International Journal of Assessment Tools in Education
2023, Vol. 10, No. 1, 56–75

https://doi.org/10.21449/ijate.981367

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate **Research Article**

Comparison of Kernel equating methods under NEAT and NEC designs

Seyma Nur Ozsoy[1,*], **Sevilay Kilmen**[2]

[1]Ankara University, PhD Student of Educational Sciences, Ankara, Türkiye
[2]Bolu Abant İzzet Baysal University, Faculty of Education, Department of Educational Sciences, Bolu, Türkiye

ARTICLE HISTORY

Received: Aug. 11, 2021

Accepted: Feb. 02, 2023

Keywords:

Test equating,
Kernel equating,
NEAT design,
NEC design,
Bandwidth.

Abstract: In this study, Kernel test equating methods were compared under NEAT and NEC designs. In NEAT design, Kernel post-stratification and chain equating methods taking into account optimal and large bandwidths were compared. In the NEC design, gender and/or computer/tablet use was considered as a covariate, and Kernel test equating methods were performed by using these covariates and considering bandwidths. The study shows that, in the NEAT design, Kernel chain equating methods exhibit higher error than the post-stratification equating methods do since the lowest error in the NEC design was obtained from the Kernel equating method with large bandwidth through the computer/tablet variable. Kernel test equating results based on the NEC design, which considers gender and computer tablet use variables as a covariate separately, showed lower SEE than that of the NEC pattern, which takes these variables together as covariates. In terms of the bandwidth, when all methods are compared within the pattern used (i.e., NEAT and NEC), it has been seen that generally Kernel test equating with large bandwidth results in fewer errors than the Kernel test equating with optimal bandwidth. When the NEAT and NEC designs are compared generally, the NEAT design has a lower SEE than that of the NEC design.

1. INTRODUCTION

In some testing practices, different test forms are used in different groups to provide test reliability. These tests consisting of different items bring along some equivalence discussions due to varying difficulties. Therefore, the need to equate tests arises in order to prevent injustice in comparing tests.

The concept of test equating has been defined and studied by many researchers for many years and still continues to be among the current research (Kolen & Brennan, 2004; von Davier et al., 2004b; Livingston, 2014). Test equating is accepted as a statistical process used by individuals who are subjected to the same assessment process to make the scores obtained from many forms of this assessment into comparable state (von Davier, 2013; Kolen & Brennan, 2004) since such a process eliminates discussions about which form of test individuals will take because differences between the obtained scores depending on the test form are prevented (Lord, 1980).

*CONTACT: Seyma Nur Ozsoy 1seymaozsoy@gmail.com Ankara University, phD Student of Educational Sciences, Ankara, Türkiye

e-ISSN: 2148-7456 /© **IJATE** 2023

Test equating is mainly divided into two categories, namely equating with observed score and true score (Lord, 1980). The observed scores equating is performed with just observed scores and includes equal percentage equating and chain equating approaches (Kolen & Brennan, 1995). On the other hand, in the true score equating, the true score covers the observed score and the standard error. Among the scaling/calibration methods for true score equating, there are approaches such as mean-mean, mean-standard deviation, and Stocking-Lord (Kolen & Brennan, 2004).

Both true score and observed score equating possess limitations. As the true score equating requires assumptions such as large sample size and local independence, in practice using it to equate different test forms can be too hard while in observed score equating using discrete distributions can cause increase in equating errors. To overcome these limitations, the Kernel equating method, a relatively new method, is recommended as an observed score equation method in which score distributions are equated by converting discrete score distributions into continuous distributions by using Gauss Kernel approach instead of the linear approach (von Davier et al., 2004a) because Kernel equating offers more realistic assumptions than the other methods do (Godfrey, 2007). Furthermore, due to the pre-smoothing, Kernel equating gives less standard error compared to other methods, is less dependent on sample size, and can be applied to all designs and equating functions (von Davier et al., 2004b).

