

Views of Teachers on the Potential Negative Effects of High Stake Tests

Mustafa Ilhan ¹, Nese Guler ², Gulsen Tasdelen Teker ^{3,*}

¹Dicle University, Ziya Gökalp Education Faculty, Department of Mathematics and Science Education, Diyarbakır, Turkey

²İzmir Demokrasi University, Faculty of Education, Department of Measurement and Evaluation, İzmir, Turkey

³Hacettepe University, Faculty of Medicine, Department of Medical Education and Informatics, Ankara, Turkey

ARTICLE HISTORY

Received: May 10, 2020

Revised: Jan. 27, 2021

Accepted: Apr. 05, 2021

Keywords:

High stake exams,
Pairwise comparisons,
Many-facet Rasch model,
Scaling method

Abstract: This study analyses teachers' viewpoints on the potential undesirable influences of high stake exams. Seven stimulants related to undesirable influences of high stake exams on education were given to 191 teachers in a pairwise comparisons format. The participating teachers in this study were asked to choose one undesirable influence by comparing the stimulants given to them in pairwise and to determine the more prominent problem stemming from high stake exams. Data were analyzed via many-facet Rasch model. As a result, it was found that teachers considered the stimulant "school assessments turn into secondary importance in the eyes of students and parents" as the foremost problem stemming from high stake exams. On the other hand, the stimulant "administrators focus on policies for increasing exam scores instead of policies for improving the learning-teaching process" ranked the last in the undesirable influences of high stake exams.

1. INTRODUCTION

Examinations are an integral part of educational processes. Individuals' proficiency levels in various fields are determined via examinations and in accordance with the results obtained and decisions are made about the success of students, the functioning of curricula, or the quality of teaching. Yet, the importance of decisions made based on exam results is not always the same for students, parents, teachers, schools, or educational policy-makers. While some of the exam results form the basis for at least one key stakeholder of education to make extremely important decisions, some of the results are used to make decisions with relatively more restricted effects on individuals. Considering the impact of the decisions it creates on students, parents, education administrators, or policy makers and the traces they leave on individuals' lives, exams can be examined under two headings; namely, low stake exams and high stake exams.

Exams whose results are not used to make important decisions for students or educators are called low stake exams. Subject screening tests and quizzes given in order to determine the deficiencies in learning and to plan the improvements to be made in teaching are the most typical examples for low stake exams. Low stake exams have such an important advantage as

*CONTACT: Gülşen Taşdelen Teker ✉ gulsentasdelen@gmail.com 📍 Hacettepe University, Faculty of Medicine, Department of Medical Education and Informatics, Ankara, Turkey

causing no stress to individuals (Simpson, 2016). On the other hand, when individuals know that the test they take will not have important results for them, they can lack motivation (Finn, 2015; Kornhauser et al., 2014), which would result in doubtful approaches towards the validity of information that low stake exams present in relation to examinees' performance (Wise & Demars, 2005).

Tests are defined as high stake exam if individuals gain or lose a lot according to the results obtained from them (Coniam & Falvey, 2007). What makes an exam at high stake is often the impact it may have on the educational life and career opportunities of the candidates (Moses & Nanna, 2007). However, sometimes their impact on teachers, educational administrators, or educational policy makers can make an exam a high stake one. For instance, even though such international exams as the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) are low stake ones for students taking the tests, they can be high stake for governments, educational policy makers, and schools (Stobart & Eggen, 2012). In a similar vein, if the scores students receive from the exams they take are used for evaluating teachers' performance and making decisions about their appointment and promotion, they have high risks for teachers but not for students (Dawson, 2012). However, this study focuses on high stake exams that affect primarily students and therefore their parents.

While high stake exams have several advantages that are referred to as justification for large areas of use, on the one hand, they also have a number of disadvantages causing them to be the focus of complaints. A number of major advantages that high stake exams offer are listed as follows: It is thought that the use of the in-class assessment results in order to make decisions about individuals causes justice problems since lecture notes can differ from teacher to teacher and from school to school (Holland, 2001; cited in Brockmeier et al., 2014). However, high stake exams are believed to make relatively fairer measurements (Phelps, 2003). Therefore, making important decisions about individuals according to the results of high stake exams is perceived by society as a more reliable (Çetin & Ünsal, 2018) and conscientious alternative (Baykal, 2014). Another advantage of high stake exams is that they can provide national and even international data on the success of students, schools, and educational systems (Acar Güvendir, 2014; Baines & Stanley, 2004). Other positive qualities of high stake exams include raising teachers' sense of responsibility and creating the need for updating themselves (Çetin & Ünsal, 2018), motivating students to study harder on the one hand, and giving feedback to students about their strengths and weaknesses on the other hand (Stecher, 2002).

What turns high stake exams into a controversial subject despite the above-mentioned advantages they offer is that the disadvantages of them are enough to overshadow their advantages. The disadvantages of high stake exams can be summarized as follows: Teaching applications (such as lab work, educational trips, etc.) and lessons not included by the exams are handled superficially (Taylor et al., 2003) and teaching process is regulated in accordance with exam content (Yeh, 2005). These exams put pressure on mainly students, parents, educators, and administrators (Kruger et al., 2007), ensuring that increase in students' exam scores becomes the primary purpose of education (Pbrreault, 2000) and such practice moves education away from collaborative mentality and turns it into a rivalry-oriented system (Polesel et al., 2012)

