Ankara University Journal of Faculty of Educational Sciences
Year: 2020, Volume: 53, Issue: 3, 931-951
DOI: 10.30964/auebfd.760362, E-ISSN: 2458-8342, P-ISSN: 1301-3718

SES Related Differential Factors in Turkey's PISA 2015 Science Literacy Data

ARTICLE TYPE	Received Date	Accepted Date	Published Date
Research Article	06.29.2020	09.23.2020	09.30.2020

Hüseyin H. Yıldırım [1]
Bolu Abant Izzet Baysal University

Eren Ceylan [2]
Ankara University

Abstract

The purpose of this study is to investigate item characteristics that may differentially affect PISA 2015 science performance of students in lower, middle, or upper SES groups in Turkey. The Profile Analysis was used in the study. Results revealed response format and knowledge aspect of PISA 2015 science items as the two main characteristics which might lead differential performance among students from different SES background. Constructed response item format and procedural knowledge aspect have been detected as the weaknesses of low SES students, and multiple choice item format and content knowledge aspect have been detected as the weaknesses of high SES students in Turkey.

Keywords: PISA, science literacy, SES; profile analysis; DIF.

The Ethical Committee Approval: Since, an international database, which is open to all researchers is used, anapproved ethical committee decision for research is not required to be submitted for this study.

[1]Assoc. Prof., Faculty of Education, Department of Mathematics and Science Education, E-mail: hhy.yildirim@gmail.com, https://orcid.org/0000-0001-9924-8281
[2]*Corresponding Author*: Assoc. Prof., Faculty of Educational Sciences, Department of Elementary Education, E-mail: ernceylan@gmail.com, https://orcid.org/0000-0001-8244-8260

One of the crucial indicators of educational fairness is the extent to which educational systems equip students irrespective of their socio-economic status (SES). However, results from the Programme for International Student Assessment (PISA) 2015 once again reveal the long-term-known fact that most of the educational systems in the world are not even close to this SES-free educational performance ideal. Among the OECD (The Organisation for Economic Co-operation and Development) countries, 13 percent of students' PISA 2015 science performance variation is related by their SES score. Besides, almost 63 percent of between-school variation in science performance is explained by students' and schools' socio-economic statuses (OECD, 2016b).

In terms of educational fairness, Turkey's situation is, on average, better than the OECD countries. The percentage values given above are 9% and 49% for Turkey, respectively (OECD, 2016b). However, when it comes to the percentage of students from disadvantaged socio-economic backgrounds, among 69 PISA 2015 participant countries, only two, Indonesia and Viet Nam, are in a poorer condition than Turkey. Fifty nine percent of students in Turkey are in the bottom decile of the international SES distribution. To give more of a sense of this percentage value, it may be worth adding the information that it is 74% in Indonesia, 76% in Viet Nam, 12% on average in OECD countries, and only 1% in Iceland. Correspondingly, the percentage of students in Turkey who are in the top decile with respect to the international SES index is only 4, whereas the percentages are 1%, 2%, 27%, and 57% in Indonesia, Viet Nam, OECD countries, and Iceland, respectively. Furthermore, a crucial finding is that, in Turkey, there are more than 100 points difference between the top decile and the bottom decile student groups' PISA 2015 mean science scores (OECD, 2016b).

The achievement gap between groups of students from different social-backgrounds, as the one indicated above between low and high SES groups of students in Turkey, are by no mean unexpected when the findings from several studies are considered. For example, Darling-Hammond (2014) underlies the large disparity in math, reading, science, and problem solving scores of low-income and high-income children in the United States. Lenkeit and Caro (2014), in their study investigating data from 4 cycles of PISA from 2000 to 2009, show that performance comparisons among educational systems without considering their socioeconomic differences are very much unfair.

This highly significant relationship between socioeconomic status and test performance of students also makes it crucial to understand in what ways students' socioeconomic status is related to their performance as measured in certain assessment programmes, such as in PISA. These sort of associations may generally be grouped into two types. Evidently, one of the types is the explicit effect in which being in a higher level of socioeconomic status effects test performance due to students' higher level at the psychological construct being measured by the tests. Simply that is, students from higher socioeconomic backgrounds are more likely to have higher educational and social opportunities, and having higher opportunities may lead to a better learning, and thus a better test performance.

The second type of effects, in line with the differential item functioning (DIF) literature (e.g., Camilli and Shepard, 1994), can be called as differential or implicit effect, to indicate the factors related to students' test performance and yet not captured by the test's measurement model (Verhelst, 2012). Here is a hypothetical example. PISA has changed its assessment mode in 2015 from paper to computer based and delivered the tests on computers. What if socio-economically advantaged students are more accustomed to computers than their disadvantaged peers, and consequently manage to handle computerized tasks more effectively in responding to science literacy test items? In such a case, disadvantaged students would perform poorer than their advantaged peers in science literacy items not because their science literacy level is lower, but because their computer skills level is. However, as the test is intended to measure science literacy, any score difference between groups in such a case would be devoted to a genuine difference between science literacy levels of students unless possible differential factors related to the achievement are investigated.

In test theory, such potential implicit factors that are related to students' test performance and yet not accounted by the tests' measurement model are named as differential factors (Camilli and Shepard, 1994). Studies focused on such implicit effects which might contribute a better understanding of the disparities between SES groups, have been underestimated when the SES literature is reviewed. Our aim in this present study is to investigate possible SES related differential factors in Turkey's PISA 2015 science data. To this purpose, we employ a recently developed innovative technique Profile Analysis (PA) (Verhelst, 2012).

PA is not an exploratory method. Thus, it requires hypotheses on item characteristics which might be related to the differential functioning of item bundles among groups of students as defined in the analyses. The hypotheses in this current study are grounded on the PISA 2015 Science Framework which provides definite item specifications. In this context, the possible SES related differential factors in this research are tested on the following item specifications: response format (constructed response, complex multiple choice, multiple choice); competency aspect (explaining phenomena, interpreting data); knowledge aspect (content, procedure); cognitive demand (low, middle) and content (living systems, earth and space, physical systems) (OECD, 2016b).

In sum, the purpose of this study can be rephrased as investigating whether PISA 2015 science items' response format, competency level, knowledge aspect, cognitive demand, and content are related to any differential performances among students from different SES backgrounds in Turkey.

Method

In this section, the population and the sample of the study, PISA 2015 test design and science assessment framework, and the analyses used in this study are described.

Since, an international database, which is open to all researchers is used, anapproved ethical committee decision for research is not required to be submitted for this study.

Universe-Sample

The major domain assessed in PISA 2015 is scientific literacy. Seventy-two countries and economies participated in the study. The population of PISA-participating students is defined by technical standards given in the OECD (2017) document.