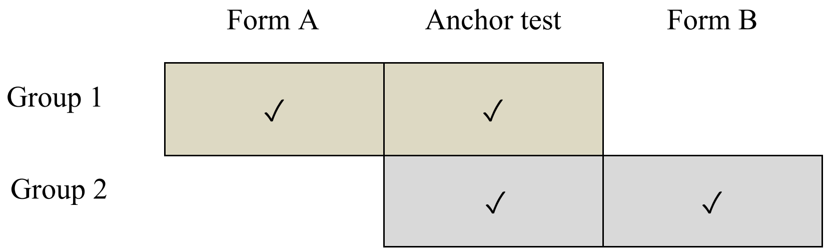
On the other hand, test equating generally requires applying an anchor test to different groups that take different tests. This test equating design is called a nonequivalent groups anchor test (NEAT). However, specifically in examinations that are applied several times in a year or term, using the same anchor test sometimes can cause some problems; for example, the use of the same items repeatedly can lead to recall of items for individuals, which can negatively affect discrimination. Recently, as a solution to this problem, there are studies suggesting that test equating can be conducted by using nonequivalent groups with covariates (NEC) design (e.g., Akın Arıkan, 2020; Albano & Wiberg, 2019; Branberg, 2010; Branberg & Wiberg, 2011; Gonzales et al., 2015; Wiberg & Branberg, 2015; Wiberg & von Davier, 2017). For example, Yurtçu (2018) equated scores obtained from different tests by using common item scores, gender, and mathematics self-efficacy scores as covariates. Their results showed that common variables could be used instead of common items to equate test scores obtained from different tests. Akın Arıkan (2020) compared NEAT design and NEC designs using gender and socioeconomic status variables as covariance variables and their study results indicated that NEC design could be taken as a practically viable alternative to the NEAT design in Kernel equating to establish the comparability of the test scores. Notwithstanding the proven utility of the NEC design for obtaining comparable test scores from different groups under the Kernel equating in a limited number of studies, it still remains a question about whether this approach can be used instead of anchor items. Therefore, there is still need for more studies that compare Kernel equating results in NEC design and NEAT design. To this end, the present study focuses on comparing the performance of Kernel test equating methods under NEAT and NEC design.

1.1. NEAT and NEC Design in Test Equating

1.1.1. *Nonequivalent groups with anchor tests (NEAT) design*

In NEAT design, common items in different forms are used to equate test scores obtained from different tests as can be seen in Figure 1. These forms are applied in different groups who do not know the equivalence due to such features as the number of individuals and item order (von Davier et al., 2004b). Anchor test is prepared in accordance with the characteristics of the main test forms. For these common substances to have a similar effect in both forms, the item numbers must also be the same (Kolen, 1988; Kolen & Brennan, 2004). Two test forms are equated by using the anchor test.

Figure 1. *NEAT design.*

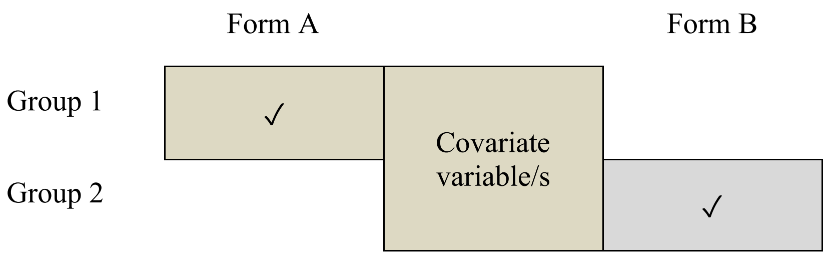


Post-stratification and chain equating in NEAT design: In this design, information from the anchor items can be provided by two different approaches, namely post-stratification or chain equating approaches. The approach in which anchor test score is used as a conditioning variable (or a covariate) for estimating the score distributions is called the poststratification approach. In this method, the conditional distributions of the X form given anchor test and of the Y form given anchor test are weighted by distribution for anchor test to estimate the score distributions for X form and Y form in a hypothetical target population (T) (von Davier & Chen, 2013). In T denoted as (wP + (1 – w)Q), w is the proportion of T that comes from P, (Braun & Holland, 1982). The second approach, the chain equating approach (von Davier et al., 2006), involves a two-stage process for the transformation of the scores of form X into scores of form Y (von Davier et al., 2004a). In Kernel chain equating, first, the X form is linked to the common items and then the common items are linked to the Y form to ensure equating (Andersson et al., 2013). An important difference between post-stratification equating and chain equating is that in the former there is an explicit target population (T) whereas in the latter T plays no explicit role (von Davier & Chen, 2013). In the present study, Kernelpost-stratification and chain equating methods under NEAT design were used.

