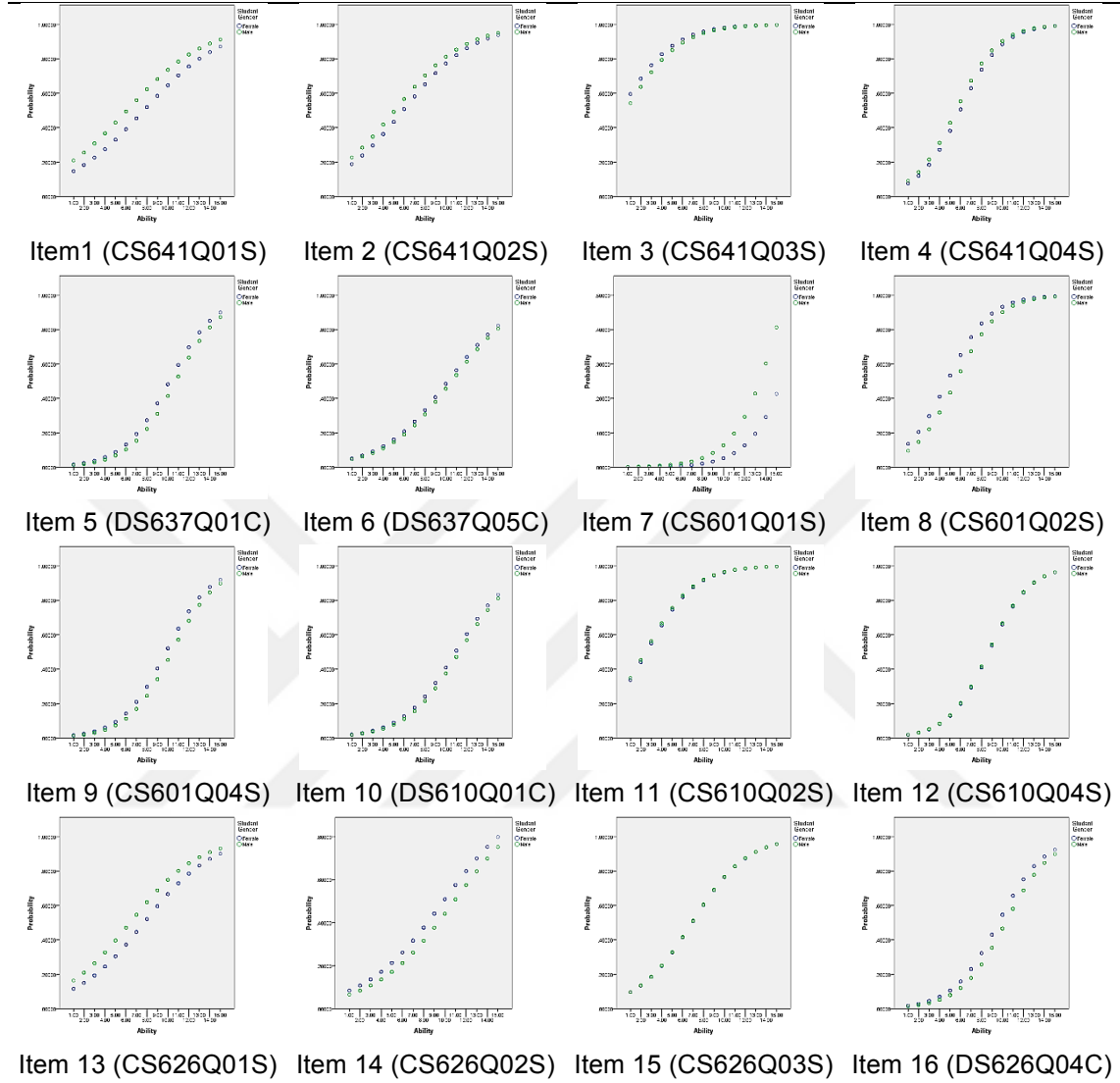
Probability of Correct Answer by Gender for Chile