High stake exams are widely applied in many countries in the world and Turkey is not exceptional in using such exams. In the Turkish educational system, exams administered at a national scale are taken into consideration in making such important decisions as to transition into secondary and higher education, the selection of personnel for employment in the public sector, and determining the individuals who will receive specialisation training in medicine and dentistry. As a result of its widespread use in the educational process, research related to high

stake exams takes a large place in the related literature. When the studies in the literature on the subject are looked into, it is figured out that the effects of high stake exams on students (Amrein & Berliner, 2003; Banks & Smyth, 2015; Segool et al., 2013; Simpson, 2016), teachers (Abrams, 2004; Assaf; 2008; Brady, 2008; Christian, 2010; Dawson, 2012), parents (Polesel et al., 2012; Saito, 2006; Westfall, 2010), teaching-learning process, and curriculum implementation (Amoako, 2019; Finkeldei, 2016; Davis, 2011; Johnson, 2007; Marchant, 2004; Ritt, 2006; Shepard & Dougherty, 1991; Taylor et al., 2002; Togut, 2004; Vogler & Virtue, 2007; Wright, 2002) are generally focused. Current research in the literature provides important information about the undesirable influences of high stake exams in the educational process but does not provide any information on which of the problems stemming from high stake exams is prioritized or which is more in the background. However, in the literature, it is pointed out that the first thing to do in order to overcome the problems caused by an application is to prioritize the problems while the first 20% of the problems are expressed as the causes of the remaining 80%. According to the 80/20 rule (Knapp, 2010), the effectiveness of the steps taken to solve a problem depends on the fact that these steps are related to the first 20% of the problems (Kane, 2014). Therefore, in order to produce the right solutions for the undesirable influences of high stake exams, it is thought that the problems arising from these exams should be sorted out.

1.1. The Purpose and Significance of the Study

The aim of the current study is to analyse teachers' viewpoints on the potential undesirable influences of high stake exams which interest almost all individuals in the society directly or indirectly in terms of their results through pairwise comparisons based on many-facet Rasch model (MFRM). The research is thought to make two important contributions to the literature: Firstly, this study differs from other research in the literature in that it intends to reveal the problems that need to be addressed primarily, beyond identifying the problems caused by high stake exams; the second feature of this research making it important for the literature is that it has a methodological difference. When the studies utilizing the scaling method through pairwise comparisons in the literature are examined (Nartgün, 2006; Anıl & Güler, 2006; Bülbül & Acar, 2012; Ekinçi et al., 2012; Güler & Anıl, 2009; Nalbantoğlu Yılmaz, 2017; İlhan, 2016; Yaşar, 2018), it can be ascertained that the analyses have been performed usually by means of Microsoft Excel and also they have been done on the basis of traditional psychometric approach. Even though the study performed by Güzeller, Eser and Aksu (2016) differs from other studies available in the literature in that it analyses the pairwise data by using R software, it is similar to other pairwise comparison studies in that traditional psychometric approach is dominant in the analysis process. In the present study, however, the collected data through pairwise comparisons are analyzed on the basis of MFRM. When the pairwise comparison data are analyzed within the framework of traditional psychometric approach, statements of indiscrimination are not permitted and the participants are always asked to make a choice between two stimulants (Turgut & Baykul, 1992). In such a case, some of the pairwise comparisons can be left unanswered and thus it becomes difficult to collect data, which may result in a probable loss of data. The first advantage in analyzing the data of pairwise comparisons emerges at this point. When the stimulants compared in pairs are analyzed using the MFRM, participants are not always expected to make a choice. In fact, they are allowed to think equivalently about the two stimulants (Linacre, 2014). Another advantage offered by analyzing the data of pairwise comparisons in MFRM is that statistics which provide evidence for psychometric properties of measurements are reported synchronically with the stimulants' scale values. The fit statistics calculated for the stimulant facet, reliability coefficient, and separation ratio along with the scale values for the stimulants are also available in the many-facet Rasch analysis outputs. There is evidence for both validity and reliability: Fit statistics

provide evidence for the former and reliability coefficient and separation ratio provide evidence for the latter.

In addition, individuals inform us about whether model-data fit exists and whether there are interactions between facets even though they are not used for measurements in pairwise comparisons based on Rasch analysis (Linacre, 2014). Considering all these advantages, it is believed that providing a sample study analyzing pairwise comparisons data according to MFRM for the literature would be important. In this context, the study is also thought to have potential to contribute to the literature in terms of methodology.

2. METHOD

2.1. Research Model

This study is a descriptive survey research. Descriptive research is based on the principle of revealing the present situation without any intervention and is mostly considered as a survey model (Erkuş, 2011). Basic characteristics of survey model include gathering information from individuals to define certain characteristics (attitude, belief, opinion, ability, etc.) of the universe to which they belong (Fraenkel et al., 2012), requiring a large sample selection to represent the universe, presenting standard information obtained by applying the same measurement tool to all individuals in the sample, and collecting quantitative data on which statistical procedures can be performed (Cohen et al., 2007).