In each country, a two-staged sampling design in which first the schools and then the students are selected is used in PISA 2015. In Turkey, 5895 15-year-old students from 187 schools participated in the study. These students were selected to represent 925,366 15-year-old students who were enrolled at grade 7 or above at the time of the study. Ninety three percent of the students selected for the study in Turkey were at 9th or 10th grade levels. Due to some reasons (e.g., functional disability) 36 students were excluded by PISA from the analysis. In addition, in this current study, 18 students having an extreme response format (i.e., all items correct or incorrect) were excluded as the PA cannot be carried on for these types of students. Thus, for the current study 5841 students' responses to PISA 2015 science items were used.

PISA 2015 Test Design and Science Assessment Framework

PISA 2015 science test forms consisted of 184 computer-based test items in total, which were grouped in 12 clusters. Two of these clusters as determined through a multistep rotated test design process were directed to each student (OECD, 2017). In this way, test forms received by each student included 25 to 35 science items. The 184 PISA science items were developed with respect to PISA's science assessment framework.

PISA 2015 science assessment framework designates the test items with respect to the construct PISA aims to measure, and also the assessment characteristics items posses. Table 1 shows the item categories in this context. Further explanation on these categories follows the table.

The major construct involved in PISA 2015 is scientific literacy. In the most general sense, although science and technology differ with respect to purposes, processes, and products, scientific literacy in PISA compromises the knowledge of both science and science-based technology. The components of the scientific literacy are defined by four interrelated aspects: context, knowledge, competency, and attitudes (OECD, 2016b). Two predominant aspects of knowledge and competency are involved in this current study.

The competency aspect of PISA's scientific literacy deals with abilities students should possess in order to understand and engage in critical discussions about issues that involve science and technology. In this context, scientific literacy in PISA 2015 is outlined by the three competencies: explain phenomena scientifically; evaluate and design scientific enquiry; and interpret data and evidence scientifically.

Table 1

Categorizations of PISA 2015 Science Items With Respect to PISA 2015 Assessment Framework

Scientific Literacy Item Categories				
Organization of the Domain		**Assessment of the Domain**		
Competency Aspect	Knowledge Aspect	Response Format	Cognitive Demand	Content
Explain (89)	Content (98)	Multiple Choice (54)	Low (56)	Physical Systems (61)
Evaluate (39)	Procedure (60)	Complex Multiple Choice (66)	Middle (113)	Living Systems (74)
Interpret (56)	**Epistemic (26)**	Constructed Response (64)	**High (15)**	Earth and Space Systems (49)

Note. The number of items in each category is given in parenthesis. In the highlighted categories, PA was not conducted due to the insufficient number of items

The first competency, explain phenomena scientifically, deals with students' ability to offer and evaluate explanations for natural and technological phenomena, and recognize their implications for society. The second competency, evaluate and design scientific inquiry, is about identifying and evaluating the quality of scientific investigations and propose scientific ways to answer the questions. The last competency, interpret data and evidence scientifically, defined in the assessment framework refers to the ability to evaluate whether the conclusions are appropriate from a scientific perspective.

The competencies defined above all require knowledge. Thus, knowledge is defined as the second aspect in the organization of the PISA's scientific literacy domain. This aspect involves three distinctive but relevant elements: content knowledge, procedural knowledge, and epistemic knowledge. Knowledge of facts, concepts, ideas and theories about science is referred to as content knowledge. This sort of knowledge is relatively more engaged to explain phenomena scientifically competency than the other two competencies. Procedural knowledge deals with the procedures used in science to establish scientific knowledge. Indeed, procedural knowledge refers to understanding how scientific knowledge is established what has been called as: the nature of science (Lederman, 2006).

Repeating measurements to minimize error and reduce uncertainty, the control of variables, and standard procedures for representing and communicating data is referred to as procedural knowledge. Finally, epistemic knowledge involves understanding of the rationale for the common practices of scientific inquiry and some fundamental scientific terms such as theories, hypotheses, and models. Both procedural and epistemic knowledge are more at work in PISA items on scientific enquiry.

The second dimension of PISA 2015's science domain organization is about the assessment of the scientific literacy. This facet is mainly about the item characteristics, and determines their three characteristics: response format, cognitive demand, and content (i.e., main area of knowledge of science).

Items used in PISA 2015 science tests are one of the three classes of response format: simple multiple choice, complex multiple choice, and constructed response. The first and the last classes are probably familiar to the reader. If additional information for the format complex multiple choice is required these type of items basically require selection of more than one response among the list of options.

The cognitive demand for an item is introduced as a new key feature in PISA 2015 study. In PISA 2015 assessment framework, it was strongly emphasized that items' difficulty is not only related to the cognitive demand of the question. For example, an item that only requires a cognitive demand of recalling would have a high difficulty (i.e., only a small group of students might give the correct answer) since it involves a specific knowledge that is not well known. In PISA 2015, an adapted version of Webb's Depth of Knowledge (Webb, 1997) which offer a taxonomy taken into account of the verbal cues used in the questions (e.g. analyse, arrange, compare) and required depth of knowledge was used. Based on this taxonomy, three cognitive demand classes defined are the low, the middle, and the high demanding items. Items to carry out a one-step procedure are labeled as low demanding items. Items that require using and applying conceptual knowledge are of medium level. The most demanding items at a high level in PISA require analyzing complex information or data.

Finally, the content categorization groups the PISA items involves three main areas such as physical systems (e.g., structure of matter, motion and forces), living systems (e.g., cells, ecosystems), and earth and space systems (e.g., structures of the Earth systems, Earth in space).

Socio Economical Status in PISA 2015

In PISA, students' socio-economic status is defined within the framework of the education system's fairness. In this context, in a fairer education system, students' achievement should be more likely to result from their abilities and effort than their socio-economic status (OECD, 2016b).

PISA estimates students' socio-economic status by the index of economic, social and cultural status (ESCS). This index is derived from the variables related to students' parents' education level and their occupations, the number of educational resources (e.g., a desk to study, a quiet place to study, a computer usable for school work, books, educational software, a dictionary), and the number of certain home possessions (e.g., a room of student's own, a link to the internet, televisions, cars, rooms with a bath, musical instruments) in the home of students (OECD, 2016b). PISA derives ESCS index from these variables via Principal Component Analysis. ESCS scores are then standardized to have a mean of 0 and standard deviation of 1 for the OECD countries (OECD, 2017).

In this current study, the groups for the Profile Analyses are determined with respect to students' ESCS scores. Students in the lower quartile in Turkey with respect to their ESCS score are specified as the low SES group. Similarly, students in the upper quartile are grouped as the high SES group, and the rest of the students are labeled as the middle SES group. The number of students analyzed in each group has been 1492, 1411 and 2938 respectively.

Profile Analysis

Profile analysis is a technique based on comparing test takers' observed and expected performances on a group of items of a test (Verhelst, 2012). In the analysis, expected performances are calculated conditionally given the individuals' total test scores and the parameters of test items. The item parameters should be estimated via an Item Response Theory (IRT) based measurement model, such as the Rasch Model or the 2-parameter IRT model (van der Linden and Hambleton, 1997).

