Üçüncü Sektör Sosyal Ekonomi Dergisi
Third Sector Social Economic Review
56(3) 2021, 1508-1521
doi: 10.15659/3.sektor-sosyal-ekonomi.21.08.1577

Research Article

Testing Frequency and Student's Mathematics Achievement Relationship

Test Sıklığı ve Öğrencilerin Matematik Başarısı Arasındaki İlişki

Ufuk GÜVEN
Dr. Öğr. Üyesi , Düzce Üniversitesi, Eğitim Fakültesi
Temel Eğitim Bölümü, Sınıf Eğitimi ABD
ufukguven75@gmail.com
https://orcid.org/0000-0003-1977-6426

Makale Gönderme Tarihi	Revizyon Tarihi	Kabul Tarihi
30.03.2021	05.07.2021	05.08.2021

Abstract

The main goal of this study was to investigate the testing effect on student achievement in fourth-grade mathematics by using Trends in International Mathematics and Science Study datasets. Another goal of this study was finding out the ideal quiz frequency (daily, weekly, monthly, no quizzes) and student achievement relationship for fourth-grade mathematics course. The study examined this relationship in four selected countries (Korea, England, Turkey, and Egypt) from different achievement levels. The study obtained data from the Trends in International Mathematics and Science Study 2015 exam and from teacher questionnaires. The study utilized multiple linear regressions to find out whether test frequency is a significant predictor of student achievement in all selected countries. Descriptive results of this study showed that quizzing/testing is a widely used tool around the world. In 2015, 99% of 4th graders took at least one test, 16% of students took tests every day, and only 1% of them did not take any tests. Regression results indicated that testing frequency improved student achievement. This improvement is significant in Korea and Turkey, but not significant in England and Egypt.

Keyword: Testing Frequency, Testing Effect, Student Achievement, Mathematics

Öz

Bu çalışmanın amacı, dördüncü sınıf matematik dersinde test/quiz sıklığı ile öğrenci başarısı arasındaki ilişki Uluslararası Matematik ve Fen Bilimleri Çalışmalarında Eğilimler sınavı aracılığıyla ölçülmüştür. Çalışmanın diğer bir amacı ise dördüncü sınıf matematik dersi için ideal sınav sıklığını (günlük, haftalık, aylık, hiç yok) belirlemektir. Çalışma, seçilen 4 ülkenin (Kore, İngiltere, Türkiye ve Mısır) dördüncü sınıflarında yukarıda belirtilen ilişkiyi araştırmıştır. Çalışma 2015 Uluslararası Matematik ve Fen Bilimleri Eğilimleri sınavından ve öğretmen anketlerinden veri elde etmiştir. Sınav sıklığının farklı başarı düzeylerinden seçilen dört farklı ülkede (Kore, İngiltere, Türkiye ve Mısır) öğrenci başarısı ile ilişkisinin olup olmadığını belirlemek için çoklu doğrusal regresyonlar gerçekleştirilmiştir. Bu çalışmanın betimleyici sonuçları, testlerin seçilen ülkelerde yaygın olarak kullanılan bir araç olduğunu göstermiştir. 2015 yılında 4. sınıf öğrencilerinin% 99'u en az bir test, öğrencilerin% 16'sı her gün test ya da quiz çözerken, sadece% 1'i hiç test ya da quiz çözmemiştir. Regresyon sonuçları, 4. Sınıf matematik derslerinde test sıklığı arttıkça öğrenci başarısının da arttığını ortaya koyarken, bu artış Kore ve Türkiye'de anlamlı bir artış iken İngiltere ve Mısır'da anlamlı bir artış etkisinin bulunmadığı sonucunu göstermektedir.

Anahtar Kelimeler: Test Sıklığı, Test Etkisi, Öğrenci Başarısı, Matematik

Önerilen Atıf /Suggested Citation
Güven, U. 2021 Testing Frequency and Student's Mathematics Achievement Relationship, *Üçüncü Sektör Sosyal Ekonomi Dergisi*, 56(3), 1508-1521

Introduction

Teachers utilize different instructional methods, but one of the common part of each instruction is the assessment. Teachers constantly employ different assessment methods in their teaching activities. Because assessment provides information for students about their progress, and informs teachers about where students may need help, feedback on the effectiveness of instructional practices, and information on student performance (NCTE, 2013). Assessment also provides information for all stakeholders of education on whether goals and standards are being met (Garrison & Ehringhaus, 2016). Assessment techniques are grouped as summative and formative assessment methods. Exit slips, question-answer technic, summarization, concept maps, and quizzes are some of the many formative assessment strategies teachers use in their instructions. A range of assessment methods can be used at different frequencies for formative assessment. The literature confirms that quizzes and tests are routinely used in formative assessment in order to assess and promote students' learning (CERI, 2008; Roediger & Karpicke, 2008; Roediger, Putnam, & Smith, 2011). However, the optimal frequency for quizzing as an assessment and learning tool is not clear. Even though many studies (Roediger and Karpicke, 2006; Gholami & Moghaddam, 2013; Salas-Morera et al., 2012), have been conducted to determine the relationship between testing frequency and student achievement, results of these studies do not provide a clear answer. While some studies shown that frequent quizzes and tests improve student achievement (Azorlosa, 2011; Beaulieu & Zar, 1986, Geiger & Bostow, 1976; Hadsell, 2009; Johnson & Kiviniemi, 2009; Landrum, 2007) some other studies found no significant relationship between frequent testing and student achievement (Azorlosa & Renner, (2006); Bell, Simone, & Whitfield, (2015); Roediger & Karpicke, 2008; Roediger, Putnam, & Smith, 2011; Burns & Vinchur, 1992; Gurung, (2003); Azorlosa, Pero, & Azorlosa (2019); Lumsden, (1976); and Wilder et al., (2001). The literature does not have a consensus on how often teachers should implement tests and quizzes for the best results in student achievement. It is not clear whether daily quizzes work better than weekly quizzes or bi-weekly quizzes. This absence in the literature requires further research to determine best quizzing frequency in fourth-grade mathematics. This study will try to answer the questions of testing effect on student achievement and ideal testing frequency by analysing TIMSS data that provides rich datasets from different countries.