2.2. The Study Group

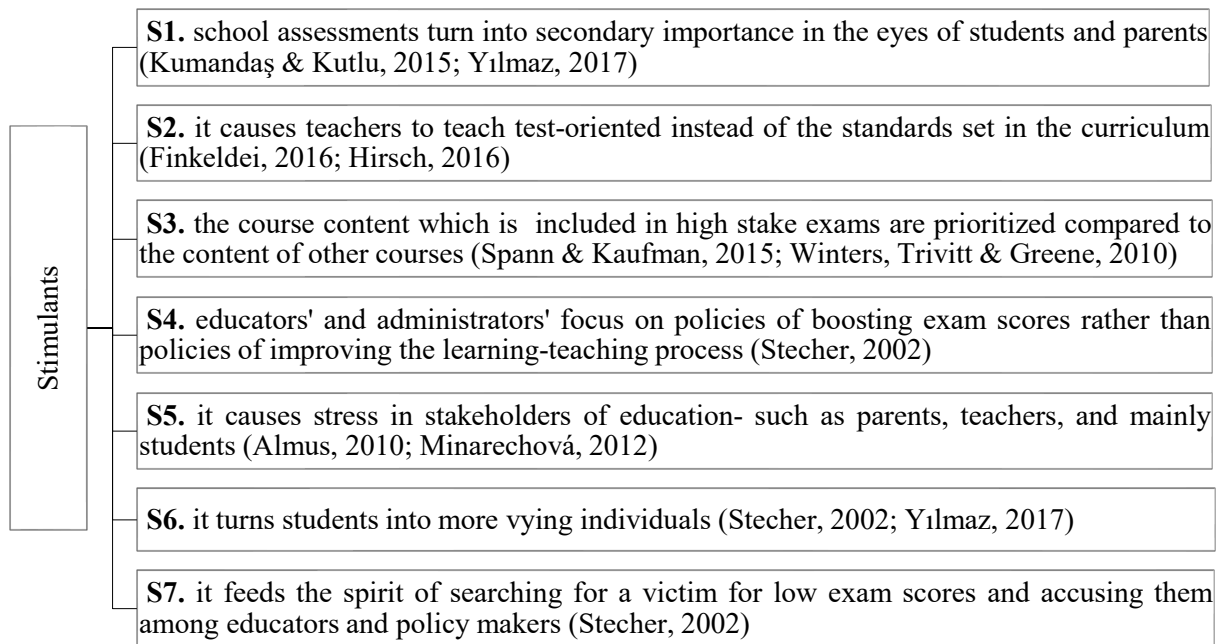
The study was conducted on 191 teachers working in Turkey. Of all the participating teachers 88 (46.07%) were female, 84 (43.49%) were male, and 19 (9.95%) did not mention their gender. The distribution of the teachers according to the stage of education they taught was as follows: 41 (21.47%) primary school teachers, 74 (38.74%) secondary school teachers, and 76 (39.79%) high school teachers. The teachers included in the study group ranged between 22 and 61 years old (\bar{X} =34.33) and they had been teaching for 1–33 years (\bar{X} =10.17).

2.3. Data Collection Tool

The relevant literature about the potential undesirable results that high stake exams can yield was reviewed prior to forming the data collection tool. It was seen in the literature review that there were many instructional and affective undesired influences associated with high stake exams. While deciding on the stimulants to be included in the data collection tool, the negative effects mentioned in almost all of the studies were examined. Consequently, the seven stimulants given in [Figure 1](#) coded as and ranked between S1 and S7 and mentioned frequently in most of the examined studies were determined as the major undesirable influences caused by high stake exams in the education process.

The above mentioned seven stimulants were arranged in a way that the teachers involved in the research could make pairwise comparisons and the data collection tool was formed. The data collection tool included 21 comparisons containing the pairwise combinations of the seven stimulants as well as demographic variables of gender, age, branch, and duration of service in the teaching profession. An instruction about the purpose of the study and how to answer the measurement tool was added to the beginning of the instrument. In the directive, the phrase central examination was adopted instead of a high stake exam. Because it was thought that the term of central exam would be more understandable for teachers compared to the concept of high stake exam. In addition, in order to make the statement of the central exam clearer, which exams are included in this scope were exemplified in parentheses; namely, after the central exam phrase a parenthesis was opened and examples of central examination implemented in Turkey was listed.

Figure 1. *The Potential Undesirable Influences of High Stake Exams in Education.*



Prior to its use, the data collection tool was presented to the opinion of a total of five experts, each with at least a PhD degree: two measurement and evaluation experts, one curriculum and instruction expert, one training management, one supervision and planning expert, and one psychological counselling and guidance expert. Experts reported that the directive and statements in the data collection tool were understandable and contained the main undesirable influences that could be associated with high stake exams. Then, two experts in the field of language education, one of whom attended doctorate education and the other had the title of associate professor, were interviewed. Data collection phase was started when experts reported that there were no spelling rules, punctuation marks, or problems related to narration.

2.4. Data Collection Procedure

The data were collected during the 2017–2018 and 2019–2020 academic years. When the participating teachers were asked to compare the stimulants pairs given to them in the instrument, they stated the more important problem arising from the high-stakes exams. They were also asked to indicate the comparisons they could not make. The data collection process was completed in this way. No missing or incorrectly filled measurement tool was found in the data set. On examining the data collected, it was detected that the teachers had found it difficult to make a choice between the two stimulants in 256 out of 4011 comparisons [(191 teachers) x (21 pairs of stimulants)]. It was also found that they did not have difficulty in the remaining 3755 comparisons.

2.5. Data Analysis

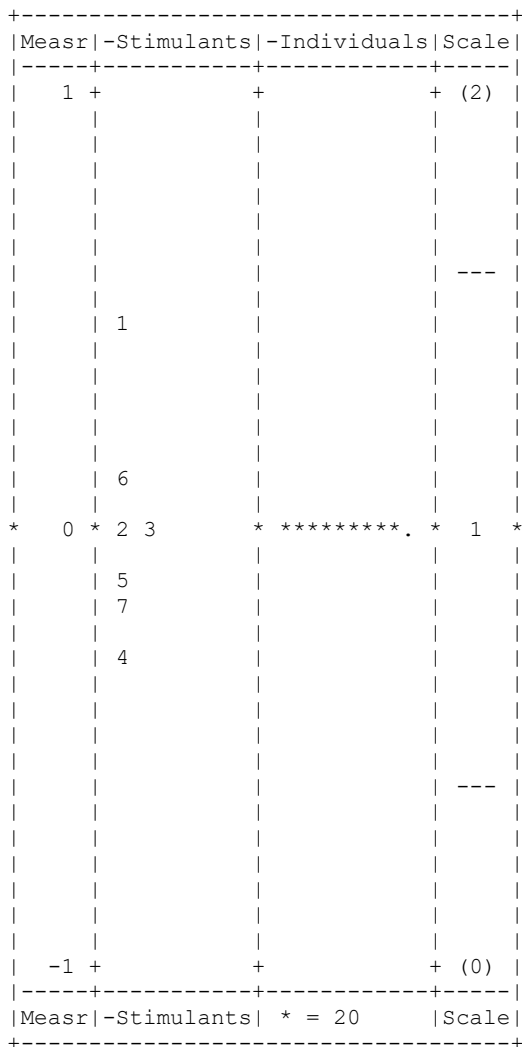
The MFRM is an extension of the basic Rasch model and is a highly functional analysis for situations where there are different sources of variability that can affect the measurement results other than items and individuals. The MFRM was initially conceptualized as consisting of item, individual, and rater facets. However, later on, the model was expanded to increase the number of facets and the model was started to be used in different problem situations and data sets other than the rater mediated assessments. For example, in order to determine whether the scores of the dependent variable differ in terms of a categorical independent variable, studies (Behizadeha & Engelhard, 2014; Güler & İlhan, 2017; İlhan & Güler, 2017; Ricketts, Engelhard ve Chang, 2015) have been conducted using the MFRM.