The basic statistics in profile analysis is the difference between an individual's observed and expected performances on a subgroup of test items. This difference value is named as a deviation profile. A positive deviation value indicates that the student performs better than expected on the subgroup of test items as specified in the profile analysis. Similarly, a negative value points to a performance lower than expected based on the measurement model. The basis for this rationale is that the expected performance of individuals depends only on the item parameters given the individual's total test score, as proved by Georg Rasch (1960), a distinguished Danish psychometrician. Thus, a significant deviation profile indicates that the measurement model misses some factors in the test data which affect students' performance at the subgroup of test items sharing a common property. The profile analyses in this study were conducted through ProfolieG software developed by Norman Verhelst (2012). The software also checks the statistical significance of the deviation values through a Chi_square based built-in algorithm.

We carried out the PA in this current study with respect to the following fundamentals which also outlines a profile analysis procedure.

1. Prior to the profile analysis, item parameters should be calibrated via a measurement model which well fits the test data. In this current study, item parameters were estimated using Turkey's PISA 2015 test data only. For the item calibration in this study, One-parameter Logistic Model (OPLM) and the corresponding software was used (Verhelst and Glas, 1995).

2. Grouping the test items in profile analysis is not an arbitrary process. Items should be clustered with respect to some valid hypotheses. In this current

study, PISA 2015 science assessment framework was used as a basis to classify the 15 item categories as given in Table 1. Twelve of the categories were investigated in the profile analyses.

3. Profile analysis basically produces results at an individual level. In this current study, for each of the 12 item clusters, individuals were identified to have one of the four possible deviation types: positive-significant, positive-non-significant, negative-significant, negative non-significant (As explained before, a positive profile indicates higher observed performance than the expected, and significance depends on a built-in PA statistical test).

On the other hand, PA results for the individuals of common property can also be aggregated to get group level results. In this study, results were aggregated at the low, middle, and high SES groups. To this purpose, for each of the item categories, the number of students of the positive deviation type observed in each group was determined. In addition, the number of positive-deviation-type students the measurement model allows (named in the analyses as: expected) were detected. It is important to note that if a student is not of a positive deviation type in an item category, she is of positive deviation type in the rest of the items. Thus, each item category was investigated with its complementary category (e.g., open response type items versus not open response type items).

At the group level, a Chi-square test with 1 degrees of freedom was used to statistically test whether certain types of students dominated the group. In other words, for each of the low, the middle and the high SES group of students, a chi-square test of goodness-of-fit was performed to determine whether the observed number of individuals was in line with the expected number of individuals. These analyses were conducted separately for each of the 12 item categories.

Table 2 presents an example for such an aggregated result at the group level for the PISA 2015 science items of constructed response type. Please note that when a subgroup of the test items are marked as of a specific type (here, constructed response type items indicated as CR+), the rest of the test items automatically determines another subgroup of items of not this type (indicated as (CR-)).

Table 2

Observed Versus Expected Number of Students Doing Better Than Estimated in Open Response Type PISA 2015 Science Test Items

SES Groups	**Observed/Expected**	**CR+**	**CR-**	**Total**
Low	Observed	657	835	1492
	Expected	696.01	795.99	1492
Middle	Observed	1349	1589	2938
	Expected	1383.33	1554.67	2938
High	Observed	681	730	1411
	Expected	676.8	734.19	1411

Note. CR+: Items of constructed response type; CR-: Items of not constructed response type

The numbers in Table 2 for the low SES group point to the following: The measurement model allows 696.01 individuals to have a positive deviation profile (i.e., performing relatively better) in the subgroup of constructed response type items. However, with respect to the observed performances, there are 657 such students. Thus, in the low SES group, the number of students who are estimated to perform relatively better in the CR+ type items falls below the expectation by 39 students.

This difference value can also be set as a percentage of the total group, -2.61% (657-696.01)/1492). This percentage value is called the excess percentage (EP) for the corresponding group of items. Please note that EP for the CR- type items is +2.61% as CR+ and CR- type items are complementary by design. A chi-square test for the difference between the observed and the expected number of individuals in a group can also be conducted. For example, this value for the low SES group in Table 2 is, $\chi^2(1, 1492) = 4.09$, $p < .05$. In this current study, EP values are used to investigate the results of PA analyses in the 3 SES groups for the 12 item categories.

Results

PISA calibrates the test items using the combined data from all participant countries. The programs named 'mdltm' and 'ConQuest' are used in the calibration processes (von Davier, 2005; Wu, Adams and Wilson, 1997). Within the scope of this current study, 184 PISA 2015 science items were recalibrated using Turkey's test data only. This Turkey specific test calibration was conducted by the software OPLM available to the authors (Verhelst and Glas, 1995).

To test that the OPLM software produces valid item parameter estimates as compared to the PISA analyses, calibrations with OPLM were also conducted using the combined test data from all PISA 2015 participant countries. The Pearson correlation coefficient between the OPLM-estimated and PISA-published item difficulty parameters for the 184 science items showed that OPLM also generates valid estimates for the PISA 2015 science test data ($r = .97$, $p < .01$).

The profile analysis was conducted for 12 item categories in three SES groups of students in Turkey. The ESCS scores in each of the three SES groups have been summarized in Table 3.

Table 3

Descriptive Statistics for ESCS Scores of Individuals in the Three SES Groups

SES Groups	N	Min.	Max.	Mean	Std. Dev.
Low	1492	-5.13	-2.31	-2.87	0.41
Middle	2938	-2.31	-0.61	-1.49	0.49
High	1411	-0.61	3.12	0.14	0.54

As explained in the preceding section, PA analyses produce a huge amount of output for individuals' conditional and observed performances on a group of test items. Excess percentage (EP) is an appropriate statistics to summarize all these findings at a group level in a compact way. Table 4 provides the EP values to sum up the findings of this current study. Please note that details for the calculation of the EP value -2.61% for the constructed response item category in the low SES group has been given in the upper row of Table 2. The rest of the EP values in Table 4 have been calculated in the same way.

Table 4

Excess Percentages (EP Values) as a Summary of the Profile Analyses in 3 SES Groups for the 12 Item Categories

Categorization	**Categories**	**EP values for this category of items (+ Type)**		
		SES GROUPS		
		Low	**Middle**	**High**
Response Format	Constructed Response	**-2.61%**	-1.17%	0.30%
	Complex Multiple Choice	**3.95%**	0.01%	-0.79%
	Simple Multiple Choice	-0.90%	1.17%	-2.33%
Competency Aspect	Explain	0.91%	-0.89%	-0.91%
	Interpret	-0.23%	1.32%	-0.25%
Knowledge Aspect	Content	1.86%	0.50%	-1.49%
	Procedural	-2.48%	0.76%	0.94%
Cognitive Demand	Low	0.66%	-0.16%	-0.22%
	Middle	0.11%	0.22%	-1.08%
Content	Living Systems	0.21%	-0.18%	0.59%
	Earth & Space	0.78%	-0.33%	1.75%
	Physical Systems	0.39%	0.80%	-1.40%

Note. EP values in bold points to a statistically significant difference at .05 level between the observed and the expected number of individuals for the corresponding PA in the group.