Testing effect or frequent testing terms are commonly used terms to describe giving tests to students very often. These tests might be used for formative or summative assessment purposes. However, in this study, frequent testing terms is used to describe daily or weekly tests and/or quizzes that are utilized for formative assessment, instead of summative tests. Even though some researchers (Gholami & Moghaddam, 2013) have used frequent quizzing term and some other (Shirvani, 2009), used the frequent testing term, both definitions are used to define daily, weekly, or monthly formative tests. Therefore, the frequent testing and frequent quizzing terms are used interchangeably in this study.

Purpose of the Study

There are more than hundreds of factors that affect student achievement. These factors can be classified as student related factors, school related factors, family related factors, and teacher related factors. The test or quiz frequency is one of the teacher related factors that the literature has a gap in the examination of its effect on student achievement in fourth-grade mathematics. The first purpose of this study was to help bridge that gap by investigating the relationship between quiz frequencies and students' learning level in mathematics in four selected countries by using TIMSS 2015 data. Another purpose of this study was examining the relationship between quiz frequency and students' mathematics achievement in different countries, to be able to compare quiz frequency effect in different countries. The study also aimed to reveal test and quiz practices as well as reveal optimum quizzing frequencies for four selected TIMSS 2015 participating countries.

Research Questions

The main research question of this study is:
Is there a significant relationship between quiz frequency and the mathematics achievement of fourth-grade students' achievement scores as measured by the TIMSS 2015?
The study will also find answers to following sub questions:

a) Does quizzing frequency effect change from country to country?
b) What is the optimum quizzing frequency in 4th grade mathematics?

While frequent testing effect on student achievement studies go back to the early 1930s (Hertzberg, Heilman, & Leuenberger, 1932; Keys, 1934; Kulp, 1933; Turney, 1932), recent studies in cognitive psychology reframed studies on testing effect. Two psychologist, Roediger and Karpicke (2006), studied effect of tests on student learning and they found that implementing frequent tests or quizzes into classroom teaching improves learning and retention. They named this strategy as Test Enhanced Learning in Classroom (TELC). Even though this strategy has been using by other researchers, some researchers (Gholami & Moghaddam, 2013; Marshall, 2007) that using frequent tests in classrooms may not very beneficial after all. Gholami & Moghaddam (2013) claim that implementation and scoring tests and quizzes can be time-consuming for teachers and students that will take the class time away from effective instructional time. They also say that testing students very often might tire out students and eventually student interest towards materials and learning in general will be reduced. Another negative notion against frequent testing is raised by Marshall (2007). He says frequent testing does not cause to fruitful and long time learning because teachers only concentrate on the tests and teach to the test, cover subjects that will be on the tests. Since teachers teach for tests and students learn only the subjects that will be on the tests, long time learning does not occur.
Another study was conducted to compare the impact of monthly tests versus weekly tests on students' achievement levels (Dustin, 1971).The study was carried out with control and experimental groups. The control group received monthly tests while the experimental group received weekly tests. Students in both control group and experimental group also received two retention tests at 7th and 10th week after the intervention in order to examine the amount of information students retain in different groups. Dustin (1971) found that students in the experimental group scored higher than control group on both exams. Moreover, Dustin (1971) studied the exam anxiety issue through a student survey in his study. He found that receiving frequent tests decreased test anxiety of students. Salas-Morera, et al. (2012) also studied the relationship between frequent testing and student achievement as well as other variables. They found that giving frequent tests to students increase their achievement. Moreover, frequent testing increases student participation and helps students to keep up with the subject. They also found that students showed a high level of desire in taking the tests (Salas-Morera et al., 2012).

A meta-analysis study was conducted by Phelps (2012) on the effect of frequent testing on student learning. They used studies that only written in English and published between1991-2010. They excluded studies that were carried out in lab settings and studies with paid volunteers in their meta-analysis study. Based on pre- determined criteria, 669 previously conducted research studies on frequent tests were selected. This meta-analysis study found that testing has a positive moderate to high main effect on student achievement. Phelps (2012) also found that small sample sized studies caused bigger effect sizes than large scale sampled studies. Bangerts- Drown et al. (1991) also conducted a meta-analysis study and they found that students that did not take tests scored less that students that took tests. However, the degree of difference in student scores are decreased as number of tests increased. Gholami & Moghaddam (2013) said that giving too much tests reduces instructional time that might explain the decrease in student scores. Even though too much testing reduces instructional time, Bangerts-Drown et al. (1991) found that students develop positive attitudes towards instruction when they are tested frequently. Another study on frequent testing was conducted by Connor-Greene (2002). The study compared daily quizzes versus announced tests. The study collected data through student questionnaires and found that giving

students a few announced tests in a semester led to procrastination and last-minute preparation. However, giving students daily tests encouraged them to complete assignments. The study determined that giving students frequent tests is an essential factor to mptivate students to learn the material.