1.1.2. *Nonequivalents groups with covariates (NEC) design*

NEAT design may not be used in many test applications for such reasons as test security and recognizing the items which are used in the anchor test of previous test applications (Wiberg, 2015). Branberg and Wiberg (2011) recommended using covariate variables to equate the two different test forms and conducted various studies using the NEC design. In the NEC design, the scores obtained from different tests are equated with the covariate variable/s associated with the test scores (see Figure 2). Covariates are considered similar to the common item scores used in the NEAT design (Wiberg, 2015). The most important feature of covariates is that they are categorical. In many studies where continuous variables are used, the variables are categorical by methods such as cluster analysis. Therefore, in the present study, gender and having a computer/tablet as variables are discussed. The gender variable is important because it is related to the learned roles of women and men in the field of science, and the computer/tablet use variable is important because it allows access to today's information.

Figure 2. *NEC design.*



1.2. Kernel Test Equating in NEAT and NEC Design

Kernel equating was first recognized by Livingston (1993) with his test equating study using log-linear smoothing. Kernel equating method is an observed score equation method in which score distributions are equated by converting discrete score distributions into continuous distributions. In these conversions Kernel equating uses Gauss Kernel approach instead of linear approach which is used in the traditional observed score equating (von Davier et al., 2004a). Kernel equating is preferred to traditional test equating methods for at least four reasons: The first is that it has realistic assumptions than other methods (Godfrey, 2007); the second is that due to the pre-smoothing, it gives less standard equating error compared to other methods; the third is that it is less dependent on sample size; and lastly, it can be applied to all designs and equating functions (von Davier et al., 2004b). Kernel equating is carried out in a five-step process (von Davier et al., 2004b), which includes pre-smoothing, estimation of score probabilities, continuization, equating, and calculation of equaling error. Kernel test equating process was explained for both NEAT and NEC designs separately as consistent with the aim of this study.

In the first step of Kernel equating, pre-smoothing is performed in order to reduce complexities in the observed score distributions depending on the sampling. In this step, the data are linked with log linear model (von Davier et al., 2004a). This process is the same in both NEAT and NEC designs. In the second step, score probability is estimated. Score probability estimation varies according to the equating design used as mentioned before. In NEAT design, score probability is estimated by common items, while in NEC design it is estimated by common categorical variable/s. Moreover, when score probabilities are estimated by using anchor test in the NEAT design, two different approaches are used, namely poststratification equating and chain equating (von Davier et al., 2006). In the present study, both approaches were used to see the possible effects of these approaches on the SEEs and to compare them. In the third step, discrete score distributions are made continuous. This process is performed in order to produce two cumulative frequency distributions. Gauss Kernel is commonly used to make the discrete distributions continuous in Kernel equating studies. In addition, in this step, the bandwidth (h parameter) is determined to make the discrete distributions continuous (Gonzales & Wiberg, 2017). The bandwidth can be chosen in two ways as optimal or large bandwidths (von Davier et al., 2006). In the current study, both optimal and large bandwidths were used to see the possible effect of the bandwidth on SEE results. In the fourth step, equating is performed between continuous distributions by using the Kernel equating methods. The Kernel equating function in which an X form is equal to the Y form is as follows (Andersson et al., 2013):

$$\hat{e}_y(x) = G_{hy}^{-1}(F_{hx}(x;\hat{r});\hat{s})$$

$$= G_{hy}^{-1}\left(F_{hx}(x_j)\right)$$

F_{hx} and G_{hy}: Cumulative distribution function
hx and hy: Bandwidths for test x and test y
r and s: Score probabilities for test x and test y

In the last step, equating error is obtained by calculating SEEs in Kernel equating. The SEE obtained by equating the X form to the Y form is calculated using the equation below (Andersson et al., 2013; Gonzales & Wiberg, 2017:

$$SEE_Y(x) = \sqrt{Var(\hat{e}_Y(x))}$$

$$SEE_X(y) = \sqrt{Var(\hat{e}_X(y)}$$

1.3. Studies Comparing NEAT and NEC Design

In Kernel test equating studies, it was seen that the NEAT design was commonly used. Over the last decade, NEC design with Kernel equating has been used; for example, Branberg (2010) investigated the use of NEC design in test-equating studies and obtained important findings of the use of covariates in the absence of an anchor test. In another study conducted by Branberg and Wiberg (2011), it was revealed with simulated data that the variables of gender and educational status can reduce the amount of test equating error. Strong evidence was also obtained showing that covariate variable/s can be used to equate different tests. In another study conducted by Gonzales, Barrientos, and Quintana (2015) gender and school type were used as covariates in NEC design. The results presented supportive evidence to previous studies that revealed that covariates can be used in test equating studies. Wiberg and Branberg (2015) compared equated scores obtained from NEC design and equated group design and their study results showed that when common variables are used together with common items, they give fewer errors. Wiberg and von Davier (2017) examined anchor tests using age, gender, and education as covariates. The results obtained in their study indicated that even if the composition of the group taking the exam changes, test results can be controlled. In the study conducted by Albano and Wiberg (2019), in which gender was used as a common variable, it was determined that frequency estimation gives less error in the presence of anchor test and covariate variables. Moreover, recent studies comparing NEAT and NEC designs show that common variables can be used instead of common items. For example, in a test equating study conducted by Yurtçu (2018), gender and mathematics self-efficacy scores were used to equate test scores besides the anchor test and the study results presented evidence that common variables can be used instead of common items. Akın Arıkan (2020) made a comparison of the NEAT design and NEC design using gender and socioeconomic status variables as covariance variables and concluded that in the absence of anchor tests, equating can be made by using covariate variables.

In sum, such studies examined NEC design and compared NEC design with NEAT design to find an alternative to anchor tests in conditions in which NEAT design cannot be used. With an aim to contribute to these studies, in this current study, Kernel equating methods under both NEAT and NEC designs were compared according to their standard errors of equating (SEE). Two booklets numbered 1 and 14 out of 14 different booklets used in the Türkiye sample of the TIMSS 8th grade science test applied in 2019 were used to compare Kernel equating methods under both NEAT and NEC designs. In this present study, gender is considered as a covariate variable for the NEC design. In addition to gender, considering the transition to eTIMSS application in 2019, the use of a computer/tablet use is also considered as a covariate variable.

1.4. The Present Study

In the current study, in NEAT design post-stratification equating and chain equating were compared and in NEC design, gender and computer/tablet use were considered as covariates. On the other hand, the selection of bandwidth was considered as a variable that could affect SEEs. Two bandwidths were used in this study, namely optimal and large bandwidths. Depending on these conditions, ten Kernel test equation SEEs were examined (see Table 1). Consequently, the present study examines the role of test equating methods, bandwidths, and the use of covariate/s on SEEs. Accordingly, three research questions are formulated:

1) Which Kernel test equating method gives less SEE when equating scores are obtained from different TIMSS booklets under NEAT design?
2) Which covariate gives less SEE when equating scores are obtained from different TIMSS booklets under NEC design?
3) How does the selection of bandwidth (optimal and large bandwidth) affect SEE when equating scores are obtained from different TIMSS booklets in both NEAT and NEC design?

As mentioned earlier, the current study focuses on comparing ten Kernel test equating conditions under NEAT and NEC designs according to their SEEs. Examining the SEEs between different Kernel equating methods under different test equating designs is crucial for at least three reasons. First, it has been known that different Kernel test equating results give different SEEs. To compare these results and create some advice about which test equating methods are more proper and in which situation, these test equating methods should be examined in various test conditions. Therefore, there is need to conduct further studies addressing test equating method comparisons in various test conditions. The current study therefore compares test equating methods by focusing on Kernel test equating, which is used under conditions that can be met in practice.

Second, although there is a number of studies that compare Kernel equating methods, except for limited research (e.g., Choi, 2009; Liang & von Davier, 2014), there is lack of research examining the performance of Kernel equating regarding the choice of bandwidth and how choices on bandwidth affect equating results in terms of SEE. Relevant literature shows that the bandwidth parameter determines the smoothness of the continuized score distributions and has a large effect on the Kernel density estimate. Relevant research results also show that there is a need to investigate how the bandwidths affect the equating results more rigorously and also to identify certain test scenarios where each different bandwidth method is particularly suitable (e.g., Wallin et al., 2018). Therefore, by considering that it is reasonable to claim that selection of bandwidth could have a noteworthy role in the performance of Kernel equating methods, the present study examines the role of bandwidth selection on Kernel test equating methods' SEEs on TIMSS data.

