Educational Sciences: Theory & Practice - 12(1) • Winter • 222-229

www.edam.com.tr/estp

An Investigation of Item Bias in Science & Tecnology Subtests and Mathematic Subtests in Level Determination Exam (LDE)

İsmail KARAKAYA[a]
Gazi University

Abstract

The aim of this study is to determine the items in Science and Technology and Mathematics subtests of 6th, 7th, and 8th grades in the 2009 LDE, which was performed to select the students for the secondary schools, exhibited item bias with regard to student gender. Mantel-Haenszel (MH) method was used in order to determine the differential item functioning (DIF) of the items in the tests. The study was carried out with 22624 students in total, 6913 of whom were 6th year students (3614 males and 3299 females), 6333 of whom were 7th year students (3277 males and 3066 females) and 9374 of whom were 8th year students (4290 males and 5084 females). When the Science and technology subtest items are examined, it is seen that 2 (items number 6 and 16) in 6th grade, 3 in 8th grade (item number 14, 15, and 17) tests have B-level DIF. Among those five items 3 are favoring girls and 2 are favoring boys. In Mathematics subtests, item number 3 in 6th grade, item number 3 in 7th grade, and items 2 and 6 in 8th grade have DIF at B level. While 3 of those items work in favor of males, remaining 1 item works in favor of females. Content experts are asked for their opinion on the items having DIF to explore if any of them are biased. Based on the expert reviews it has been concluded that none of them includes gender bias.

Key Words

Differential Item Functioning, Item Bias, Logistic Regression, Mantel-Haenszel.

In Turkey, the central examinations have been performed with the aim of selecting and placing the students in the various education institutions for a long time. Among these central examinations come the ones that are performed in the transition from primary education to secondary education first.

In the transition from primary education to secondary education, Secondary Education Student Selection and Placement Exams (SESSPE) were performed by the Ministry of National Education (MNE) between 1998 and 2008.

a **İsmail KARAKAYA, Ph.D.**, Ismail Karakaya studies at Department of Educational Measurement and Evaluations as PhD. His research interests include scale development, alternative assessment in education. Correspondence: PhD. Ismail KARAKAYA, Gazi University, Faculty of Education, Department of Educational Measurement and Evaluations, 06500 Teknikokullar/ANKARA. E-mail: ikarakaya2002@yahoo.com Phone: +90 0312 202 8230 Fax: +90 312 222 8483.

Developed based on the primary school curriculum and quantifying the academic skills (Milli Eğitim Bakanlığı [MEB], 1998). The aim of SESSPE in general is trying to measure the intellectual processes through the questions in the tests such as; *a*) to make assumptions from the concepts, graphics, tables and principles used in the sum of the given information, *b*) to come to the conclusion by relating the components of the information pattern provided as figure, graphic, or table, *c*) to understand the written text, graphic and table, to make generalization, to interrelate among constituents and to reach a conclusion by using the relations, *d*) to be able to evaluate a given problem, written text, figure, graphic or table by certain measures, *e*) to make an inference from observations of daily life by using the basic concepts and principles, to give similar examples and to solve the problem no matter what the subject area was (Kutlu & Karakaya, 2007; MEB, 2003).

In order to select students and place them in secondary schools, "Level Determination Exam (LDE)"has been performed for the 6th, 7th, and 8th grades since the academic year 2007 – 2008. The LDE includes five subtests, which are Turkish, Mathematics, Science and Technology, Social Sciences, and Foreign Language. These subtests in LDE are prepared in order to determine to what extent students have educational attainment defined in educational programs of each related class level. The scores obtained from this exam are used for the placement of the students in relevant secondary schools (MEB, 2007). In other words, the questions in the LDE only determine the success level of students in the year they receive education, and it does not include their other class status. Since the subtests in the LDE are generally prepared in the way targetting the gains of the educational programs, they are expected to be qualified to measure the knowledge and skills that teaching programmes try to develop.