Method

The current study used correlational research design that is one of the quantitative research models to analyse the relationship between testing frequency and student achievement in 4th grade mathematics.

Participants

The Trends in International Mathematics and Science Study (TIMSS) 2015 assessments was the source of this study. This study selected four countries (Korea, England, Turkey, and Egypt) among TIMSS 2015 4th grade participant countries. Countries selection was completed in two phases. In the first phase, countries were grouped based on their achievement levels. TIMSS determined achievement levels in four categories; advanced level: 625 and above, high level: 550-625, intermediate level: 475-550, and low level: 400-475 (National Center for Educational Statistics, 2015). Even though TIMSS do not categorize countries that have achievement scores of below 400, it can be categorized as under low level. This study ordered countries based on their achievement scores in the first phase and aimed to select one country from each category. However, none of the 4th grade participant countries scored 625 or higher. Therefore, the study selected four countries from other four categories. In the second phase, countries were randomly selected from four achievement categories. Table 1 presents selected countries information in details.

Table 1. Study sample

Countries	Number of Students	Achievement Scores	Achievement Rank	Category Scores	Category Labels
Korea	5.547	606	2nd	550-625	High
England	4.129	518	10th	475-550	Intermediate
Turkey	6.038	458	24th	400-475	Low
Egypt	7.653	392	34th	Below 400	Under Low

Table shows that Korea was selected among high achieving countries, England was selected among intermediate achieving countries, Turkey was selected among low achieving countries, and Egypt was selected among under low achieving countries. A total of 23.367 students participated to TIMSS 2015 administration from four selected countries. In 2015, 39 countries attended to TIMSS exam at fourth grade. Korea was ranked as 2nd, England was ranked as 10th, Turkey was ranked as 24th, and Egypt was ranked as 34th.

Data Collection Tools

Testing frequency variable is derived from TIMSS 2015 teacher questionnaire which consists of 26 main questions and more than 100 sub-questions. Testing frequency question is the 18th question of teacher questionnaire and it was stated as: “In teaching mathematics to this class, how often do you ask students to take a written test or quiz?” Teachers needed to respond to this question with: “Every day or almost every day = 1”, “About half the lessons = 2”, “Some lessons = 3”, and “Never = 4”.

Testing frequency variable is a categorical variable and this categorical variable was dummy coded by The IDB Analyzer tool to enable comparison between testing frequencies. Thus, it can reveal which testing frequency is the most helpful in affecting student achievement in 4th grade mathematics. The constant (reference level) was never testing frequency in hierarchical linear

regression analyses. Table 2 shows the number of students that take tests and quizzes in different quiz frequencies for four selected countries.

Table 2. Testing frequency in four selected countries

Test Frequencies	Korea		England		Turkey		Egypt	
	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Every day or almost every day	400	%7,2	96	%2,3	2133	%35,3	1190	%15,5
About half the lessons	1.211	%21,8	392	%9,5	1887	%31,2	2088	%27,3
Some lessons	3.768	%68	3616	%87,5	1967	%32,5	4176	%54,6
Never	168	%3	25	%0,7	61	%1	199	%2,6
Total	5.547	%100	4.129	%100	6.038	%100	7.653	%100

As can be seen in Table 2, there are some important differences in testing practices among four countries. Giving tests in everyday or almost everyday frequency differ between low achieving countries (Turkey and Egypt) and high achieving countries (Korea and England). Less than 10% of students take tests everyday in high achieving countries but more than 10% percent in low achieving countries. It is even highest testing frequency in Turkey (%35.3 percent of students take tests and quizzes everyday or almost everyday). Another important finding is the total number of students that take tests or quizzes at least once (%97 in Korea, %99 in England and Turkey, and %97,5 in Egypt). It shows that tests and quizzes are common as in class assessment technic in all of the four selected countries.

Data Analysis

TIMSS data structure is constructed as hierarchical where students nested under classrooms. Classrooms are nested under schools, and schools are nested under countries. Therefore, the study implemented hierarchical linear regression method to analyse the relationship between testing frequency and student achievement in 4th grade mathematics. To analyse the data, SPSS in conjunction with The IDB Analyzer, developed by International Association for the Evaluation of Educational Achievement (IEA), tools were utilized. In the first phase of the analysis, a syntax was created, countries and test frequency variable, using merge module of IDB Analyzer and SPSS. In the second phase, linear regression method was conducted through analysis module of IDB Analyzer with the help of SPSS. In this study, testing frequency variable was taken as independent variable and students' math achievement scores were taken as dependent variable.

Research Ethics

TIMSS data is publicly available for every researchers and educators to use through their website. Since the study used secondary dataset, ethical review board permissions are not required for this study.

Findings

As it was mentioned before, SPSS and IDB Analyser tools were used to analyse TIMSS data because SPSS alone is not powerful enough to handle TIMSS data, because TIMSS is constructed hierarchically. It means that students are nested under teachers, teachers are nested under schools, and schools are nested under countries. TIMSS data are connected with linking codes and files

that are created to connect students with teachers, teachers with schools. Unfortunately, SPSS cannot organize these data files correctly. Therefore, TIMSS developed the IDB analyser tool to correctly analyse its hierarchical data. However, SPSS output, in conjunction with IDB Analyser, does not give an ANOVA table in regression analyses (E. J. Gonzales, personal communication, October 27, 2016). Moreover, SPSS output tables does not include p-values in coefficient tables of regression analyses. But, significance tests can be calculated by using standard errors. IADB (2016) says, "If the absolute value of the group difference divided by the standard error of the difference exceeds a t-value of 1.96, the result can be regarded as statistically significant on the 95% level" (p. 25). Therefore, in this study, t-value was used for significance tests.