Discussion, Conclusion and Suggestions

Results from this current study show that students' performance in a number of subcategories of PISA 2015 science test items is related to their SES status. Among these categories, response format of the test items is the category with the biggest EP value, which also reaches a statistically significant level. However, statistical significance is only one side of the story (Yıldırım and Yıldırım, 2011). The design of the current study required dividing the students into three SES groups which not only decreased the sample size but also the power of the statistical tests. In addition, due to the incomplete test design used in PISA tests, a student replies only a small portion (about 15%) of the available test items (OECD, 2017). Thus, in analyses like PA which requires separating test items as well into subgroups, due to the loss of power, statistical significance would be relatively more difficult to reach despite the considerable effect sizes.

Consequently, evaluating the results in Table 4 with respect to their effect size would be more suitable within the context PA as conducted in this study. EP values in Table 4 can also be regarded as effect sizes in an absolute sense. If these 36 EP values are ordered with respect to their value, the four EP values above 1.20% lay in the upper decile while the four values below -1.50% make up the bottom decile. Although an arbitrary approach, this may contribute to detecting some significant findings in PA which otherwise would be lost in the lack of statistical power. Table 5 gives the eight EP values on the edges.

Table 5

Excess Percentages in the Bottom and the Top Deciles, and the Corresponding Item/Group Category

Negative EP Values (Weaknesses)		**Positive EP Values (Strengths)**	
EP	Group/Item Category	EP	Group/Item Category
-2.61%	Low SES Constructed Resp.	**3.95**%	Low SES C. Multiple Choice
-2.48%	Low SES Procedural Know.	1.86%	Low SES Content Know.
-2.33%	High SES Multiple Choice	1.75%	High SES Earth & Space
-1.49%	High SES Content Know.	1.32%	Middle SES Interpret

The results reveal that the most significant finding in PA is related to the category response format. The item response format is also among the variables that is commonly studied in research on detecting students' test performance related variables (e.g., Liou and Bulut, 2017; Mingo, Chang and Williams, 2018). These studies provide evidence for the differential function of item format. In line with these findings, the result of this current study also reveals that item format effects students' science performance differentially with respect to their SES background.

In terms of response format the EP values indicate that low SES group students are performing better than expected in items of complex multiple-choice format. On the other hand, students of low SES group are performing below the expectations on the constructed response format items. First thing that comes to mind is naturally that constructed response items require students to communicate their ideas, and students from lower SES group may lack this ability lowering their science performance. However, we should also note that PA is an associational technique and does not offer causal relationships.

Another remarkable finding on the response format is the relatively poor performance of high SES group students on multiple-choice items. In Turkey, all high stakes tests (e.g., transition to high school examination, university admission exams) are of multiple-choice format. Thus, students in Turkey are accustomed to this response type items, and this makes such a result surprising. However, students from high SES background have broader educational opportunities such as continue at private universities and having an education abroad. Therefore, these students may lack some testing skills contributing to detect the correct response of a multiple-choice item. This comment may be extended for the items requiring content knowledge as well, which are frequently used in high stakes tests in Turkey. Results show that science performances of high SES group students are below the expectations in this sort of items too.

On the other hand, this relatively lower performance of high SES group on multiple choice items does not mean that high SES group students' performance falls behind the performance of low or middle SES group students. In thinking of the results, one should be aware of that what is analyzed in PA is students' conditional performance. For example, positive EP value of 1.86% in Table 5 indicates that low SES students perform better than expected on items requiring content knowledge. However, mean PISA 2015 content knowledge items based science performances of low, middle and high SES group students in Turkey are 401.32, 420.96, and 459.46, respectively. These average performances show that students of low SES group are performing lower than the other two SES group of students. On the other hand, PA reveals that given the total performance level of the students and the difficulty level of the items, even this low performance of low SES group individuals on content knowledge items is beyond the expectations in general. These sort of positive differential performances are indicated as strength in Table 5. Similarly, negative differential performances are indicated as weaknesses.

In this manner, another weakness drawing attention in Table 5 is the low SES students' performance in procedural knowledge items. As it was mentioned before, procedural knowledge type of questions are strongly related to the concept of nature of science (NOS). Although there is no agreement among the researchers about the definition of nature of science, it has been described as "the epistemology of science, science as a way of knowing, or the values and beliefs" (Abd-El-Khalick and Lederman, 2000, p. 666). In PISA 2015, explaining scientific phenomena competency requires content knowledge of science. On the other hand, the second and third competencies demand an understanding of how knowledge is constructed is strongly related to understanding the nature of science. There is an agreement in the literature that scientific knowledge is tentative, empirically based, subjective, comprises human inference, imagination, creativity, and socially and culturally embedded (Lederman, Lederman, and Antink, 2013).

In a study carried out on 6th to 8th grade level students in Turkey, Hacıeminoğlu, Ertepınar, Yılmaz-Tüzün, and Çakır (2015) revealed that the relationship between students' science achievement and their NOS views, especially the empirical NOS, was weaker in low SES schools than the relationship in high SES schools. The students' achievement level in high and in low SES schools was matched in this study. The study also showed that students in high SES schools were likely to have higher level of tentative NOS, whereas students in low SES schools were likely to have lower level of empirical NOS.

Thus, when it is considered that procedural knowledge items in PISA 2015 are related to students' views of NOS, weakness of low SES students on procedural knowledge items as detected in this current study might be due to the disadvantage of low SES students in understanding the nature of science. Moreover, these results may indicate the lack of some opportunities and facilities such as science related extracurricular activities, science club activities, and well-equipped laboratories that foster students' NOS views and also their procedural knowledge in low SES schools.

All but one strengths or weaknesses in Table 5 are about item format or knowledge aspects. The exception is the content of Earth and Space for which high SES group's strength is observed. There are studies in the literature which might shed a light on this finding. National Assessment Governing Board (2014) reports that the latest technological developments, such as professional telescopes, visualization tools, and models found at science and technology centers, web based programs to access the satellite images of the Earth and Space, shuttle's camera have also enriched how students learn about Earth and Space. But it is highly possible that low SES students are disadvantaged in reaching these sort of costly assistance. For example, in PISA 2015, it was found that some extracurricular activities such as joining science clubs or competitions are more offered on average in advantaged schools than the disadvantaged ones across the OECD countries. Especially in Turkey, schools that are at the top quarter with regard to SES offered almost as twice science club activities as offered in schools at the bottom quarter (OECD, 2016b). On the other hand, studies report that extracurricular activities such as participating in a science club, competitions, and field trips, visiting museums, zoos, science and technology centers improve students' science learning outcomes positively (Bellipanni and Lilly, 1999; OECD, 2016c). Thus, such opportunities students of high SES background may reach would be the reason for this differential performance detected on Earth and Space items.

In sum, all but one of the EP values in Table 5 are detected either in low or in high SES groups. This may indicate that the more obvious the SES difference among the students the more likely the test items to function differentially for the students. In addition, all but one of the EP values in Table 5 are regarding the response format or knowledge aspect categorizations. Thus, the results of this study provide empirical-based evidence that science items' response format and knowledge aspect may interact with students' SES backgrounds in Turkey.