Each year, hundreds of thousands of applicants take the student selection and placement exams such as LDE. While some applicants are placed in relevant programs in accordance with their success level, some cannot be placed in anywhere because of their insufficient scores. In these exams, decisions are expected to be error-free since they shape the future of the individuals. This will be primarily possible through measuring tools whose validity and reliability were ensured. The validity and reliability of these exams will be possible through operating them in accordance with their purposes.

When the literature is reviewed, various studies are encountered on the SESSPEs and the LDEs. It is seen that the studies on the SESSPE and the LDE are intended to estimate the predictive validity, concurrent validity, and construct validity (Anıl & Güzeller, 2010; Anıl, Güzeller, Çokluk, & Şekercioğlu, 2010; Berberoğlu, Kaptan, & Kutlu, 2002; Güzeller, 2005; Kutlu & Karakaya, 2007; Örs, 2010). It is also seen that several studies have been carried out on the differential item functioning (DIF) in the items of SESSPE subtests and item bias (Doğan & Öğretmen, 2008; Gök, Kelecioğlu, & Doğan, 2010; Öğretmen & Doğan, 2004; Yurgudül, 2003).

A valid and reliable selection exam indicates that the exam has been carried out in accordance with its target and the test items were prepared in line with the purpose of the exam. In addition, it can be commented on whether structures and contents of the items in the tests provide neither advantage nor disadvantage to any of the groups taking the exam. If the items in the tests provide an advantage for a group because of various features such as gender, socio-economic status, religion, etc., it can be said that the exam has bias in favor of that group. This will cause the decisions made based on the test scores to be discussed, and the validity of the decisions to be affected negatively. In other words, if the tests provide an advantage or a disadvantage to a group, there is a systematic error (Zumbo, 1999). This will directly affect the validity of the exam in a negative way.

Determination of the item bias of the items in the tests is one of the important works to increase the test validity and reliability. Kristjansson, Aylesworth, Mcdowell, and Zumbo (2005), pointed out that item bias was an important factor to threaten the validity of the measurements, and test and item bias detection methods should be used as much as possible.

There has not been any study carried out on any item bias or the differential item functioning in the subtests of the LDE. This study is important due to its contributions to the validity of the subtests of the LDE.

Applying the secondary school selection and placement exams like LDE to another group in advance, and making arrangements for the items of the subtests as a result of this application does not seem possible in terms of the reliability and validity of the exams. The items in the subtests, which are prepared only by field experts, are evaluated by field, language and measurement experts, and after the evaluation, necessary corrections are made and the test booklets are created by taking into account the features like content, number of the questions etc. Direct application of these tests and items according to its purpose without any trial is only possible by people who are specialized in the item writing. Determination of the items with DIF in the tests and knowing if these items work in favor of various sub-groups will contribute the item writing process. This is important in terms of the tests to be applied serving the purpose more.

The item bias studies start with a statistical process that determines whether item function differs for the individuals who are at the same skill levels in different groups. The items that demonstrate the differential item functioning in the sub-groups with same skill levels are determined (Gök et al., 2010; Roever, 2005). After that, by obtaining opinions of the field experts on the reasons why items demonstrate DIF, it is tried to expose if the group differentiation actually results from the differences among the skills or the measuring process (Camilli & Shepard, 1994; Zumbo, 1999)

Taking into consideration of the factors causing the item bias at test development stage will contribute to give more accurate decisions about the students on scoring and evaluation of tests (Allalouf, Hambleton, & Sireci, 1999). The probability of answering the item correctly might differ if individuals vary in knowledge and skill levels. In other words, the differences in performances of individuals with different skill levels cannot be taken into account as a sign of item bias (Schumacher, 2005).