Descriptive Results

Table 3 indicates the mean scores and standard deviations of testing frequency variable for four selected countries.

Table 3. Mean scores of Testing Frequency for four selected countries

Countries	Means	sd
Korea	2,64	,65
England	2,90	,38
Turkey	2,00	,84
Egypt	2,43	,79

As it seen in Table 3, the lowest mean score of testing frequency was observed in Turkey and highest score was observed in England. It indicates that students take tests or quizzes most often in Turkey and least often in England. Table 3 also indicates that low achieving (Turkey and Egypt) countries implement tests and quizzes more often than high achieving countries (Korea and England).

Regression Results

A hierarchical linear regression analysis was utilized to see the relationship between testing frequency and student achievement in fourth-grade mathematics for four selected countries. Table 4 shows this relationship for each selected country. Table 4 also shows coefficient estimates of testing frequency for each participating country. The results indicate that there is a significant relationship between testing frequency and 4th grade student achievement in mathematics in Turkey and Korea, but no relationships in England and Egypt. Even though there is no significant relationship between testing frequency and student achievement in England and Egypt, the study found that student achievement increases as testing frequency increases in these countries as well. The results also indicate that testing frequency variable explained 1% ($R^2 = .01$) of the variance in Korea and Turkey, but did not explain any ($R^2 = .00$) of the variance in student achievement in England and Egypt.

Table 4. Regression Results

Country	R^2	Variable	Achievement Score	SE B	B	t-value
Korea	.01	Never	586,78	13,03	.	.
		Some lessons	607,77	13,43	,07	1,58
		Half the lessons	596,90	14,23	,07	,71
		Every lesson	620,39	15,84	,05	2,10*
England	.00	Never	474,41	35,44	.	.
		Some lessons	517,03	35,13	,12	1,29
		Half the lessons	521,55	39,94	,13	1,16
		Every lesson	521,84	43,90	,08	1,07
Turkey	.01	Never	347,53	57,89	.	.
		Some lessons	453,71	59,46	,13	1,80
		Half the lessons	457,00	58,55	,12	1,87
		Every lesson	465,74	59,80	,10	1,97*
Egypt	.00	Never	376,65	25,67	.	.
		Some lessons	385,92	26,22		,35
		Half the lessons	400,37	27,38	,04	,87
		Every lesson	395,48	27,39	,05	,69

* Indicates statistical differences

Table 4 also shows that student achievement increased as testing and quizzing frequency increased in all four selected countries. The average score of Korean students that never take a test or quiz is 586,78, students that take test and quizzes in some lessons frequency scored 607,77, student that take tests and quizzes in half the lessons frequency averaged 596,90 points, and students that take tests and quizzes almost everyday scored an average score of 620,39. The only statistical difference occurred between students that never take a test or quiz and students that take tests and quizzes in almost everyday (t value > 1,96±). The average score of British students that never take a test or quiz is 474,41, students that take test and quizzes in some lessons frequency scored 517,03, student that take tests and quizzes in half the lessons frequency averaged 521,55 points, and students that take tests and quizzes almost everyday scored an average score of 521,84. None of these testing frequencies resulted a statistical difference in England. Regression results of the Turkey has a similar pattern with Korea. The average score of Turkish students that never take a test or quiz is 347,53, students that take test and quizzes in some lessons frequency scored 453,71, student that take tests and quizzes in half the lessons frequency averaged 457 points, and students that take tests and quizzes almost everyday scored an average score of 465,74. The only statistical difference occurred between students that never take a test or quiz and students that take tests and quizzes in almost everyday (t value > 1,96±) as it was in Korea. On the other hand, regression results of Egypt is similar with England. Student achievement increased with testing frequency but none of those increases are statistically significant. The average score of Egyptian students that never take a test or quiz is 376,65, students that take test and quizzes in some lessons frequency scored 385,92, student that take tests and quizzes in half the lessons frequency averaged 400,37points, and students that take tests and quizzes almost everyday scored an average score of 395,48. Regression results in table 4 also shows that increase in student

achievement with testing frequency is linear in Turkey and England. However, the linear increase is not observed in Korea and Egypt.

Discussion and Conclusion

The current study found that there is a positive relationship between testing frequency and student achievement in 4th grade mathematics in all of the selected countries. The positive relationship is at statistically significant level in two of the selected countries (Korea and Turkey), but non-significant in the other two selected countries (England and Egypt). According to Roediger's (2006) test enhanced learning in classrooms (TELC) model frequent tests can help students to retain information for a longer time and improve students' academic success. This study found a similar results with Roediger's TELC model. On the other hand, Basol and Johanson's (2009) meta-analyses study found that most studies found a positive but little improvement in student achievement with frequent tests. They (2009) investigated the effect of frequent testing on student achievement and they did not find a significant effect of frequent testing on student achievement. However, the current study found that the relationship between testing frequency and student achievement is significant in two countries but not in all of them. The study also found that tests and quizzes are a common instructional practice because the current study found that they are widely used in four selected countries (See Table 1).