Third, as an alternative to NEAT test design, relevant literature shows that test equating can be conducted by using NEC design (e.g., Albano & Wiberg, 2019; Akın Arıkan, 2020; Branberg, 2010; Branberg & Wiberg, 2011; Gonzales et al., 2015; Wiberg & Branberg, 2015; Wiberg & von Davier, 2017). Indeed, a number of studies revealed that the covariate variables can reduce the amount of equating error (e.g., Branberg & Wiberg, 2011) and covariates can be used when equating different tests (e.g., Gonzales et al., 2015). Furthermore, in some studies, results showed that common variables can be used instead of common items. For example, in Yurtçu's (2018) study, scores were equated with gender and mathematics self-efficacy scores as covariates and common item scores and the study results showed that common variables can be used instead of common items. Akın Arıkan (2020) made a comparison with the NEAT design and NEC design using gender and socioeconomic status variables as covariance variables and concluded that in the absence of anchor tests, equating can be made by using common variables. Therefore, it may be argued that using covariates instead of common items in test equating may have a role in SEEs when equating scores are obtained from different TIMSS booklets. Hence, examining the role of test equating methods and bandwidths by taking into account test equating designs (i.e., NEAT and NEC) on SEEs when equating scores obtained from different TIMSS booklets is important to decide the eligible test equating approach.

2. METHOD

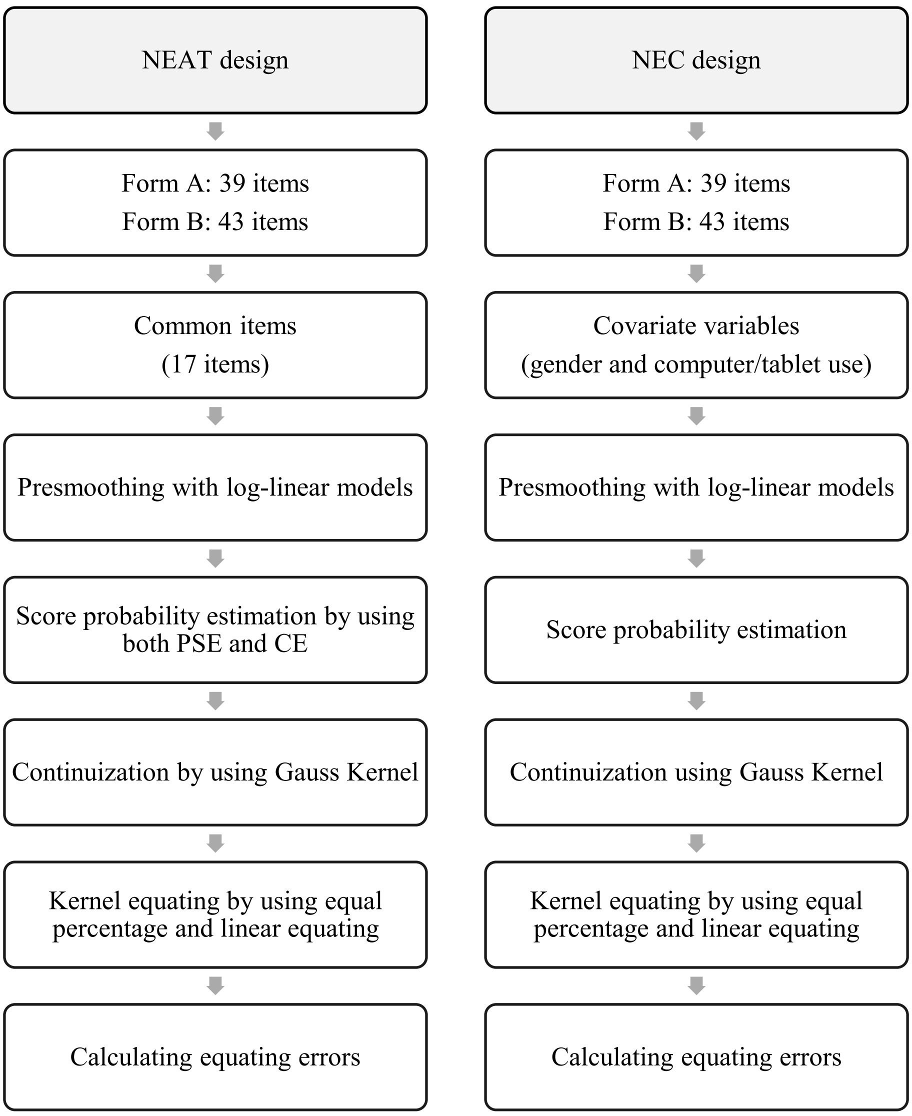
2.1. Study Group

In the study, two booklets numbered 1 and 14 out of 14 different booklets used in the Türkiye sample of the TIMSS 8th grade science test applied in 2019 were included in the analysis. 288 and 295 students took the specified tests, respectively. However, those students who did not answer the items in the student questionnaire were not included in the analysis. Therefore, the study group of the research consisted of 577 students, of whom 284 answered booklet number 1 and 293 answered booklet number 14.