When the user manual of FACETS, which is a package program where MFRM is performed, is examined, it is understood that many-facet Rasch analysis can be used in pairwise comparisons of objects by individuals. Three components are available for each observation in data sets obtained in pairwise comparisons. While one of these components is the individuals making the comparison, the remaining two components are the two stimulants compared (Linacre, 2014). Thus, MFRM emerges as the suitable alternative usable in analysing the data of pairwise comparisons. Accordingly, the data were analysed by using the MFRM in a design of three facets: stimulant 1, stimulant 2, and individuals who made the comparisons. In the data set of the pairwise comparisons based on the MFRM, a three-category scale is available in which 2 is used for choosing stimulant one, 1 for failure to make a choice between the two stimuli, and 0 for choosing stimulant two. Since the analysis outputs reported in the findings section provide evidence for the validity and reliability of the measures, the psychometric properties of the data collection tool are not included under this heading.

3. RESULTS / FINDINGS

The logit map obtained by analysing the pairwise comparison data according to MFRM is shown in [Figure 2](#). As is clear from [Figure 2](#), there are four columns in the logit map.

Figure 2. *Logit map.*



The column to the far left of the logit map contains scale levels. In what range the scale values in this column are reported is dependent on the measures of the components in the facets

included in the study. An examination of [Figure 2](#) shows that the scale values for this study are reported to be in ± 1 range. The second column of the logit map corresponds to the stimulants. There is a distribution in this column from the top to the bottom extending between the most significant undesirable influence of high stake exams and relatively less significant influences of those exams. Accordingly, S1 was considered as the most significant undesirable influence of high stake exams by teachers, while S4 was considered as a problem less severe than the other six undesirable influences. Column three in the logit map shows the measurements for the facet of individuals. The ranking in terms of examinees' ability levels can be seen by looking at this column in studies aiming to measure individuals' levels of ability. Yet, when pairwise comparison data are analysed in MFRM, scores for abilities cannot be found for individuals as in analyses in Excel. For this reason, all the teachers in the column of individuals are at the same point in this logit map. Column four in the logit map shows information on the scale categories used in this study. After the logit map, the measurement reports for the stimulant facet were analysed and the findings obtained are shown in [Table 1](#).

Table 1. *Measurement Reports for the Facet of Stimulant.*

Stimulant	Measure	Model S.H	Infit MnSq	Outfit MnSq
S1	.46	.03	1.00	1.00
S6	.11	.03	1.02	1.02
S2	.02	.03	1.00	1.00
S3	-.01	.03	.98	.98
S5	-.14	.03	.97	.96
S7	-.20	.03	1.04	1.05
S4	-.31	.03	.99	.98
Mean	-.01	.03	1.00	1.00
Standard deviation	.25	.00	.02	.03
Chi square = 335.10	$df = 6$	$p = .00$	Separation Index = 7.75	Reliability = .98

As can be seen from [Table 1](#), the potential undesirable influences of high stake exams were differentiated significantly by the teachers [$\chi^2_6 = 335.10, p < .001$]. Having separation ratio above 2 and reliability coefficient above .80 (Linacre, 2012) indicated that the measures made in the study were reliable. The infit and outfit statistics in [Table 1](#) were found to take on values between .96 and 1.05. The acceptable range for infit and outfit statistics is .5 to 1.5 (Wright & Linacre, 1994). When the number of stimulants seen by the participants as equal is too high, the fit statistics fall below the acceptable range, and in this case, the assumptions about the MFRM are not met and validity issues arise. The fit statistics in the [Table 1](#) are within the recommended range and make a sign that the stimulants ratio seen as equal is not at a size that will negatively affect model-data fit in unfavourable ways. This finding regarding the fit indices provides evidence for the validity of the measurements.

The results of the stimulant presented visually in the logit map are shown numerically in [Table 1](#). Apparently, S1 is the stimulant having the highest scale value (.46) with a considerable difference. Accordingly, the teachers considered the stimulant “school assessments turn into secondary importance in the eyes of students and parents” as the foremost undesirable influence of high stake exams in the process of education, which was followed by the stimulant “it turns students into more vying individuals”. “It causes teachers to teach test-oriented instead of the standards set in the curriculum” ranked the third as an undesirable influence. The stimulant “the course content which is included in high stake exams are prioritized compared to the content of other courses” ranked the fourth and the stimulant “it causes stress in students, teachers and parents” ranks the fifth. The stimulant “it feeds the spirit of searching for a victim for low exam scores and accusing them among educators and policy makers” and the stimulant “educators’

and administrators' focus on policies of boosting exam scores rather than policies of improving the learning-teaching process" ranked the sixth and the seventh, respectively.