Probability of Correct Answer by Gender for Costa Rica



Probability of Correct Answer by Gender for Mexico



APPENDIX-B: SPSS Syntaxes for LR and Probability Graphs

SPSS Syntax Used for LR by Country Pairs

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q03S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q01C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q05C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS610Q01C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q03S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS626Q04C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL CNTRYID
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

```

SPSS Syntax Used for LR by Country

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS641Q01S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS641Q02S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS641Q03S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS641Q04S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES DS637Q01C  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES DS637Q05C  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS601Q01S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.  
LOGISTIC REGRESSION VARIABLES CS601Q02S  
  /METHOD=ENTER TOTAL  
  /METHOD=ENTER TOTAL ST004D01T  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS610Q01C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q03S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS626Q04C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER TOTAL ST004D01T
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

```

SPSS Syntax Used for LR Probability Data by Gender

* Encoding: UTF-8.

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q03S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q01C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q05C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER ST004D01T
  /CONTRAST (ST004D01T)=Indicator
  /SAVE=PRED
```

```

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q01S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q04S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS610Q01C
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q04S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q01S
/METHOD=ENTER TOTAL

```

```
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q03S
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS626Q04C
/METHOD=ENTER TOTAL
/METHOD=ENTER ST004D01T
/CONTRAST (ST004D01T)=Indicator
/SAVE=PRED
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

SPSS Syntax Used for Graphs by Gender

```
* Encoding: UTF-8.

* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=TOTAL PRE_1
  ST004D01T MISSING=LISTWISE
  REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: TOTAL=col(source(s), name("TOTAL"), unit.category())
  DATA: PRE_1=col(source(s), name("PRE_1"))
  DATA: ST004D01T=col(source(s), name("ST004D01T"),
unit.category())
  GUIDE: axis(dim(1), label("Ability"))
  GUIDE: axis(dim(2), label("Probability"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior),
label("Student Gender"))
  SCALE: linear(dim(2), include(0))
  SCALE: cat(aesthetic(aesthetic.color.exterior), include("1",
"2"))
  ELEMENT: point(position(TOTAL*PRE_1),
color.exterior(ST004D01T))
END GPL.
```

SPSS Syntax Used for LR Probability Data by Culture

* Encoding: UTF-8.

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q01S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q02S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q03S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS641Q04S
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q01C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

```
DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS637Q05C
  /METHOD=ENTER TOTAL
  /METHOD=ENTER CNTRYID
  /CONTRAST (CNTRYID)=Indicator
  /SAVE=PRED
```

```

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q01S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS601Q04S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS610Q01C
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS610Q04S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5) .

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q01S
/METHOD=ENTER TOTAL

```

```

/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q02S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES CS626Q03S
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

DATASET ACTIVATE DataSet1.
LOGISTIC REGRESSION VARIABLES DS626Q04C
/METHOD=ENTER TOTAL
/METHOD=ENTER CNTRYID
/CONTRAST (CNTRYID)=Indicator
/SAVE=PRED
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

```

SPSS Syntax Used for Graphs by Culture for Chile and Costa Rica

```
* Encoding: UTF-8.

* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=TOTAL PRE_1
  CNTRYID MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: TOTAL=col(source(s), name("TOTAL"), unit.category())
  DATA: PRE_1=col(source(s), name("PRE_1"))
  DATA: CNTRYID=col(source(s), name("CNTRYID"),
unit.category())
  GUIDE: axis(dim(1), label("Ability"))
  GUIDE: axis(dim(2), label("Probability"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior),
label("Country"))
  SCALE: linear(dim(2), include(0))
  SCALE: cat(aesthetic(aesthetic.color.exterior),
include("152", "188"))
  ELEMENT: point(position(TOTAL*PRE_1),
color.exterior(CNTRYID))
END GPL.
```