Additionally, the study found that low achieving countries, in general, utilized tests/quizzes more frequently than high achieving countries. It is possible that teachers and students in low achieving countries spend too much time on tests and they have less time for instruction or other instructional activities. Another reason might be the improper use of tests in low achieving countries. These countries may need to analyse assessment practices of high achieving countries. It is known that providing feedback after tests and quizzes improves student learning (Butler and Roediger, 2008). This might be one of the reasons why testing and quizzing work in some countries but not in other countries.

The study also investigated to find optimal test/quiz frequency. It found that giving tests and quizzes every day is the optimal testing frequency in Korea, England, and Turkey. Giving tests about half the lessons frequency is the optimal frequency in Egypt. It can be concluded that giving students test and quizzes every day is the most favourable testing frequency for best achievement scores. However, frequent testing opponents claim that extreme use of tests/quizzes may not be very beneficial for student learning because frequent testing reduces instructional time, leads students to score better on exams, and emphasizes exams rather than learning (Gholami & Moghaddam, 2013). Therefore, teachers need to be careful when implementing tests and quizzes very often. They need to keep in mind that simply increasing test frequency will not increase student achievement. They also need to increase the quality of tests as well as implement other assessment strategies.

Future studies should focus on the countries that significantly benefitted from frequent testing to analyse their applications of tests frequency, how they implement tests and quizzes, when they implement, the goal of tests, type of tests, and etc. in their educational system, so the rest of the globe can learn from their testing practices. Studies can also examine cultural differences, such as attitudes (Erdoğan, 2021) that may prevent the positive effect of frequent tests and quizzes.

References

Azorlosa, J. L. & Renner, C. H. (2006). The effect of announced quizzes on exam performance. Journal of Instructional Psychology, 33, 278-283.

Azorlosa, J. L. (2011). The effect of announced quizzes on exam performance: II. Journal of Instructional Psychology, 38(1), 3.

Azorlosa, J. L., Pero, A. N., & Azorlosa, J. E. (2019). The Effect of Announced Quizzes on Exam Performance: Extra-Credit Quizzes vs. Traditional Quizzes. DOI: 10.13140/RG.2.2.36747.23842

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bangert-Drowns, R. L. Kulik, J. A. & Kulik, G. L. C. (1991). The effects of frequent classroom testing. Journal of Educational Research, 85, 89-99.

Basol, G. & Johanson, G. (2009). Effectiveness of frequent testing over achievement: A meta analysis study. International Journal of Human Sciences, 6(2), 99–121.

Beaulieu, R. P., & Zar, M. C. (1986). The effects of examination frequency on student performance. Journal of Instructional Psychology, 1381-89.

Bell, M. C., Simone, P. M., Whitfield, L. C. (2015). Failure of online quizzing to improve performance in introductory psychology courses. Scholarship of Teaching and Learning in Psychology, 1(2), 163-171.

Butler, A.C. and Roediger, H.L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition 36,* 604-616.

Büyüköztürk, Ş., Akgün, Ö. E., Kahveci, Ö., & Demirel, F. (2004). Güdülenme ve öğrenme stratejileri ölçeğinin Türkçe formunun geçerlik ve güvenirlik çalışması [The validity and reliability study of the Turkish version of the motivated strategies for learning questionnaire]. *Educational Sciences: Theory & Practice, 4*(2), 207-239.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri (18. bs.) [Scientific research methods (18th ed.)]*. Ankara: Pegem Academy Publishing.

CERI (2008). Assessment for learning formative assessment. Retrieved from http://www.oecd.org/site/educeri21st/40600533.pdf

Connor-Greene, P. A. (2000). Assessing and promoting student learning: Blurring the line between teaching and testing. Teaching of Psychology, 27, 2, 84-88.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Dustin, D. S. (1971). Some effects of exam frequency. *The Psychological Record, 21,* 409-414.

Ellis, A. (1973). Rational-emotive psychotheraphy, In H. C. Patterson (Ed.), *Theories of counseling and psychotherapy* (pp. 49-57). New York, Harper and Row Publishers.

Erdoğan, M. (2021). A Research On Determination of Primary School Students' Attitudes Toward

Environment: The Case of Çanakkale. *Third Sector Social Economic Review, 56(1),* 249-271.

Garrison, C. & Ehringhaus, M. (2016). Formative and summative assessment in the classroom. Retrieved from https://www.amle.org/BrowsebyTopic/WhatsNew/WNDet/TabId/270/ArtMID/888/ArticleID/286/Formative-and-Summative-Assessments-in-the-Classroom.aspx

Geiger, O. G., & Bostow, D. E. (1976). Contingency-managed college instruction: Effects of weekly quizzes on performance on examination. *Psychological Reports,* 39, 707-710.

Gholami, V. & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science, 5(1),* 36-41.

Hadsell, L. (2009). The effect of quiz timing on exam performance. *Journal of Education for Business,* 84(3), 135-140.

Hertzberg, O. E., Heilman, J. D., & Leuenberger, H. W. (1932). The value of objective tests as teaching devices in educational psychology classes. Journal of Educational Psychology, 23(5), 371–380. https://doi.org/10.1037/h0072896

Johnson, B. C., Kiviniemi, M. T. (2009). The effect of online chapter quizzes on exam performance in an undergraduate social psychology course. *Teaching of Psychology,* 36(1), 33-37.

Keys, N. (1934). The influence on learning and retention of weekly as opposed to monthly tests. Journal of Educational Psychology, 25, 6, 427-436.