4. DISCUSSION and CONCLUSION

According to the research results, the teachers reported that "school assessments turn into secondary importance in the eyes of students and parents" was the foremost undesirable influence stemming from high stake exams by obvious difference. The clear difference detected means that the primary problem to be resolved regarding the negative effects of high stake exams is that these tests push school assessments to the second plan. Additionally, the fact that this stimulant was placed the first by far can be interpreted as a remarkable consensus among the teachers on the negative effects of these exams. Findings regarding the first stimulant's scale values being significantly different compared to the scale values of other stimulants are in parallel to the ones reported in the literature. Atılgan (2018), in a study on high stake exams administered in transition into the next stages in Turkey historically, points out that schools and curricula have become dysfunctional due to those exams. According to Atılgan (2018), high stake exams have become the goal and schools have become the instrument giving diplomas to achieve the goal in the current system of education. In a similar vein, Can (2017) also states that the great majority of students stated that the primary goal for them was success in the high stake exams and they attended classes in their institutions just to get a diploma. Accordingly, the fact that school assessments are considered as the foremost problem stemming from high stake exams in the eyes of parents and students is a reflection that school assessments are perceived as the tasks which must be fulfilled for high stake exams.

The teachers in the study group put the fact that "it turns students into more vying individuals" in rank two as the undesirable influence of high stake exams. The fact that high stake exams have an almost vital impact on individuals' future lives makes it inevitable that high stake exams trigger rivalry among students. However, the second order among the potential undesirable influences of this rivalry caused by the high stake exams suggests that the psychological impact of these exams on students may be greater than expected. In fact, the vying environment created by high stake exams may make it difficult to transfer the fundamental values such as love for charity, sharing, solidarity, and cooperation (Board of Education and Training, 2017) that the Ministry of National Education of Turkey aims to bring to students.

According to the results obtained in this study, the stimulant "it causes teachers to teach test-oriented instead of the standards set in the curriculum" was ranked the third in the undesirable influences of high stake exams. It was followed by the stimulant "the course content which is included in high stake exams are prioritized compared to the content of other courses" with a scale value very close to the one ranking the third. These stimulants, which ranked the third and fourth, can be regarded as the results of "considering school assessment as of secondary importance" and of "the increase in rivalry between students". More clearly, the fact that the importance students and parents attach to school assessment is shadowed by high stake exams can lead teachers to shape teaching according to high stake exams. Additionally, teachers can choose to plan their teaching according to the test content instead of the curriculum to support their students in the high rivalry environment caused by high stake exams. Thus, which of the stimulant is perceived as the more primary problem stemming from high stake exams can be connected with the cause-effect relationships between the stimulants. This view is supported by the Pareto principle (Jenny, 2007), which argues that priority issues are the cause of other problems with lower priority.

The stimulant "stress caused by high stake exams in students, teachers, and parents" ranked the fifth in the undesirable influences of high stake exams. Accordingly, teachers consider the

affective influences of high stake exams on the stakeholders of education as a problem less important than the effects on learning-teaching process. Yet, it should not be forgotten that the situation might have stemmed from the fact that the study was conducted with the teachers, not with the students or parents. This is because the examinations administered in Turkey form the basis for decisions to be made about students but they are not used in decisions for teachers. Therefore, it is thought that if a study is conducted with students and parents or if such a study is performed in a country where high stake exams influence teachers' wages (Brooke, 2016) and their position (Nichols & Berliner, 2005), the stress caused by exams can be considered as a more important problem. It was found in relevant literature that the stress teachers have due to high stake exams can differ from country/state to country/state. Abrams (2014), for instance, compared the pressure teachers working in and outside Florida were exposed to and found that 80% of the teachers working in Florida felt the pressure caused by high stake exams but that 40% of the teachers working in the other states felt the pressure.

Searching for victims among educators, administrators, and policy makers for low exam scores and accusing them ranked the sixth as an undesirable influence of high stake exams. Administrators' focussing on policies for boosting exam scores instead of policies for improving the learning-teaching process ranked the last. The fact that teachers ranked these two stimulants at the bottom meant that they considered the grade level and school level effects of those exams as more important than the effects on educational policies. This finding was quite different from the one obtained in Adedoyin (2013) who analysed university students' viewpoints on high stake exams in Botswana educational system. It was found that the fact that those exams caused politicians to search for victims for low exam scores and that the exams offered misleading information causing politicians to make inadequate decisions about the process of education were the undesirable influences of high stake exams. On the other hand, it was also found that high stake exams did not have such effects as causing school assessment to lose its importance or causing the subjects included in test content to be prioritised. The great differences between the findings obtained in this study and those in Adedoyin (2013) can be considered as an indication that the effects of high stake exams on the education system vary from country to country.

A review of relevant literature shows that there are studies concluding that the results of high stake exams influence educational policies. For instance, Buyruk (2014) reports that students' achievement in high stake exams is associated with teachers' accomplishment and that the results of those exams are used like instruments informing us of school and teacher performance according to provinces, districts, and schools. Due to this, high stake exams can lead educators and administrators to policies for boosting exam scores to be in a better position in comparisons between schools. Therefore, it would be mistaken to see the results of the study as high stake as exams in Turkey do not have undesirable influences such as searching for victims for low exam scores or educators', administrators', and policy-makers' focusing on policies for boosting exam scores instead of policies for improving the learning-teaching process. The reality is that teachers do not consider these two undesirable influences as primary as the other stimulants.

4.1. Implications for Practice

This study concentrates on seven undesirable influences of high stake exams. However, the undesirable influences are not restricted to the ones considered in this study. There are several undesirable influences mentioned in the literature that affect all the stakeholders of education such as students, teachers, and administrators. In this context, the undesirable influences of high stake exams should be revised again in the light of scientific studies performed. The attention of stakeholders who can make regulations to reduce the undesirable influences on the educational system should be called to the problems that are considered more primary. Such

exams should no longer be the turning points for individuals so that the undesirable influences could be minimised. Schools should be varied and the differences between schools arising from physical conditions (the number of students per classroom and teacher and specially equipped learning environments such as laboratory, library, and gymnasium), artistic, sporting, cultural and social activities, social-economic conditions of the school district, and teacher qualifications should be reduced, and the meaning attached to exams should be minimized without getting away from the fact that a system without exams is impossible at the moment.