One extension for this study would be to re-conducting the analyses in other countries participated in PISA 2015 and providing aggregated results for the countries. This extension may also contribute to understanding the nature of SES. For example, it would be informative to understand if relatively weaker performance of low SES group students on constructed response items, as detected in this study for Turkey, is specific to some countries or a general phenomenon. Another extension of this study might be to other subject areas such as mathematics.

Lastly, it may be worth commenting on how to group students with respect to their ESCS scores in similar studies. This may especially be of critical importance in comparative studies among countries. Grouping students regarding the quartiles, as we did in this study, may create very diverse SES groups among the countries. For example, in this study, the mean ESCS score in Turkey even for the high SES group is 0.14, which is almost the average OECD level. Thus, for international comparisons, a criterion based cut-off scores for the ESCS index to assembly the SES groups would be a more accurate option.

References

Abd-El-Khalick, F., and Lederman, N. G. (2000). Improving science teachers' conceptions of the nature of science: A critical review of the literature. *International Journal of Science Education*, 22 (7), 665-701. doi: 10.1080/09500690050044044

Bellipanni, L. J., and Lilly J. E. (1999). What have researchers been saying about science fairs? *Science and Children*, *36*(8), 46-50. Retrieved from https://search.proquest.com/openview/662ea6b47b0f3ce5974b0c9595f9da08/1?pq-origsite=gscholar&cbl=41736

Camilli, G., and Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publications.

Darling-Hammond, L. (2014). What can PISA tell us about U.S. education policy? *New England Journal of Public Policy*, *26*(1), 1-14. Retrieved from http://scholarworks.umb.edu/nejpp/vol26/iss1/4

Hacıeminoğlu, E., Ertepınar, H, Yılmaz-Tüzün, O., and Çakır, H. (2015). Students and school characteristics related to elementary school students' views of the nature of science. *Education 3-13 International Journal of Primary, Elementary and Early Years Education 43*(6), 700-721, doi: 10.1080/03004279.2013.865655

Lederman, N. G. (2006). Nature of science: Past, present and future. In S. Abell and N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831-879). Mahwah, NJ:Lawrence Erlbaum.

Lederman, N. G., Lederman, J. S., and Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *International Journal of Education in Mathematics, Science and Technology, 1*(3), 138-147. Retrieved from https://ijemst.net/index.php/ijemst/article/view/19

Lenkeit, J., and Caro, D. H. (2014). Performance status and change – measuring education system effectiveness with data from PISA 2000–2009. *Educational Research and Evaluation, 20*(2), 146-174. doi: 10.1080/13803611.2014.891462

Liou, P. Y., and Bulut, O. (2017). The effects of item format and cognitive domain on students' science performance in TIMSS 2011. *Research in Science Education, 50*, 99-121. doi: 10.1007/s11165-017-9682-7

Mingo, M. A., Chang, H. H., and Williams, R. L. (2018). Undergraduate students' preferences for constructed versus multiple-choice assessment of learning. *Innovative Higher Education, 43*(2), 143-152.

National Assessment Governing Board. (2014). *Science framework for the 2015 national assessment of educational progress*. Washington, DC: U.S. Government Printing Office.

Organisation for Economic Co-operation and Development. (2016a). *PISA 2015 assessment and analytical framework: science, reading, mathematics and financial literacy*. Paris: PISA OECD Publishing.

Organisation for Economic Co-operation and Development. (2016b). *PISA 2015 Results (Volume I): Excellence and equity in education*. Paris: PISA OECD Publishing. doi: 10.1787/9789264266490-en

Organisation for Economic Co-operation and Development. (2016c). *PISA 2015 Results (Volume II): Policies and practices for successful schools*. Paris: PISA OECD Publishing. doi: 10.1787/9789264267510-en

Organisation for Economic Co-operation and Development. (2017). *PISA 2015 technical report*. Paris: OECD Publishing.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute of Educational Research (Expanded edition, 1980, Chicago: The University of Chicago Press).

van der Linden, W. J., and Hambleton R. K. (1997). *Handbook of modern item response theory*. New York: Springer.

Verhelst, N. D. (2012). Profile analysis: A closer look at the PISA 2000 reading data. *Scandinavian Journal of Educational Research, 56*(3), 315-332. doi: 10.1080/00313831.2011.583937

Verhelst, N. D., and Glas, C.A.W. (1995). One-parameter logistic model. In G. Fischer and I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 215–237). New York: Springer.

von Davier, M. (2005). *A general diagnostic model applied to language testing data* [Research Report No. RR-05-16]. Princeton, NJ: Educational Testing Service.

Webb, N. L. (1997). *Research monograph number 6: Criteria for alignment of expectations and assessments on mathematics and science education*. Washington, D.C.: Council of Chief State School Officers. Retrieved from https://files.eric.ed.gov/fulltext/ED414305.pdf

Wu, M. L., Adams, R. J, and Wilson, M. R. (1997), *ConQuest: Multi-aspect test software* [Computer program]. Camberwell, Australia: Australian Council for Educational Research.

Yıldırım, H. H. and Yıldırım, S. (2011). Hipotez testi, güven aralığı, etki büyüklüğü ve merkezi olmayan olasılık dağılımları üzerine. *İlköğretim Online*, *10*(3), 1112-1123.

The Ethical Committee Approval

Since, an international database, which is open to all researchers is used, anapproved ethical committee decision for research is not required to be submitted for this study.

Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi
Yıl: 2020, Cilt: 53, Sayı: 3, 931-951
DOI: 10.30964/auebfd.760362, E-ISSN: 2458-8342, P-ISSN: 1301-3718

Türkiye'de PISA 2015 Fen Maddelerinde SES'le İlişkili Yanlı Faktörler

MAKALE TÜRÜ	Başvuru Tarihi	Kabul Tarihi	Yayım Tarihi
Araştırma Makalesi	29.06.2020	23.09.2020	30.09.2020

Hüseyin H. Yıldırım [1]
Bolu Abant İzzet Baysal Üniversitesi

Eren Ceylan [2]
Ankara Üniversitesi

Öz

Bu çalışmanın amacı Türkiye'de alt, orta veya üst sosyoekonomik statüdeki (SES) öğrencilerin PISA 2015 fen testindeki performansını yanlı olarak etkileyen madde özelliklerini araştırmaktır. Çalışmada profil analizi kullanılmıştır. Elde edilen sonuçlar PISA 2015 fen testinde, farklı SES düzeylerine sahip öğrenciler arasında yanlılığa neden olabilecek temel madde özelliklerinin cevap formatı ve bilgi boyutu olduğunu göstermektedir. Genel olarak, cevabın yazılmasını gerektiren ve prosedür bilgisiyle ilgili maddelerin düşük SES düzeyindeki öğrencilerin aleyhine işlediği görülürken, çoktan seçmeli ve içerik bilgisi gerektiren maddelerin de üst sosyoekonomik düzeydeki öğrencilerin zayıflığı olduğu belirlenmiştir.