This study is conducted to determine if the items of Science and Technology, and Mathematics subtests used in 6th, 7th, and 8th grades in the LDE exam demonstrated item bias. For this purpose, answers were sought for the following questions;

1. Do the items in the Science and Technology subtests of 6th, 7th, and 8th grades in the LDE function differently by gender?
2. Do the items in the Mathematics subtests of 6th, 7th, and 8th grades in the LDE function differently by gender?
3. Do the items that have differential item functioning exhibit bias by gender according to the expert opinions?

Method

The aim of this study is to determine if the items in Science and Technology, and Mathematics subtests of 6th, 7th, and 8th grades in the 2009 LDE exhibited item bias with regard to the variation of gender. Since the research was intended for determining an existing situation, it is a descriptive study.

The research population was the students who lived in Ankara in 2009 and took the LDE exam. The randomly selected students among the students lived in Ankara province were the samples of the research. Class and gender of the students who were research samples and their numbers according to subtests are provided in Table 1.

As it is seen in Table 1, the answers of Science and Technology and Mathematics subtests in the LDE of 22620 students (6913 6th graders, 6333 7th graders and 9374 8th graders) were used in this study.

In this study, the data belonged to the Science and Technology and Mathematics subtests of 6th, 7th, and 8th grades of the LDE, which was performed for selection and placement of the students in 2009, were used. There were 16, 18, and 20 questions for 6th, 7th, and 8th grades in the Science and Technology and Mathematics subtests, respectively. The data used in this study were obtained from MNE.

Considering the techniques to determine the differential item functioning, it can be seen that classical test theory is categorized as a technique based on item response theory. Between the Mantel-Haenszel (MH) and logistic regression (LR) methods, which are the techniques based on the classical test theory, the MH is the most preferred method for determination of DIF because it has easy application and statistical interpretation, gives effective results for small groups, and is most efficient to determine uniform DIF (Guilera, Gomez-Benito & Hidalgo, 2009). Therefore, the MH was preferred at determination phase of DIF in this study. The EZ-DIF program (Waller, 1998) was used for the MH analysis at the stage of analyzing the data.

Results

In this section, in order to determine whether items in Science and Technology and Mathematics subtests of 6th, 7th, and 8th grades in the 2009 LDE exhibited item bias, the differential item functioning with regard to variation of gender was investigated. Findings obtained from the analyses were tabulated and interpreted below. After determination of the items with DIF, it was examined if the regarded items demonstrated item bias by obtaining opinions from field experts and measurement and evaluation experts.

Table 1.
Number of Students by Subtests, Grade and Gender

Class Level	Science and Technology Subtests			Mathematic Subtests		
	Male	Female	Total	Male	Female	Total
6th	3614	3299	6913	3592	3274	6866
7th	3277	3066	6333	3061	3052	6113
8th	4290	5084	9374	4290	5084	9374

Do the Items in the Science and Technology Subtests of 6th, 7th, and 8th Grades in the LDE Function Differently by Gender?

The results belonging to the MH method related to the items giving DIF in regard to the variation of gender of the items in the 6th grade Science and Technology subtests are provided in Table 2.

Table 2.
MH Analysis Results of the 6th Grade Science and Technology Subtest Items

Items	α*	χ^2	p**	Δ-MH	SE DMF	DIF Level***
1	0.736	29.852	0.000	0.721	0.132	A
2	1.227	16.390	0.000	-0.480	0.119	A
5	0.795	11.796	0.001	0.538	0.155	A
6	0.584	57.896	0.000	1.264	0.166	B
7	0.868	4.957	0.026	0.334	0.148	A
8	1.169	6.252	0.012	-0.367	0.145	A
9	1.150	6.216	0.013	-0.328	0.130	A
11	0.829	11.548	0.001	0.440	0.129	A
13	1.155	4.828	0.028	-0.339	0.152	A
14	1.415	43.504	0.000	-0.816	0.124	A
15	1.136	4.588	0.032	-0.300	0.138	A
16	1.601	33.183	0.000	-1.105	0.191	B

Focus group: Male (n=3614) Referance group: Female (n=3299)

* Odds ratio represents the performance diffefrrence between focus and reference groups (Gök et al., 2010)

** In MH analysis, a significant Chi-Square means there is A, B, or C DIF.

*** If [Δ-MH] <1 DIF is ignorable (A); if 1< [Δ-MH] <1.5 DIF is moderate (B), if 1.5<[Δ-MH] DIF is at an impotrtant level (C) (Zieky, 1993).