4.2. Future Directions

The first proposal that could be brought into the scope of the study is that the researchers would prefer to use MFRM instead of the traditional psychometric approach when scaling with pairwise comparisons. Since this study aimed to show that paired comparison data can be analyzed with MFRM, MFRM and traditional method comparison was not performed. Another research proposal that can be brought in this context is to test the agreement between the scale values obtained from the paired comparisons performed according to the MFRM and the traditional method. Finally, this study was conducted with a study group of 191 teachers that was not very large. Therefore, it may be suggested that a similar study be conducted with different samples in order to increase the generalizability of the findings obtained in the study.

Acknowledgments

The author(s) received no financial support for the research, authorship, and/or publication of this article. The first draft of the paper was presented at VIth International Congress on Measurement and Evaluation in Education and Psychology, Prizren, Kosova.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

Authorship Contribution Statement

Mustafa İlhan: Investigation, Resources, Methodology, Analysis, Writing the original draft.

Neşe Güler: Investigation, Resources, Supervision, Writing the original draft.

Gülşen Taşdelen Teker: Investigation, Resources, Methodology, Writing the original draft.

ORCID

Mustafa İLHAN  <https://orcid.org/0000-0003-1804-002X>

Neşe GÜLER  <https://orcid.org/0000-0002-2836-3132>

Gülşen TAŞDELEN TEKER  <https://orcid.org/0000-0003-3434-4373>

5. REFERENCES

- Abrams, L. M. (2004). *Teachers' views on high-stakes testing: Implications for the classroom*. Arizona: Policy Brief: Education Policy Studies Laboratory. Arizona State University College of Education. <http://epsl.asu.edu/epru/documents/EPSSL-0401-104-EPRU.pdf>
- Acar Güvendir, M. (2014). Student and school characteristics' relation to Turkish achievement in student achievement determination exam. *Education and Science*, 39(172), 163–180. <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/2839>
- Adedoyin, O. O (2013). Public examinations and its influence on the Botswana educational system: Views of undergraduate education students at the University of Botswana. *Asian Journal of Humanities and Social Sciences*, 1(2), 124–134. <https://ajhss.org/pdfs-1/Public%20Examinations%20and%20its.....pdf>

- Almus, K. (2010). *The beliefs of principals and assistant principals regarding high-stakes testing*. [Doctoral dissertation, University of Houston, Houston, United States]. <https://uh-ir.tdl.org/handle/10657/ETD-UH-2010-12-78>
- Amoako, I. (2019). What's at stake in high-stakes testing in Ghana: Implication for curriculum implementation in basic schools. *International Journal of Social Sciences & Educational Studies*, 5(3), 72–82. <https://doi.org/10.23918/ijsses.v5i3p72>
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stake testing on student motivation and learning. *Educational Leadership*, 60(5), 32-38. http://www.wou.edu/~girodm/611/testing_and_motivation.pdf
- Anıl, D., & Güler, N. (2006). An example of the scaling study by pair-wise comparison method. *Hacettepe University Journal of Education*, 30, 30-36. <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/722-published.pdf>
- Assaf, L. Z. (2008) Professional identity of a reading teacher: Responding to high-stakes testing pressures. *Teachers and Teaching: Theory and Practice*, 14(3), 239-252. <https://doi.org/10.1080/13540600802006137>
- Atılğan, H. (2018). Transition among education levels in Turkey: Past-present and a recommended model. *Ege Journal of Education*, 19(1), 1-18. <https://doi.org/10.12984/egeefd.363268>
- Banks, J., & Smyth, E. (2015). Your whole life depends on it': academic stress and high-stakes testing in Ireland. *Journal of Youth Studies*, 18(5), 598-616. <http://dx.doi.org/10.1080/13676261.2014.992317>
- Baykal, A. (2014). *Sınavlardan sınav beğen*. *Eğitim sisteminde kademeler arası geçiş ve sınavlar* [Like the exam from the exams. Transition between stages and exams in the education system]. Ege'den Eğitime Bakış Paneli, Ege Üniversitesi, İzmir. https://www.academia.edu/11610273/SINAVLARDAN_SINAV_BE%C4%9EEN
- Behizadeha, N., & Engelhard, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing*, 21, 18-36. <https://doi.org/10.1016/j.asw.2014.02.001>
- Board of Education and Training. (2017). *Our renewal and change studies in curriculum*. https://tkkb.meb.gov.tr/meb_iys_dosyalar/2017_07/18160003_basin_aciklamasi-program.pdf
- Brady, A. L. (2008). *Effects of standardized testing on teachers' emotions, pedagogy and professional interactions with others*. [Doctoral dissertation, Cleveland State University, Cleveland, United States]. <https://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1038&context=etdarchive>
- Brockmeier, L. L., Green, R. B., Pate, J. L., Tsemunhu, R. T., & Bochenko, M. J. (2014). Teachers' beliefs about the effects of high stakes testing. *Journal of Education and Human Development*, 3(4), 91-104. <http://dx.doi.org/10.15640/jehd.v3n4a9>
- Brooke, N. (2016). High-stakes accountability using teacher salary incentives in Brazil: An update. *Profesorado, Revista de Currículum y Formación del Profesorado*, 20(3), 207–250. <https://recyt.fecyt.es/index.php/profesorado/article/view/54598>
- Bülbül, T., & Acar, M (2012). A pair-wise scaling study on the missions of education supervisors in Turkey. *International Journal of Human Sciences*, 9(2), 623–640. <https://www.j-humansciences.com/ojs/index.php/IJHS/article/viewFile/2327/941>
- Buyruk, H. (2014). Standardized examinations as a teacher performance indicator and performance evaluation in education. *Trakya University Journal of Education*, 4(2), 28–42. <http://dergipark.gov.tr/trkefd/issue/21480/230201>
- Can, E. (2017). Determination of the effects of central exams according to the view of students. *The journal of Academic Social Science*, 5(58), 108-122. <http://dx.doi.org/10.16992/ASOS.12842>