Anahtar sözcükler: PISA 2015, fen okuryazarlığı, SES, profil analizi, DIF.

Etik Kurul Kararı: Bu araştırmada, tüm araştırmacılara açık, uluslararası veri tabanında yer alan veriler kullanıldığından etik kurul kararı gerektirmemektedir.

[1]Doç. Dr., Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Matematik Eğitimi Anabilim Dalı, E-posta: hhy.yildirim@gmail.com, https://orcid.org/0000-0001-9924-8281
[2]*Sorumlu Yazar*: Doç. Dr., Eğitim Bilimleri Fakültesi, Temel Eğitim Bölümü, Sınıf Eğitimi Anabilim Dalı, E-posta: ernceylan@gmail.com, https://orcid.org/0000-0001-8244-8260

Eğitimde eşitliğin önemli göstergelerinden biri, eğitim sistemlerinin öğrencileri sosyoekonomik statülerinden (SES) bağımsız olarak yetiştirebilme derecesidir. Ne var ki, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2015'in de gösterdiği üzere, birçok eğitim sistemi SES'ten etkilenmeyen eğitim performansı idealinin uzağındadır. OECD ülkeleri içinde, öğrencilerin PISA 2015 fen performanslarındaki varyansın ortalama olarak % 13'ü öğrencilerin sosyoekonomik statüleriyle ilişkilidir. Bununla beraber, okullar arası fen performans farklılığının yaklaşık % 63'ü öğrencilerin ve okulların SES düzeyleriyle açıklanmaktadır (OECD, 2016b).

Öğrencilerin sosyoekonomik statüleri ve test performansları arasındaki bu ilişki düşünüldüğünde SES'in, PISA gibi uluslararası çalışmalarda öğrenci performanslarıyla ne şekilde ilişkili olduğunun anlaşılması da önem kazanır. Bu ilişkiler genel olarak iki gruba ayrılabilir. İlk grupta, öğrencinin sosyoekonomik avantajından ötürü testte ölçülen becerilere daha fazla sahip olmasını sağlayan etmenlerden kaynaklı ilişkiler yer alır. Bunlara genel olarak aşikar veya açık etkiler diyebiliriz. İkinci grupta ise, DIF (differential item functioning) alanyazınıyla uyumlu olarak, öğrencilerin test performanslarını etkilemekle birlikte, kullanılan ölçme modelinin belirleyemediği etkenlerden kaynaklanan, bu nedenle saklı veya örtük olarak belirtilebilecek etkiler yer alır (Verhelst, 2012). Test kuramında (teorisinde), öğrencilerin test performansıyla ilişkili olan fakat kullanılan ölçme modeliyle açıklanamayan bu tür potansiyel saklı etmenler, farklı veya yanlı işleyen faktörler genel başlığı altında incelenir (Camilli ve Shepard, 1994).

Amaç ve Önem

Çalışmanın amacı, öğrencilerin PISA 2015 fen testindeki performansını etkileyen ve SES'le ilişkili görünen saklı etmenleri Türkiye özelinde incelemektir. Analizlerde, bu amaçla geliştirilmiş yenilikçi bir teknik olan Profil Analiz (PA) kullanılmıştır (Verhelst, 2012).

Çalışmanın önemini varsayıma dayalı bir örnekle şöyle açıklanabilir: Bilindiği üzere PISA 2015'te, kağıt-kalemden bilgisayara geçildi ve testler bilgisayar ortamında uygulandı. Peki, sosyoekonomik olarak üstünlüğü olan (avantajlı) öğrenciler, sınırlılığı olan (dezavantajlı) gruptaki akranlarına kıyasla bilgisayara daha fazla alışkınlarsa ve bu da onların fen sorularını yanıtlarken bilgisayar tabanlı işleri daha etkili yürütmelerine neden oluyorsa ne söylenebilir? Böyle bir durumda, avantajlı ve dezavantajlı gruplar arasında, bilgisayara daha fazla aşina olmaktan kaynaklanan performans farkı da öğrenciler arasındaki fen okuryazarlığı farkı gibi görünecektir. Görüldüğü üzere, öğrenci puanlarındaki farklılığın gerçek düzey farklılığına atfedilebilmesi için, test performansını etkileyebilecek saklı etmenler de incelenmelidir.

PA açımlayıcı bir yöntem değildir. Dolayısıyla, incelenen gruplar arasında yanlılığa neden olabileceği düşünülen madde özellikleri üzerine hipotezler gerektirir. Bu araştırmanın hipotezleri, madde özelliklerinin betimlendiği PISA 2015 fen değerlendirme çerçevesine göre oluşturulmuştur. Bu bağlamda, çalışmada, farklı SES grupları arasında yanlılık olup olmadığı şu madde özelliklerine ve sayılarına göre incelenmiştir: cevap formatı (cevabı yazılan - 64 soru, karmaşık çoktan seçmeli – 66 soru, çoktan seçmeli – 54 soru), yeterlik boyutu (olayları açıklama – 89 soru, veriyi yorumlama – 56 soru), bilgi boyutu (içerik – 98 soru, süreçsel - 60), bilişsel gereklilik (düşük – 56 soru, orta – 113 soru) ve içerik (canlılar sistemi – 74 soru, yeryüzü ve uzay – 49 soru, fiziksel sistemler – 61 soru) (OECD, 2016a).

Özetle bu çalışmanın amacı, PISA 2015 fen sorularının Türkiye'de, cevap formatı, yeterlik boyutu, bilgi boyutu, bilişsel gereklilik ve içerik açılarından, farklı sosyoekonomik statüdeki öğrenciler arasında yanlı biçimde işleyip işlemediğini araştırmaktır.

Yöntem

PISA 2015'te, her ülkede, ilkinde okulların ve sonrasında öğrencilerin seçildiği iki basamaklı örnekleme yöntemi kullanılmıştır. Türkiye'de bu araştırmaya 187 okuldan on beş yaşında 5895 öğrenci katılmıştır. PISA tarafından örneklemden çıkarılan öğrecilerin (36 kişi) ve PA'nın yürütülemeyeceği 18 öğrencinin verisi bu çalışma özelinde kullanılmamış, sonuç olarak bu çalışmada 5841 öğrencinin PISA 2015 fen sorularına vermiş olduğu yanıtlar incelenmiştir. Bu araştırmada, tüm araştırmacılara açık, uluslararası veri tabanında yer alan veriler kullanıldığından etik kurul kararı gerektirmemektedir.

PISA 2015 fen test formları 12 kümeye ayrılmış toplam 184 bilgisayar temelli test sorusunu içermektedir. Her öğrenciye bu kümelerden ikisi, çok basamaklı döndürülmüş test deseni modeline göre belirlenerek uygulanmıştır (OECD, 2017). Öğrencilerin cevapladığı formlarda 25 ila 35 fen sorusu bulunmaktadır. Araştırmada uygulanan 184 fen sorusu, PISA'nın fen okuyazarlığı değerlendirme çerçevesine göre geliştirilmiştir (OECD, 2016a).