Kulp, D. H. II (1933). Weekly tests for graduate students? School and Society, 38, 157-159.

Landrum, E. R. (2007). Introductory psychology student performance: Weekly quizzes followed by a cumulative final exam. *Teaching of Psychology,* 34(3), 177-180.

LaRoche, S., Joncas, M., & Foy, P. (2016). Sample Design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), Methods and Procedures in TIMSS 2015 (pp. 3.1-3.37). Retrieved from Boston College, TIMSS & PIRLS International Study.

Munneke, L., Andriessen, J., Kirschner, P., & Kanselaar, G. (2007). Effects of synchronous and asynchronous CMC on interactive argumentation. In *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning 16-21 July 2007* (pp. 532-541), New Brunswick, NJ, USA.

National Center for Educational Statistics (2015). International Benchmarks. Retrieved from https://nces.ed.gov/timss/timss15technotes_intlbenchmarks.asp#:~:text=International%20benchmarks%20for%20achievement%20were,score%20of%20555%20or%20480).

NCTE (2013). Formative assessment that truly informs instruction. Retrieved from http://www.ncte.org/positions/statements/formative-assessment/formative-assessment_full

Ong, W. J. (2012). *Sözlü ve yazılı kültür: Sözün teknolojileşmesi* (6. bs.) *[Orality and literacy: The technologizing of the word (6th ed.)]* (Sema Postacıoğlu Banon, trans.). İstanbul: Metis Publishing.

Richard P. Phelps (2012) The Effect of Testing on Student Achievement, 1910–2010, International Journal of Testing, 12:1, 21-43, DOI: 10.1080/15305058.2011.602920

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 1(3),* 181-210.

Roediger H. L., III, Putnam A. L., Smith M. A. (2011). Ten benefits of testing and their applications to educational practice. Psychology of Learning and Motivation, 44, 1–36.Semerci, Ç. (1999). Öğrencilerin öğretmenlik mesleğine ilişkin tutum ölçeği [A scale for students' attitude towards the teaching profession]. *Education and Science, 23*(111), 51-55.

Salas-Morera, L, Arauzo-Azofra, A & García-Hernández, L. 2012. Analysis of online quizzes as a teaching and assessment tool. Journal of Technology and Science Education 2(1), 39-45.

Semerci, Ç., Semerci, N., & Yazıcıoğlu, S. (2000). Duyuşsal boyutun ölçülmesi [Measuring the affective dimention]. In *Proceedings of the 8th National Educational Sciences Congress 01-03 September 2000* (pp. 334). Karadeniz Technical University, Trabzon, Turkey.

Shirvani, H. (2009). Examining an assessment strategy on high school mathematics achievement: Daily quizzes vs. weekly tests. American Secondary Education, 38(1), 34-45.

Turney, A. H. (1931). The effect of frequent short objective tests upon the achievement of college students in educational psychology. School and Society, 33, 760-762

Wilder, D. A., Flood, W. A., & Stromsnes, W. (2001). The use of random extra-credit quizzes to increase student attendance. *Journal Instructional Psychology, 28*, 117-120.

Vicks, M. E. (2013). *An examination of internet filtering and safety policy trends and issues in south carolina's k-12 public schools*. Doctorate Dissertation, Nova Southeastern University, Florida, USA.

Research Article

Testing Frequency and Student's Mathematics Achievement Relationship

Test Sıklığı ve Öğrencilerin Matematik Başarısı Arasındaki İlişki

Ufuk GÜVEN

Dr. Öğr. Üyesi , Düzce Üniversitesi, Eğitim Fakültesi

Temel Eğitim Bölümü, Sınıf Eğitimi ABD

ufukguven75@gmail.com

https://orcid.org/0000-0003-1977-6426

Genişletilmiş Özet

Giriş

Değerlendirme, her öğretim yönteminin temel parçalarından biridir. Öğretmenler, öğretim faaliyetlerinde sürekli olarak farklı değerlendirme yöntemleri kullanırlar. Çünkü değerlendirme, öğrencileri kendi öğrenmeleri hakkında bilgilendirir ve öğretmenleri öğrenci performansı, öğretim uygulamaları ve öğrencilerin yardıma ihtiyaç duydukları konular hakkında bilgilendirir (NCTE, 2013). Değerlendirme aynı zamanda hedeflere ve standartlara ulaşılıp ulaşılmadığı konusunda eğitimin tüm bileşenlerini bilgilendirir (Garrison ve Ehringhaus, 2016). Değerlendirme yöntemleri, özetleyici ve biçimlendirici değerlendirme yöntemleri olarak gruplandırılmıştır. Çıkış fişleri, soru-cevap tekniği, özetleme, kavram haritaları ve quizler, öğretmenlerin derslerinde kullandıkları biçimlendirici değerlendirme stratejilerinden bazılarıdır. Biçimlendirici değerlendirme amacıyla farklı sıklıklarda bir dizi değerlendirme yöntemi kullanılabilir. Literatür, quizlerin ve testlerin rutin olarak öğrencilerin öğrenmesini değerlendirmek ve teşvik etmek için biçimlendirici değerlendirmede kullanıldığını göstermektedir (CERI, 2008; Roediger ve Karpicke, 2008; Roediger, Putnam ve Smith, 2011). Bununla birlikte, bir değerlendirme ve öğrenme aracı olarak quiz ve testler için en uygun sıklık net değildir. Test sıklığı ile öğrenci başarısı arasındaki ilişkiyi belirlemek için birçok çalışma (Roediger ve Karpicke, 2006; Gholami ve Moghaddam, 2013; Salas-Morera vd., 2012) yapılmış olsa da, bu çalışmaların sonuçları net bir bilgi vermemektedir. Bazı araştırmalar, sık sık yapılan sınavların ve testlerin öğrenci başarısını artırdığını gösterirken (Azorlosa, 2011; Beaulieu ve Zar, 1986, Geiger ve Bostow, 1976; Hadsell, 2009; Johnson ve Kiviniemi, 2009; Landrum, 2007) diğer bazı araştırmalar test sıklığı ve öğrenci başarısı arasında anlamlı bir ilişki bulamamıştır (Azorlosa & Renner, (2006); Bell, Simone ve Whitfield, (2015); Burns & Vinchur, 1992; Gurung, (2003); Azorlosa, Pero ve Azorlosa (2019); Lumsden, (1976); ve Wilder vd., (2001)). Görüldüğü gibi literatürde, öğrenci başarısında en iyi sonuçları elde etmek için öğretmenlerin testleri ve quizleri ne sıklıkla uygulaması gerektiği konusunda fikir birliği yoktur. Günlük sınavların haftalık sınavlardan veya aylık yapılan sınavlardan daha iyi çalışıp çalışmadığı net değildir. Literatürdeki bu eksiklik, dördüncü sınıf matematikte en iyi sınav sıklığını belirlemek için daha fazla araştırma yapılmasını gerektirir. Bu çalışma, farklı ülkelerden zengin veri setleri sağlayan TIMSS verilerini analiz ederek farklı ülkelerde test ve quiz sıklığının öğrenci başarısı üzerindeki etkisini incelemek ve ideal test sıklığını bulmak için yapılmıştır.