It was found that 10 items at the level of "A" and 2 items at the level of "B" with a total of 12 out of the 16 questions in the 6th grade Science and Technology subtest included DIF. Only the DIFs in 10 items at the level of "A" can be tolarated. Among the items included DIF at the level of "B", 6th item was seen to work in favor of boys, whereas 16th item worked in favor of girls.

The results of the MH analysis related to the materials giving DIF in regard to the variation of gender in the 7th grade Science and Technology subtests are provided in Table 3.

According to the results of the MH analysis regarding to the variation of gender in the 7th grade Science and Technology subtest, DIF was found in only 12 items out of the 18 questions at the level "A". There was not observed DIF at the level of "B" or level of "C" in the items in the test. Not having any question with DIF at the level of "B" or level of "C" in the 7th grade Science and Technology subtests can be interpreted as not having any material that shows item bias.

Table 3.
MH Analysis Results of the 7th Grade Science and Technology Subtest İtems

Items	α	χ^2	p	Δ-MH	SE DMF	DIF Level
1	1.195	9.356	0.002	0.418	0.136	A
2	1.452	41.768	0.000	0.876	0.136	A
3	1.135	4.852	0.028	0.297	0.133	A
4	1.271	12.978	0.000	0.564	0.156	A
5	1.523	52.167	0.000	0.989	0.137	A
7	1.141	4.314	0.038	0.311	0.147	A
9	0.885	4.371	0.037	-0.287	0.135	A
11	0.781	15.720	0.000	-0.580	0.145	A
14	0.703	28.043	0.000	-0.827	0.155	A
16	0.656	26.524	0.000	-0.989	0.191	A
17	0.868	6.173	0.013	-0.333	0.132	A
18	0.815	9.405	0.002	-0.481	0.155	A

Focus group: Male (n=3277) Referance group: Female (n=3066)

The results belonging to the MH method relating to the materials giving DIF in regard to the variation of gender in the 8th grade Science and Technology subtests are provided in Table 4.

Table 4.
MH Analysis Results of the 8th Grade Science and Technology Subtest İtems

Items	α	χ^2	p	Δ-MH	SE DMF	DIF Level
1	1.281	30.060	0.000	-0.582	0.106	A
2	0.805	3.924	0.048	0.511	0.252	A
3	1.185	10.721	0.001	-0.400	0.121	A
5	0.763	3.969	0.046	0.637	0.311	A
6	0.871	7.829	0.005	0.326	0.115	A
8	0.772	27.448	0.000	0.609	0.116	A
10	1.137	6.987	0.008	-0.302	0.113	A
12	1.280	5.868	0.015	-0.581	0.235	A
13	0.845	14.438	0.000	0.395	0.103	A
14	0.623	46.297	0.000	1.111	0.163	B
15	1.679	13.378	0.000	-1.218	0.330	B
16	1.397	26.743	0.000	-0.786	0.151	A
17	1.735	13.568	0.000	-1.295	0.349	B
19	0.836	17.415	0.000	0.421	0.100	A
20	1.167	5.343	0.021	-0.363	0.155	A

Focus group: Male (n=3277) Referance group: Female (n=3066)

Out of 20 questions in the 8th grade Science and Technology subtest, 13 items were found to have DIF at the level of "A" and 3 items were found to have DIF at the level of "B", whereas there was not any item included DIF at the level of "C". It can be said that while 14th item was seen to work in favor of boys, 15th and 17th items were seen to work in

favor of girls among the all questions included DIF at the level of "B".

Do the Items in the Mathematics Subtests of 6th, 7th, and 8th Grades in the LDE Function Differently by Gender?