- Çetin, A., & Ünsal, S. (2018). Social, psychological effects of central examinations on teachers and their reflections on teachers' curriculum implementations. *Hacettepe University Journal of Education*, 34(2), 304-323. <http://dx.doi.org/10.16986/HUJE.2018040672>
- Christian, S. C. (2010). *High-stakes testing and its relationship to stress levels of secondary teachers*. [Doctoral dissertation, The University of Southern Mississippi, Hattiesburg, United States]. <https://aquila.usm.edu/dissertations/932/>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education (6th ed.)*. Routledge Falmer.
- Coniam D., & Falvey P. (2007) High-stakes testing and assessment. In J. Cummins J., & C. Davison (Eds.), *International handbook of English language teaching* (pp. 456-471). Springer.
- Davis, M. F. (2011). *The influence of high-stakes testing on science teacher perceptions and practices*. [Doctoral dissertation, Walden University, Minneapolis, United States]. <https://eric.ed.gov/?id=ED528115>
- Dawson, H. S. (2012). *Teachers' motivation and beliefs in a high-stakes testing context*. [Doctoral dissertation, The Ohio State University, Ohio, United States]. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1338399669&disposition=inline
- Ekinci, A., Bindak, R., & Yıldırım, M. C. (2012). A research regarding the empathic approaches of school managers about professional problems of teachers by pair-wise comparisons method. *Gaziantep University Journal of Social Sciences*, 11(3), 759-776. <http://dergipark.gov.tr/download/article-file/223316>
- Erkuş, A. (2011). *Scientific research process for behavioral sciences*. Seçkin.
- Finkeldei, J. (2016). *The influence of high stakes testing on elementary classroom instruction*. [Doctoral dissertation, Wichita State University, United States]. https://soar.wichita.edu/bitstream/handle/10057/12632/d16008_Finkeldei.pdf?sequence=1
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, 2(2), 1-17. <https://doi.org/10.1002/ets2.12067>
- Fraenkel, J. R., Wallen, N. E. & Hyun, H. H. (2012). *How to design and evaluate research in education*. McGraw-Hill.
- Güler, N., & Anıl, D. (2009). Scaling through pair-wise comparison method in required characteristics of students applying for post graduate programs. *International Journal of Human Sciences*, 6(1), 627-639. <https://www.j-humansciences.com/ojs/index.php/IJHS/article/view/673>
- Güler, N., & İlhan, M. (2017, October). *Comparison of the results obtained from t-test and ANOVA, and many facet Rasch analysis for difference based statistics*. Paper presented at the 13th International Conference on Social Sciences, Vienna.
- Güzeller, C. O., & Eser, M. T., & Aksu, G. (2016). Pair-wise comparison method application via R project and Microsoft Excel. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 96-108. <http://dx.doi.org/10.21031/epod.80072>
- Hirsch, P. J. (2016). *High stakes testing and its effect on teacher methodologies*. [Master's thesis, Caldwell University].
- İlhan, M. (2016). An analysis of researchers' difficulties in quantitative data analysis with the use of pairwise comparisons. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 73-84. <http://dx.doi.org/10.21031/epod.28154>
- İlhan, M., & Güler, M. (2017). The use of Rasch model in Likert types scales: An application on the fear of negative evaluation scale-Student form (FNE-SF). *Trakya Journal of Education*, 8(4), 756-775. <http://dx.doi.org/10.24315/trkefd.357367>
- Johnson, P. (2007). *High stakes testing and No Child Left Behind: Conceptual and empirical considerations*. Long Island Economic & Social Policy Institute, Dowling College School