PISA'da öğrencilerin sosyoekonomik statüsü ekonomik, sosyal ve kültürel statü indeksi (ESCS) adı verilen bir indeksle hesaplanır. Bu çalışmada Profil Analizi için gerekli gruplar öğrencilerin ESCS puanlarına göre oluşturulmuştur. ESCS puanlarına göre Türkiye'de alt çeyreklikteki öğrenciler alt SES grubu olarak tanımlanmıştır. Benzer şekilde, üst çeyreklikte bulunan öğrenci grubu ise üst SES grubu olarak belirlenmiştir. Geri kalan öğrencilerse orta SES grubu olarak alınmıştır. Alt, üst ve orta gruplardaki öğrenci sayıları sırasıyla 1492, 1411 ve 2938'dir.

PA, bir testte belirlenen bir grup soru üzerinden, testi alan öğrencilerin gözlenen ve beklenen performanslarını karşılaştırmaya dayalı bir tekniktir (Verhelst, 2012). Beklenen puanlar bireylerin toplam test puanlarına ve testteki madde parametrelerine bağlı olarak hesaplanmaktadır. Madde parametreleri Rasch Model veya 2 parametreli model gibi Madde Tepki Kuramı'na (MTK) dayalı bir ölçme modeliyle kestirilmiş olmalıdır (van der Linden ve Hambleton, 1997).

Profil Analizde temel istatistik, bireylerin testin belirli bir özelliğe göre gruplanmış sorularındaki gözlenen ve beklenen performansları arasındaki farktır. Bu fark sapma profili (deviation profile) olarak adlandırılır. Pozitif sapma değeri öğrencinin, söz konusu sorulardaki performansının beklenenden iyi olduğunu gösterir. Benzer şekilde negatif bir değer, performansın beklenenden düşük olduğuna işaret eder. Bu yöntem, Danimarkalı tanınmış psikometrist Georg Rasch (1960) tarafından ispatlanan, toplam test puanı bilinen bir öğrencinin beklenen performansının sadece madde parametrelerine bağlı olduğu olgusuna dayanır. Dolayısıyla kayda değer büyüklükte bir sapma profili, belirli bir ortak özelliğe göre oluşturulan madde grubunda öğrencilerin performansını etkileyen ancak ölçme modelinin hesaba katmadığı etmenler olduğuna işaret eder. Bu çalışmadaki profil analizi Norman Verhelst (2012) tarafından geliştirilen ProfileG yazılımı ile yürütülmüştür. Yazılım, sapma değerlerinin istatiksel anlamlılığını ki-kare temelli bir algoritmayla kontrol etmektedir. Bu çalışmada profil analizi, aynı zamanda profil analizi sürecini de gösteren temel basamaklar İngilizce tam netinde örneğiyle birlikte ayrıntılı olarak açıklanmıştır.

PA sonuçları, ortak bir özelliğe sahip bireyler (örneğin, düşük sosyoekonomik statüde olanlar) için bir araya getirilerek grup düzeyinde sonuçlar elde edilebilir. Bu çalışmada PA sonuçları alt, orta ve üst SES gruplarında birleştirilmiştir. Bu amaçla, söz konusu üç öğrenci grubunda, 12 madde kategorisinin her biri için, PA sonuçlarına göre pozitif sapma profilindeki öğrenci sayısı belirlenmiştir. Buna ek olarak, yine üç öğrenci grubunda ölçme modelinin izin verdiği pozitif sapma profiline sahip öğrenci sayısı (bu sayı aşağıdaki analizlerde beklenen adıyla verilmiştir) belirlenmiştir. Profil analizinde bir öğrenci incelenen madde grubunda pozitif sapma profiline sahip değilse, testin geriye kalan maddelerinden oluşan tümleyen madde grubunda mutlaka pozitif sapma profiline sahip olmaktadır. Bu sebeple analizlerde her madde kategorisi tümleyeniyle birlikte ele alınmıştır (örneğin, açık uçlu sorular – açık uçlu olmayan sorular).

İlgili grupta ve ilgili madde kategorisinde beklenen ve gözlenen birey sayıları arasındaki fark değeri yüzde olarak da hesaplanabilir. Bu değere, söz konusu madde grubundaki *aşma yüzdesi (AY)* denir. AY değerinin istatiksel anlamlığı beklenen ve gözlenen birey sayısı arasındaki farka göre ki-kare testiyle belirlenebilir. Bu çalışmada, üç SES grubunda ve 12 madde kategorisinde yürütülen Profil Analiz sonuçları AY değerlerine göre incelenmiştir.

Bulgular

PISA 2015 fen sorularının alt kategorilerinde, öğrencilerin performanslarının sosyoekonomik statüleriyle ilişkili olduğunu göstermektedir. Bu kategoriler arasında, en yüksek AY değerleri cevap formatında görülmüştür ve bu değerler istatistiksel olarak da anlamlıdır. Ancak, istatistiksel anlamlılık madalyonun sadece bir yüzüdür (Yıldırım ve Yıldırım, 2011). Bu çalışmanın kurgusunun bir gerekliliği olarak üç SES grubu oluşturmak, her gruptaki öğrenci sayısını azalttığı için istatistiksel testlerin gücünü de azaltmıştır. Bununla beraber, PISA'da tamamlanmamış test deseni kullanıldığından bir öğrenci mevcut olan test sorularının sadece küçük bir bölümünü (yaklaşık %15) cevaplamaktadır (OECD, 2017). Dolayısıyla, test maddelerini alt gruplara ayırmayı gerektiren profil analizi gibi yöntemlerde, gücün kaybedilmesinden dolayı istatiksel olarak anlamlılığa ulaşmak kayda değer etki büyüklüklerine rağmen görece olarak zordur.

Sonuç olarak, bulguları etki büyüklüklerine göre yorumlamak, bu çalışmada kullanılan PA benzeri analizlerde daha uygundur. AY değerlerinin mutlak değerlerine göre etki büyüklükleri olarak görülebilir. Yukarıda belirlenen istatistiksel güç kaybından ötürü gözden kaçabilecek anlamlı bulguları belirlemek için kesme puanları belirlenmiştir.

% 1.20'den büyük olan değerler ve -% 1.50'den küçük olan değeler bu bağlamda değerlendirilmiştir.