SPSS Syntax Used for Graphs by Culture for Chile and Mexico

```
* Encoding: UTF-8.

* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=TOTAL PRE_1
  CNTRYID MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: TOTAL=col(source(s), name("TOTAL"), unit.category())
  DATA: PRE_1=col(source(s), name("PRE_1"))
  DATA: CNTRYID=col(source(s), name("CNTRYID"),
unit.category())
  GUIDE: axis(dim(1), label("Ability"))
  GUIDE: axis(dim(2), label("Probability"))
  GUIDE: legend(aesthetic(aesthetic.color.exterior),
label("Country"))
  SCALE: linear(dim(2), include(0))
  SCALE: cat(aesthetic(aesthetic.color.exterior),
include("152", "484"))
  ELEMENT: point(position(TOTAL*PRE_1),
color.exterior(CNTRYID))
END GPL.
```

SPSS Syntax Used for Graphs by Culture for Costa Rica and Mexico

* Encoding: UTF-8.

* Chart Builder.

GGRAPH

```
/GRAPHDATASET NAME="graphdataset" VARIABLES=TOTAL PRE_1  
CNTRYID MISSING=LISTWISE REPORTMISSING=NO  
/GRAPHSPEC SOURCE=INLINE.
```

BEGIN GPL

```
SOURCE: s=userSource(id("graphdataset"))  
DATA: TOTAL=col(source(s), name("TOTAL"), unit.category())  
DATA: PRE_1=col(source(s), name("PRE_1"))  
DATA: CNTRYID=col(source(s), name("CNTRYID"),  
unit.category())  
GUIDE: axis(dim(1), label("Ability"))  
GUIDE: axis(dim(2), label("Probability"))  
GUIDE: legend(aesthetic(aesthetic.color.exterior),  
label("Country"))  
SCALE: linear(dim(2), include(0))  
SCALE: cat(aesthetic(aesthetic.color.exterior),  
include("188", "484"))  
ELEMENT: point(position(TOTAL*PRE_1),  
color.exterior(CNTRYID))  
END GPL.
```

APPENDIX-C: Ethics Committee Approval

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

13 / 01 / 2017

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

Tez Başlığı / Konusu:	Şili, Meksika ve İspanya'nın PISA 2015 Fen Maddelerinin üzerine Değişen Madde Fonksiyonu Analizi
-----------------------	--

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmamasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Cozby Janelle DZUL GARCIA
(Öğrencinin Adı Soyadı, İmzası)



Öğrenci Bilgileri

Adı Soyadı	Cozby Janelle DZUL GARCIA
Öğrenci No	N14120424
Anabilim Dalı	Eğitim Bilimleri Anabilim Dalı
Programı	Ölçme ve Değerlendirme
Statüsü	☒ Yüksek Lisans ☐ Doktora ☐ Bütünleşik Dr.

Danışman Görüşü ve Onayı

Doç. Dr. Burcu ATAR
(İmza)
(Danışmanın Ünvanı, Adı ve Soyadı)



APPENDIX D: Declaration of Ethical Conduct

I hereby declare that...

- I have prepared this thesis in accordance with the thesis writing guidelines of the Graduate School of Educational Sciences of Hacettepe University;
- all information and documents in the thesis/dissertation have been obtained in accordance with academic regulations;
- all audio visual and written information and results have been presented in compliance with scientific and ethical standards;
- in case of using other people's work, related studies have been cited in accordance with scientific and ethical standards;
- all cited studies have been fully and decently referenced and included in the list of References;
- I did not do any distortion and/or manipulation on the data set,
- and **NO** part of this work was presented as a part of any other thesis study at this or any other university.

06 / 07 / 18



Cozby Janelle DZUL GARCIA

APPENDIX-E: Thesis/Dissertation Originality Report

26/02/2019

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: DIFFERENTIAL ITEM FUNCTIONING ANALYSIS OF CHILE, COSTA RICA AND MEXICO'S PISA 2015 SCIENCE ITEMS

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
26/02/2019	64	82453	06/07/2018	12%	1084032112

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: COZBY JANELLE DZUL GARCIA

Student No.: N14120424

Department: Educational Sciences

Program: Educational Measurement and Evaluation

Status: ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.



ADVISOR APPROVAL

APPROVED



Assoc. Prof. Burcu ATAR

APPENDIX-F: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- o Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- o Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

06 / 07 / 2018



Cozby Janelle DZUL GARCIA

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