- of Education, Long Island, NY. http://martincantor.com/files/High%20Stakes_LIESPW_hitepaperMay4.pdf
- Kane, G. (2014). *Accelerating sustainability using the 80/20 rule*. Oxford: Do Sustaniablity.
- Knapp, D. (2010). *A guide to service desk concepts*. Course Technology, Cengage Learning.
- Kornhauser, Z. G. C., Minahan, J., Siedlecki, K. L., & Steedle, J. T. (2014, April). *A strategy for Increasing student motivation on low-stakes assessments*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia. <https://files.eric.ed.gov/fulltext/ED582123.pdf>
- Kruger, L. J., Wandle, C., & Struzziero, J. (2007). Coping with the stress of high stakes testing. *Journal of Applied School Psychology, 23*(2), 109-128, https://doi.org/10.1300/J370v23n02_07
- Kumandaş, H., & Kutlu, Ö. (2015). High stake tests. *Journal of Educational Science Research, 5*(2), 63–75. <http://dx.doi.org/10.12973/jesr.2015.52.4>
- Linacre, J. M. (2012). *Many-facet Rasch measurement: Facets tutorial2*. <http://www.winsteps.com/a/ftutorial2.pdf>
- Linacre, J. M. (2014). *A user's guide to FACETS Rasch–model computer programs*. <https://www.winsteps.com/facetman/pairedcomparisons.htm>
- Marchant, G. J. (2004). What is at stake with high stakes testing? A discussion of issues and research. *The Ohio Journal of Science, 104*(2), 2-7. <https://pdfs.semanticscholar.org/ca83/c29a37b1389c737a35b3454808bacaf97128.pdf>
- Minarechová, M. (2012). Negative impacts of high–stakes testing. *Journal of Pedagogy, 3*(1), 82–100. <https://doi.org/10.2478/v10159-012-0004-x>
- Moses, M. S., & Nanna, M. J. (2007). The testing culture and the persistence of highstakes testing reforms. *Education and Culture, 23*(1), 55-72. <https://doi.org/10.1353/eac.2007.0010>
- Nalbantoğlu Yılmaz, F. (2017). Investigation of the factors affecting occupation choices of high school students with paired comparison method. *Journal of Measurement and Evaluation in Education and Psychology, 8*(2), 224–236. <http://dx.doi.org/10.21031/epod.303882>
- Nartgün, Z. (2006). Öğretmenlik meslek bilgisi derslerinin önem düzeyinin ikili karşılaştırmalarla ölçeklenmesi [Scaling the importance level of teaching profession courses using paired comparisons]. *Abant İzzet Baysal University Journal of Faculty of Education, 6*(2), 161–176. <http://efdergi.ibu.edu.tr/index.php/efdergi/article/view/964>
- Nichols, S. L., & Berliner, D. C. (2005). *The inevitable corruption of indicators and educators through high–stakes testing*. Education Policy Research Unit (EPRU), Education Policy Studies Laboratory, Arizona State University. <https://files.eric.ed.gov/fulltext/ED508483.pdf>
- Pbrreault, G. (2000). The classroom impact of high-stress testing. *Education, 120*(4), 705-710.
- Phelps, R. P. (2003). *Kill the messenger: The war on standardized testing*. New Brunswick, Transaction.
- Polesel, J., Dulfer N., & Turnbull, M. (2012). *The experience of education: The impacts of high stakes testing on school students and their families*. Sydney: Whitlam Institute Report, University of Western Sydney, https://www.researchgate.net/publication/265752469_The_Experience_of_Education_The_impacts_of_high_stakes_testing_on_school_students_and_their_families_Literature_Review
- Ricketts, S.N., Engelhard, G., & Chang, M.L. (2015). Development and validation of a scale to measure academic resilience in mathematics. *European Journal of Psychological Assessment, 33*(2), 79–86. <https://doi.org/10.1027/1015-5759/a000274>
- Ritt, M. (2016). *The impact of high–stakes testing on the learning environment*. [Masters' thesis, St. Catherine University/University of St. Thomas, United States of America]. https://sophia.stkate.edu/cgi/viewcontent.cgi?article=1660&context=msw_papers

- Saito, Y. (2006). Consequences of high stakes testing on the family and schools in Japan. *KEDI Journal of Educational Policy (KJEP)*, 3(1), 101–102.
- Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in Schools*, 50(5), 489-499. <https://doi.org/10.1002/pits.21689>
- Shepard, L., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. <https://files.eric.ed.gov/fulltext/ED337468.pdf>
- Simpson, C. (2016). *Effects of standardized testing on students' well-being*. https://projects.iq.harvard.edu/files/eap/files/c._simpson_effects_of_testing_on_well_being_5_16.pdf
- Spann, P., & Kaufman, D. (2015). *The negative effects of high-stakes testing*. <https://www.luc.edu/media/lucedu/law/centers/childlaw/childed/pdfs/2015studentpapers/Spann.pdf>
- Stanley, G. K. (2004). High stakes hustle: Public schools and the new billion dollar accountability. *The Educational Forum*, 69(1), 8-16. <https://doi.org/10.1080/00131720408984660>
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.). *Making sense of test-based accountability in education* (pp. 79-100). RAND Corporation. <https://larrycuban.files.wordpress.com/2012/01/mr1554-ch4.pdf>
- Stobart, G., & Eggen, T. (2012). High-stakes testing – value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, 19(1), 1-6. <https://doi.org/10.1080/0969594X.2012.639191>
- Taylor, G., Sheppard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost*. (CSE Tech. Rep. 588). Center for Research on Evaluation, Standards, and Student Testing. <https://files.eric.ed.gov/fulltext/ED475139.pdf>
- Togut, T. D. (2004). High stakes testing: *Educational barometer for success, or false prognosticator for failure*. <https://www.harborhouselaw.com/articles/highstakes.togut.htm>
- Turgut, M. F., & Baykul, Y. (1992). *Ölçekleme Teknikleri [Scaling techniques]*. ÖSYM.
- Vogler, K. E., & Virtue, D. (2007). “Just the Facts, Ma’am”: Teaching social studies in the era of standards and high-stakes. *Testing, The Social Studies*, 98(2), 54–58. <https://doi.org/10.3200/TSSS.98.2.54-58>
- Westfall, D. M. (2010). *Parental perceptions of the effects of the high-stakes TAKS test on the home lives of at-risk fifth grade students*. [Doctoral dissertation, University of Houston, Houston, Texas]. <https://uh-ir.tdl.org/handle/10657/170>
- Winters, M. A., Trivitt, J. R., & Greene, J. P. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review*, 29(1), 138-146. <https://doi.org/10.1016/j.econedurev.2009.07.004>
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17. https://doi.org/10.1207/s15326977ea1001_1
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370. <http://www.rasch.org/rmt/rmt83b.htm>

- Wright, W. E. (2002). The effects of high stakes testing in an inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education*, 5(5). <https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1622/663>
- Yaşar, M. (2018). Scaling of ideal teachers characteristics with pairwise comparison judgments according to pre-service teachers' opinions. *International Journal of Assessment Tools in Education*, 5(1), 130–145. <https://doi.org/10.21449/ijate.369233>
- Yeh, S. S. (2005). Limiting the unintended consequences of high-stakes testing. *Education Policy Analysis Archives*, 13(43). <http://dx.doi.org/10.14507/epaa.v13n43.2005>
- Yılmaz, S. (2017). *Evaluating the reflections of high stakes tests on school culture*. [Master's thesis, Trakya University].