The MH analysis results, which showed whether the items belonged to the Mathematics subtests had DIF regarding to the variation of gender, are given below.

Table 5.
MH Analysis Results of the 6th Grade Mathematics Subtests Items

Items	α	χ^2	P	Δ-MH	SE DMF	DIF Level
3	1.651	65.728	0.000	-1.179	0.145	B
5	0.880	5.020	0.025	0.301	0.133	A
6	0.883	5.313	0.021	0.292	0.125	A
7	0.857	7.498	0.006	0.361	0.131	A
10	1.327	15.654	0.000	-0.664	0.167	A
12	1.151	5.124	0.024	-0.330	0.144	A
13	0.860	7.884	0.005	0.356	0.126	A
16	0.827	11.858	0.001	0.445	0.129	A

Focus group: Female (n= 3274) Referance grup: Male (n=3592)

In the MH analysis results, DIF was found in 8 items out of 16 questions in the 6th grade Mathematics subtest. Considering DIFs in the items, 7 of them were found as at the level of "A" and 1 of them was at the level of "B". It can be said that the item at the level of "B" worked in favor of boys.

MH results relating to the materials giving DIF in regard to the variation of gender in the 7th grade Mathematics subtests are provided in Table 6.

Table 6.
MH Analysis Results of the 7th Grade Mathematics Subtests Items

Items	α	χ^2	p	Δ-MH	SE DMF	DIF Level
1	0.667	49.991	0.000	-0.950	0.134	A
3	0.645	50.772	0.000	-1.032	0.145	B
4	0.746	21.486	0.000	-0.690	0.148	A
7	1.222	10.017	0.002	0.470	0.147	A
8	1.340	27.392	0.000	0.688	0.131	A
10	1.189	6.418	0.011	0.407	0.159	A
11	1.157	4.710	0.030	0.343	0.156	A
12	1.281	17.590	0.000	0.582	0.138	A
13	1.297	18.684	0.000	0.610	0.140	A
15	1.241	10.787	0.001	0.508	0.153	A
16	1.188	7.771	0.005	0.406	0.144	A
18	0.868	5.141	0.023	-0.333	0.145	A

Focus group: Female (n=3052) Referance group: Male (n=3061)

According to the variation of gender of items in the 7th grade Mathematics subtest, as 11 items at the level of "A" and 1 item at the level of "B", a total of 12 items included DIF. Since DIFs in 11 items at the level of "A" were negligible, only 1 item included DIF. When the Table 6 is examined, it is seen that this one question was in favor of boys.

The results of the analyses related to the materials giving DIF in regard to the variation of gender in the 8th grade Mathematics subtests are provided in Table 7.

Table 7.
MH Analysis Results of the 8th Grade Mathematics Subtests Items

Items	α	χ^2	p	Δ-MH	SE DMF	DIF Level
1	0.729	17.238	0.000	0.742	0.178	A
2	1.530	55.484	0.000	-1.000	0.134	B
3	0.747	28.748	0.000	0.687	0.128	A
4	0.804	17.367	0.000	0.512	0.122	A
5	1.261	6.292	0.012	-0.545	0.214	A
6	0.639	50.425	0.000	1.052	0.148	B
7	1.288	20.215	0.000	-0.594	0.132	A
8	1.293	30.829	0.000	-0.604	0.108	A
9	1.298	21.560	0.000	-0.612	0.131	A
10	1.153	4.740	0.029	-0.334	0.151	A
12	1.217	13.225	0.000	-0.462	0.126	A
14	1.131	6.824	0.009	-0.290	0.110	A
15	1.284	23.138	0.000	-0.587	0.122	A
16	0.771	16.930	0.000	0.610	0.147	A
17	1.366	43.240	0.000	-0.733	0.111	A
18	0.878	7.429	0.006	0.305	0.111	A
19	1.483	25.972	0.000	-0.927	0.181	A
20	0.836	16.451	0.000	0.420	0.103	A

Focus group: Female (n= 5084) Referance grup: Male (n= 4290)

In regard to the variation of gender of materials in the 8th grade Mathematics subtest, as 16 items at the level of "A" and 2 items at the level of "B", a total of 18 items included DIF. Presence of DIF at the level of A in 18 items out of 20 questions was notable. Because DIFs in 16 items were negligible, only 2 items included DIF. When the Table 7 is examined, it can be said that one of the items at the level of "B" worked in favor of boys, and the other one for girls.