PA'daki en anlamlı bulgunun cevap formatı kategorisiyle ilgili olduğunu göstermektedir. Öğrencilerin test performansları ile ilişkili değişkenleri saptayan araştırmalarda madde cevap biçiminin bu değişkenler arasında yer aldığı görülmektedir (örneğin; Liou ve Bulut, 2017; Mingo, Chang ve Williams, 2018). Cevap formatı bağlamında AY değerleri alt SES grubundaki öğrencilerin karmaşık çoktan seçmeli sorularda beklenenden daha iyi performans gösterdiklerini belirtmektedir (Alt SES, karmaşık çoktan seçmeli, AY = % 3.95). Diğer taraftan, alt SES grubundaki öğrenciler cevabı yazılan sorularda beklenenden düşük bir performans göstermişlerdir (Alt SES, cevabı yazılan, AY = - % 2.61). Bu durumla ilgili ilk akla gelen, cevabın yazılmasını gerektiren sorularda öğrencilerin kendilerini ifade etmek durumunda kaldıkları ve alt SES grubundaki öğrencilerin bu açıdan zorlanmalarından dolayı fen performanslarının yanlı olarak düştüğüdür. Bununla birlikte, PA'nın nedensellik öneren bir teknik olmadığı da göz ardı edilmemelidir.

Diğer taraftan, üst SES'te bulunan öğrencilerin çoktan seçmeli sorulardaki bu görece düşük performansları (Üst SES, Çoktan Seçmeli, AY = - % 2.33), üst SES grubundaki öğrencilerin alt ve orta SES de bulunan öğrencilerden geride olduğu anlamında gelmemektedir. Pozitif AY değerlerine karşılık gelen performanslar ilgili grubun ilgili kategorideki güçlü yönleri, negatif AY değerlerine karşılık gelen performanslar ise ilgili grubun ilgili kategorideki zayıf yönleri şeklinde yorumlanmalıdır.

Tartışma, Sonuç ve Öneriler

Üst SES grubundaki öğrencilerin görece olarak çoktan seçmeli sorulardaki zayıflığı ve bu SES grubundaki öğrencilerin karmaşık çoktan seçmeli sorularda beklenenden daha iyi performans göstermeleri şöyle açıklanabilir: Türkiye'de, tüm ulusal merkezi sınavlar (LGS, TEOG, Üniversite giriş sınavları) çoktan seçmeli sorulardan oluşmaktadır. Dolayısıyla, Türkiye'deki öğrencilerin bu tipteki sorulara alışkın olduğu düşünüldüğünde bu bulgu şaşırtıcı görünebilir. Ancak üst SES grubunda bulunan öğrencilerin, özel üniversitelere girmek veya yurtdışında eğitim alabilmek gibi daha geniş olanakları vardır. Bu nedenle, bu öğrencilerin çoktan seçmeli soruların yanıtlarını doğru seçmeye katkı sağlayan bazı test teknikleri eksik kalmış olabilir. Bu yorum içerik bilgisi gerektiren ve merkezi ulusal sınavlarda sıklıkla kullanılan sorular için de uygun olabilir. Sonuçlar, üst SES grubundaki öğrencilerin fen performanslarının içerik bilgisi gerektiren bu tür sorularda da beklenenin altında kaldığını göstermektedir (Üst SES, içerik bilgisi, AY = -% 1.49).

Bununla beraber, dikkat çeken bir diğer zayıf yön ise alt SES gruptaki öğrencilerin prosedürel bilgi sorularındaki performanslarıdır. Prosedürel bilgi soruları bilimin doğası (NOS) kavramı ile oldukça ilgilidir. Türkiye'de altıncı ve sekizinci sınıf düzeyindeki öğrencilerle yürütülen bir çalışmada Hacıeminoğlu, Ertepınar, Yılmaz-Tüzün ve Çakır (2015), öğrencilerin başarılarıyla bilimin doğası hakkında görüşleri arasındaki ilişkinin (özellikle bilimin doğasının deneysel boyutu) alt SES'te bulunan okullarda üst SES'te bulunan okullara göre daha zayıf olduğunu ortaya koymuşlardır. Diğer taraftan alt SES okullarındaki öğrenciler, bilimin doğası boyutlarından bilimin deneysel temelli olması görüşünde düşük düzeydedir. Dolayısıyla, PISA 2015'te prosedürle ilgili soruların öğrencilerin bilimin doğası görüşleriyle ilişkili olduğu düşünüldüğünde, alt SES grubundaki öğrencilerin bu sorulardaki zayıflıklarının bilimin doğası anlayışlarındaki eksiklerle ilişkili olduğu söylenebilir. Bununla beraber, sosyoekonomik statüsü düşük okullarda öğretim programı dışı etkinlikler, bilim kulübü etkinlikleri, donanımlı laboratuvarlar gibi öğrencilerin bilimin doğası anlayışlarını ve dolayısı ile prosedürsel bilgilerini besleyen fırsatların ve olanakların eksikliği de bu çalışmadaki bulguyla ilişkili olabilir.

Bulgular tüm güçlü ve zayıf yönlerin madde formatı veya bilgi boyutlarıyla ilgili olduğunu göstermektedir. Tek istisna, üst SES grubunun güçlü yönlerinden biri olarak belirlenen içerik kategorisindeki yeryüzü ve uzay konusudur (Üst SES, Yeryüzü ve Uzay, AY = % 1.75). Son teknolojik gelişmeler, profesyonel teleskoplar, görüntü araçları ve bilim ve teknoloji merkezlerindeki modeller, dünya ve uzayın uydu görüntülerine ulaşan web tabanlı programlar, uzay mekiği kameraları öğrencilerin dünya ve uzay hakkındaki öğrenmelerini zenginleştirmektedir. Ancak bu tür maliyeti yüksek desteklere alt SES'teki öğrencilerin ulaşabilme olasılıkları düşüktür. Örneğin, PISA 2015'te, OECD ülkelerinde, bilim kulüplerine veya yarışmalara katılmak gibi müfredat dışı bazı etkinliklerin avantajlı olan okullarda önerilme oranının dezavantajlı olan okullardakine oranla fazla olduğu görülmektedir (OECD, 2015a). Dolayısıyla, üst SES grubundaki öğrencilerin bu olanaklara ulaşabilmeleri yeryüzü ve uzay sorularında avantajlı performans göstermelerinin bir nedeni olabilir.

Özetle, bulgular biri hariç tüm AY değerlerinin alt veya üst SES gruplarında görüldüğüne işaret etmektedir. Buradan hareketle, öğrenciler arasında sosyoekonomik farklık ne kadar çoksa, belirli madde kategorilerinde yanlılığın olma ihtimalinin de o kadar fazla olduğu söylenebilir. Bununla beraber, yanlılıkların baskın olarak madde formatı ve bilgi boyutu kategorilerinde olması, Türkiye'de özellikle bu kategorilerin öğrencilerin sosyoekonomik statüleriyle ilişkili yanlılıklar üretebileceğini gösteren ampirik bir kanıt olarak değerlendirilebilir. Bu çalışma PISA'ya katılan diğer ülkeler için de tekrarlanarak, ülkeler için bütünleştirilmiş bulgular sağlayacak şekilde genişletilebilir. Bu genişletme, SES'in doğasını anlamaya da katkı sağlayabilir. Diğer bir genişletme biçimi ise çalışmayı matematik gibi başka konu alanlarında yapmak olabilir.

Etik Kurul Kararı

Bu araştırmada, tüm araştırmacılara açık, uluslararası veri tabanında yer alan veriler kullanıldığından etik kurul kararı gerektirmemektedir.