Çalışmanın Amacı

Öğrenci başarısını etkileyen yüzlerce faktör vardır. Bu faktörler öğrenciyle ilgili faktörler, okulla ilgili faktörler, aile ile ilgili faktörler ve öğretmenle ilgili faktörler olarak sınıflandırılabilir. Test ve quiz verme sıklığı dördüncü sınıf matematik derslerinde öğrenci başarısı üzerindeki etkisi öğretmenle ilişkili faktörlerden biridir ve bu faktörün ne kadar etkili olduğu alanyazın tarafından net bir şekilde cevaplandırılamamıştır. Bu çalışmanın ilk amacı, TIMSS 2015 verilerini kullanarak seçilen dört ülkede sınav sıklıkları ile öğrencilerin matematikteki başarı düzeyi arasındaki ilişkiyi araştırarak alanyazında ki boşluğu kapatmaya yardımcı olmaktır. Bu çalışmanın bir diğer amacı, farklı ülkelerdeki sınav sıklığı etkisini karşılaştırabilmek için sınav sıklığı ile öğrencilerin matematik başarısı arasındaki ilişkiyi farklı ülkelerde incelemektir. Çalışma aynı zamanda farklı ülkelerin test ve sınav uygulamalarını ortaya çıkarmayı ve seçilen dört TIMSS 2015 katılımcı ülkesi için optimum sınav sıklıklarını ortaya çıkarmayı hedeflemektedir.

Yöntem

Bu çalışmada, 4. sınıf matematik dersinde öğrenci başarısı ile test sıklığı arasındaki ilişkiyi analiz etmek için nicel araştırma modellerinden biri olan ilişkisel araştırma tasarımı kullanılmıştır. Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) 2015 değerlendirmeleri bu çalışmanın veri kaynağını oluşturmaktadır. Bu araştırma verileri ikincil bir veri kaynağından elde ettiği için etik kurul onayına gerek duyulmamaktadır. Çalışmada, TIMSS 2015 4. sınıf katılımcı ülkeler arasından dört ülke (Korea, İngiltere, Türkiye ve Mısır) seçilmiştir. Ülke seçimi iki aşamada tamamlanmıştır. İlk aşamada ülkeler başarı seviyelerine göre gruplandırıldı. TIMSS başarı düzeylerini dört kategoride belirlemektedir; ileri seviye: 625 ve üstü, yüksek seviye: 550-625, orta seviye: 475-550 ve düşük seviye: 400-475 (Ulusal Eğitim İstatistikleri Merkezi, 2015). TIMSS başarı puanları 400'ün altında olan ülkeleri kategorize etmese de, bu seviyenin altında puan alan ülkeler düşük altı seviye olarak kategorize edilebilir. Bu çalışma, ülkeleri ilk aşamada başarı puanlarına göre sıraladı ve her kategoriden bir ülke seçmeyi amaçladı. Ancak, katılan ülkelerin hiçbiri 625 veya üzeri puan alamadığı için bu seviyeden herhangi bir ülke seçilemedi. Bu nedenle, çalışma diğer dört kategoriden birer ülke seçme yoluna gidildi. İkinci aşamada, dört farklı kategoride buluna ülkelerden 4 ülke (Korea, İngiltere, Türkiye ve Mısır) rastgele seçilerek araştırmaya dahil edildi. Korea, yüksek başarı gösteren ülkeler arasından, İngiltere orta düzeyde başarılı ülkeler arasından, Türkiye düşük başarılı ülkeler arasından ve Mısır, düşük başarı gösteren ülkeler arasından seçilerek araştırmaya dahil edildi. Seçilen bu dört ülkeden toplam 23.367 öğrenci TIMSS 2015 sınavına katıldı. 2015 yılında 39 ülke dördüncü sınıfta TIMSS sınavına katıldı. Sınav sonuçlarına göre Korea 2. sırada, İngiltere 10. sırada, Türkiye 24. sırada ve Mısır 34. sırada yer aldı.