Since the 6th and 16th items in the 6th grade Science and Technology subtest, 14th, 15th and 17th items in the 8th grade Science and Technology subtest, 3rd item in both 6th and 7th grade Mathematics subtests, and 2nd and 6th items in the 8th grade Mathematic subtest included gender DIF at the level of "B, these items were presented to the experts' opinion.

DIF at the level of "B" was found in total of 4 items in the 6th, 7th and 8th grades Mathematics subtests. The experts, who were asked for their opinions in order to determine whether the source of DIF in these items arose from the item influence or item bias, expressed their opinions about not seeing any item bias in the questions.

When it is considered the expert opinions on the items with DIF in the Science and Technology subtests, it can be said that the 5 items with DIF did not have any content to be bias in favor of either boys or girls.

Discussion

According to the results of the MH analysis that was conducted to see if the item function of the materials in the Science and Technology subtest changed regarding the gender difference, it was found that 10 items at the level of "A" and 2 items at the level of "B" with a total of 12 items out of the 16 questions in the 6th grade included DIF. It is seen that the 6th and 16th items containing DIF at the level of "B" worked in favor of girls and boys, respectively. Looking at the content of these items, it can be said that the item working in favor of boys was about natural and artificial monuments, whereas the item working in favor of girls was about health information and also visual comprehension reading skill. The DIF at the level of "A" was found only in 12 items out of 18 in the 7th grade Science and Technology subtests. There were 13 items with DIF at the level of "A" and 3 items with DIF at the level of "B" among the total of 20 questions in the 8th grade Science and Technology subtests. As a result of the analyses, no DIF at the level of "C" was found in the 6th, 7th and 8th grade Science and Technology subtests. It can be said that the item (14) working in favor of boys was on the subject of temperature and melting in the field of chemistry. In this item, it was asked to make an inference from what is given. Considering the items working in favor of girls, the 15th item is seen to be on the subject of cell division in the field of biology. Considering the items with DIF working in favor of girls and boys in the Science and Technology subtests, it was seen that the items in the field of physics and chemistry usually worked in favor of boys, whereas the items in the field of biology worked in favor of girls. This finding shows parallelism with the other research findings on this subject (Kalaycıoğlu & Berberoğlu, 2010; Zenisky, Hambleton, & Robin, 2004).

According to the results of the analysis conducted to determine if the items in Mathematics subtest had the differential item functioning with regard to the variation of gender, the DIF was found in the 8 questions out of 16 in the 6th grade test. While seven of these were at the level of "A", 1 item was at the level of "B". In the 7th grade in mathematics subtest, 11 items included DIF at the level of "A" and 1 item included DIF at the level of "B" out of 12 items. As for the questions in 8th grade subtest, 16 items at the level of "A" and 2 items at the level of "B", a total of 18 items within 20 questions included DIF. As in the Science and Technology subtests, there was no item with DIF at the level of "C" in the Mathematics subtests.

This finding is similar to the findings obtained from the study titled "Comparison of Generalized Progressive Linear Modeling, Logistic Regression and Likelihood Ratio Techniques in Determining the Changing Functioning of Item", conducted on the SESSPE data by Acar (2008). A certain number of items including DIF were found in both studies; however, the DIF was generally at tolerable level. The 6th item in the 6th grade Science and Technology subtest, and 3rd item in the 6th grade Mathematics subtest worked in favor of boys. Considering these two questions, the common feature in both was the nature and the environment. The fact that male students are more successful in items about nature in the field of mathematics and in the questions in the field of geometry, and included DIF in favor of male students shows parallelism with the findings obtained from the research of Abedalaziz (2010) on the investigation of the differential item functioning according to variation of gender of the items in mathematics tests. Moreover, when the Table 8 was examined, male students were seen to be more successful in all the items that included DIF in Mathematics subtest. This finding shows similarities to the studies with findings that male students are more successful in mathematics at primary education level (Geary, 1996). Looking at the items included DIF in Mathematics subtest, it is seen that the 2nd item required the algorithmic calculations and worked in favor of boys. Kalaycıoğlu and Berberoğlu (2010) demonstrated that one question measuring the algorithmic calculation in Mathematics subtest applied in the University Entrance Examination worked in favor of girls. Therefore, this finding contradicts the finding obtained from that previous study. This situation might result from the examinations performed being at different education levels. Ding, Song, and Richardson (2007) emphasized

that male students were more successful at primary education level, whereas female students were more successful at secondary education or university level especially in problem solving and application. In another research, it was explained that female students were more successful than males in mathematics and lessons based on language and parole skills at primary education level, whereas males are more successful in the field of geometry from secondary to university education (Amrein & Berliner, 2002). It can be also seen that the 6th question, which measured spatial skills, in the same test worked in favor of girls.

It was concluded that the items including DIF at the level of "B" in Science and Technology, and Mathematics subtests did not constitute bias according to the opinions of experts. Why items included DIF was due to the item impact. Furthermore, Zierky (2003) indicated that in the tests measuring success and having no possibility of prior testing, even the DIFs at the level of "B" could be ignored. Considering this case, it can be said that the Mathematics, and Science and Technology subtests within the subtests in the LDE, which performed for the selection and placement of students in the transition from primary education to secondary education, did not show any item bias. This result is similar to the findings obtained from the results of the studies titled "The Investigation of the Student Selection and Placement Examination for Secondary Education in Terms of Item Bias" conducted by Yurdugül (2003), and "The Investigation of the Student Selection and Placement Examination for Secondary Education in Terms of Differential Item Functioning According to the Students' Residential Areas" conducted by Yurdugül and Aşkar (2004).

In the current study, the item biasness in Science and Technology, and Mathematics subtests in the LDE was investigated in terms of gender variation. In line with the findings obtained from the study, it can be suggested to examine if these two subtests and other subtests show biasness depending on various variables such as socioeconomic levels and residential areas of students and school types.

References/ Kaynakça

Abedalaziz, N. (2010). A Gender- related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Assessment, 5,* 101-116

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10,* 18. Retrieved March 20, 2011, from http://epaa.asu.edu/epaa/v10n18/.

Anıl, D. ve Güzeller, C. O (2010, Mayıs). *Seviye belirleme sınavı yedinci sınıf fen ve teknoloji alt testi ile diğer alt testler arasındaki ilişkinin yol analizi ile incelenmesi.* Eğitimde ve Psikolojide Ölçme ve Değerlendirme II. Ulusal Kongresi'nde sunulan bildiri, Mersin Üniversitesi, Mersin.

Anıl, D., Güzeller, C. O., Çokluk, Ö., & Şekercioğlu, G. (2010) "Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub-test". *Procedia Social and Behavioral Sciences*, *2* (2), 5292–5298.

Acar, T. (2008). *Maddenin farklı fonksiyonlasmasını belirlemede kullanılan genellestirilmis asamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oranı tekniklerinin karsılastırılması.* Yayımlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*, *36*, 185-198.

Berberoğlu, G., Kaptan, F. ve Kutlu, Ö. (2002, Eylül). *Türkiye genelinde sekizinci sınıf öğrencilerinin fen bilgisi dersindeki üst düzey zihinsel becerilerinin incelenmesi.* V. Ulusal Fen ve Matematik Eğitim Kongresi'nde sunulan bildiri, ODTÜ, Ankara.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage.