TIMSS veri yapısı, öğrencilerin sınıfların altına yerleştirildiği hiyerarşik olarak oluşturulmuştur. Sınıflar okulların altında, okullar ise ülkelerin altında yer alır. Bu nedenle, çalışmada 4. sınıf matematikte öğrenci başarısı ile test sıklığı arasındaki ilişkiyi analiz etmek için hiyerarşik doğrusal regresyon yöntemi uygulanmıştır. Verileri analiz etmek için Uluslararası Eğitim Başarısını Değerlendirme Derneği (IEA) tarafından geliştirilen IDB Analyzer programı ile birlikte SPSS programlarından yararlanılmıştır. Analizin ilk aşamasında, ülkeler ve test sıklığı değişkeni kullanılarak IDB Analyzer ve SPSS'nin birleştirme modülü kullanılarak bir syntax oluşturuldu. İkinci aşamada ise SPSS yardımıyla IDB Analyzer'ın analiz modülü üzerinden doğrusal regresyon yöntemi uygulandı. Bu çalışmada, test sıklığı değişkeni bağımsız değişken olarak ve öğrencilerin matematik başarı puanları ise bağımlı değişken olarak alınmıştır.

Bulgular

Betimleyici analizler seçilen dört ülke için test sıklığı değişkeninin seçilen 4 ülkede uygulanma sıklığını ortaya koymuştur. Bu analizlere göre en düşük ortalama test sıklığı puanı Türkiye'de, en yüksek puan İngiltere'de görülmüştür. Öğrencilerin en çok Türkiye'de ve en az İngiltere'de test ve quiz çözdüklerini göstermektedir. Betimleyici analizler aynı zamanda, düşük başarı gösteren

ülkelerin (Türkiye ve Mısır), test ve quizleri, yüksek başarı gösteren ülkelere (Korea ve İngiltere) göre daha sık uyguladıklarını göstermektedir.

Regresyon analizleri ise matematik derslerinde test sıklığı ile 4. sınıf öğrenci başarısı arasında Türkiye ve Korea'de anlamlı bir ilişki olduğunu, ancak İngiltere ve Mısır'da hiçbir ilişki olmadığını göstermektedir. İngiltere ve Mısır'da sınav sıklığı ile öğrenci başarısı arasında anlamlı bir ilişki olmasa da, çalışma bu ülkelerde de sınav sıklığı arttıkça öğrenci başarısının arttığını bulmuştur. Sonuçlar ayrıca test sıklığı değişkeninin Korea ve Türkiye'deki varyansın% 1'ini (R^2 = .01) açıkladığını, ancak İngiltere ve Mısır'daki öğrenci başarısındaki varyansın %0'ını (R^2 = .00) açıkladığını göstermektedir. Regresyon analizleri ayrıca seçilen dört ülkenin tamamında test ve quiz sıklığı arttıkça öğrenci başarısının da arttığını göstermektedir.

Tartışma ve Sonuç

Bu çalışma, seçilen tüm ülkelerde 4. sınıf matematikte test sıklığı ile öğrenci başarısı arasında pozitif bir ilişki olduğunu bulmuştur. Pozitif ilişki, seçilen ülkelerin ikisinde (Korea ve Türkiye) istatistiksel olarak anlamlı düzeyde iken seçilen diğer iki ülkede (İngiltere ve Mısır) anlamlı değildir. Roediger'in (2006) testlerle arttırılan öğrenme (TELC) modeline göre, sık verilen testler öğrencilerin bilgileri daha uzun süre saklamalarına ve öğrencilerin akademik başarısını artırmalarına yardımcı olabilir. Bu çalışma, Roediger'in TELC modeliyle benzer sonuçlar bulmuştur. Öte yandan, Basol ve Johanson'un (2009) meta-analiz çalışması, çoğu araştırmanın, sık yapılan testlerle öğrenci başarısında olumlu ancak çok az bir iyileşme bulduğunu bulmuştur. (2009), sık yapılan testlerin öğrenci başarısı üzerindeki etkisini araştırmış ve sık yapılan testlerin öğrenci başarısı üzerinde anlamlı bir etkisini bulamamışlardır. Bununla birlikte, mevcut çalışma, test sıklığı ve öğrenci başarısı arasındaki ilişkinin iki ülkede önemli olduğunu, ancak hepsinde olmadığını ortaya koymuştur. Çalışma ayrıca testlerin ve quizlerin yaygın bir değerlendirme aracı olduğunu bulmuştur çünkü bu çalışmada seçilen dört ülkenin tamamında ki öğrencilerin en az bir test ya da quiz çözme oranı % 97'nin üzerinde olduğu bulunmuştur.

Gelecekteki çalışmalar, test çözme uygulamalarının öğrenci başarısıyla ilişkisini daha net ortaya koyabilmek için testlerden önemli ölçüde yararlanan ülkelere odaklanarak, onların eğitim sistemlerin de testleri ve quizleri nasıl uyguladıkları, ne zaman uyguladıkları, testlerin amacı, testlerin türü gibi konuları ortaya çıkararak diğer ülkelerinden bunlardan faydalanması sağlanabilir. Çalışmalar ayrıca sık yapılan testlerin ve sınavların olumlu etkisini engelleyebilecek kültürel farklılıkları da inceleyebilir.