Doğan, N. ve Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel - Haenszel, Ki-Kare ve

Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim, 33* (148), 100-112.

Ding. C. S., Song, K., & Richardson, L. (2007). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies 40* (3), 279-295.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences 19,* 229-284

Gök, B., Kelecioğlu, H. ve Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel–Haenszel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim Dergisi, 35* (156), 3-16

Guilera, G., Gomez-Benito, J., & Hidalgo, M. D. (2009). Scientific production on the Mantel-Haenszel procedure as a way of detecting DIF. *Psicothetrta, 2* (3), 492-498.

Güzeller, C. (2005). Ortaöğretim kurumları öğrenci seçme sınavının Türkçe dil yeterlilikleri açısından modellenmesi. *Kastamonu Eğitim Dergisi, 14* (2), 403-412.

Kalaycıoğlu, D. B., & Berberoğlu, G. (2010). Differential Item Functioning Analysis of the Science and Mathematics Items in the University Entrance Examinations in Turkey. *Journal of Psychoeducational Assessment.* Retrieved December 29 2010 from http://jpa.sagepub.com/content/early/2010/12/14/0734282910391623

Kristjansson , E., Aylesworth, R., Mcdowell, I., & Zumbo, D. A. (2005). Comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement, 65, 935-953.*

Kutlu, Ö. ve Karakaya, İ. (2007). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavında kullanılan testlerin faktör yapılarına ilişkin bir araştırma. *İlköğretim Online Dergisi,* 6(3), 397-410

Milli Eğitim Bakanlığı (MEB). (1998). *Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu* (OKÖSYS). Ankara: Türk Tarih Kurumu Basımevi.

Milli Eğitim Bakanlığı (MEB). (2003). *Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu (OKS) ve Özel Okullar Sınavı (ÖOS) Tanıtım Kitapçığı.* Ankara: MEB-EĞİTEK Ölçme ve Daire Başkanlığı.

Milli Eğitim Bakanlığı (MEB). (2007). "*64 Soruda OGES, ve Seviye Belirleme Sınavı Örnek Sorular*". http://oges.meb.gov.tr/docs/64_soru.pdf adresinden 24 Aralık 2010 tarihinde edinilmiştir.

Öğretmen, T. ve Doğan, N. (2004). OKÖSYS matematik alt testine ait maddelerin yanlılık analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *5* (8), 61-76.

Örs, S. (2010). *6., 7. ve 8. Sınıf Seviye Belirleme Sınavı fen ve teknoloji alt testlerinin faktör yapılarının belirlenmesi.* Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Roever, C. (2005). *"That's not fair!" Fairness, bias, and differential item functioning in language testing*. Rerieved December 10, 2010 from http://www2.hawaii.edu/~roever/brownbag.pdf.

Schumacher, R. (2005). *Test bias and differential item functioning*. Rerieved November 23, 2010, from http://www.appliedmeasurementassociates.com,pdf.

Waller, N. G. (1998) EZDIF: A computer program for detecting uniform and non uniform differantial item functioning with the Mantel-Haenszel and Logistic Regression procedures. *Applied Psychological Measurement, 22, 391.*

Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*. Yayımlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara

Yurdugül, H. ve Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın öğrencilerin yerleşim yerlerine göre differensiyel madde fonksiyonu açısından incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 27*, 268-275.

Zenisky, A. L., Hambleton, R. K., & Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item writing practices. *Educational Assessment, 9* (1), 61-78.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Zieky, M. (2003). *A DIF Primer*. Princeton, NJ: Educational Testing Service. Retrieved April 18, 2011, from http://www.ets.org/Media/Tests/PRAXIS/pdf/DIF_primer.pdf

Zumbo, D. B. (1999). *A handbook on the theory and methods of differential ıtem functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type item scores*. Ottowa: Directorate of Human Resources Research and Evaluation, Department of National Defense.