2.2. Procedure

In the study, NEAT and NEC designs were used for Kernel test equating. The first booklet has 43 items, 17 of which are common, and the 14th booklet has 39 items, 17 of which are common. Science data belonging to booklets numbered 1 and 14 were converted into items with double scores as 1-0. For this, correct, partial credit, and full credit answers were coded as 1 point, while blank or wrong answers were coded as 0 point. While the gender variable was coded as 1=Girl and 2=Male, the computer/tablet variable was coded as 1=Yes and 2=No. In addition, gender and computer/tablet use variables were used as covariates in the NEC design in this study (see Figure 3).

Figure 3. *Research schema.*



For test equating methods in both NEAT and NEC designs, in the first stage, the datasets of the two groups were smoothed with log-linear models. In the second stage, the score probability distributions were estimated using the smoothed score distributions obtained in the first stage. At this stage, score probability estimation was made by means of chain and post-stratification equating in the NEAT design. Chain equating starts by creating two separate single group patterns. Then, the first test form is linked to the common items, and from the common items to the other test form. In the post-stratification equating, the two groups are combined to form the target population. In the post-stratification equating, marginal distributions in the target population were obtained for the two test forms. In the third stage, the continuation stage, the Gaussian Kernel was used to make the discrete score distributions continuous in both NEAT and NEC designs. In the fourth stage, the tests were equalized by using the optimal and large bandwidths between the score distributions that became continuous in both NEAT and NEC designs. Finally, the SEE value was calculated.

In sum, in the current study, ten Kernel test equating methods were compared under NEAT and NEC designs. Kernel test equating methods used in the NEAT design are Kernel post-stratification equating with optimal bandwidth, Kernel post-stratification with large bandwidth, Kernel chain equating with optimal bandwidth, and Kernel chain equating with large bandwidth. Kernel equating methods used in the NEC design are Kernel equating with optimal bandwidth using gender as a covariate, Kernel equating with optimal bandwidth using computer/tablet use as a covariate, Kernel equating with large bandwidth using gender as a covariate, Kernel equating with large bandwidth using computer/tablet use as a covariate, Kernel equating with optimal bandwidth using both gender and computer/tablet use as covariates, and lastly Kernel equating with large bandwidth using both gender and computer/tablet use as covariates (see Table 1).

Table 1. *Ten different Kernel test equating methods compared in the present study.*

		NEAT design		Bandwidth		NEC design	
		Chain equating	Post-stratification equating	Optimal	Large	Gender	Computer/ tablet use
NEAT design	Chain equating with optimal bandwidth under NEAT design	x		x			
	Chain equating with large bandwidth under NEAT design	x			x		
	Post-stratification equating with optimal bandwidth under NEAT design		x	x			
	Post-stratification equating with large bandwidth under NEAT design		x		x		
NEC design	Equating with optimal bandwidth using gender as covariate under NEC design			x		x	
	Equating with large bandwidth using gender as covariate under NEC design				x	x	
	Equating with optimal bandwidth using computer/tablet use as covariate under NEC design			x			x
	Equating with large bandwidth using computer/tablet use as covariate under NEC design				x		x
	Equating with optimal bandwidth using gender and computer/tablet use as covariates under NEC design			x		x	x
	Equating with large bandwidth using gender and computer/tablet use as covariates under NEC design				x	x	x

2.3. Data Analysis

In this specific research, the performance of Kernel test equating and bandwidth selection was examined under two different test equating designs (i.e., NEAT and NEC designs). Reliability coefficients and descriptive statistics for test forms were calculated using SPSS software before the analysis for equating. The *kequate* package (Andersson et al., 2013) was used through the R program (R Core Team, 2013) to equate the two test forms using kernel equation methods. Equation methods were compared using standard equation errors (SEE).

3. RESULT

In this study, Kernel test equating methods were compared under NEAT and NEC designs. Kernel test equating methods used in the NEAT design are Kernel post-stratification equating with optimal bandwidth, Kernel post-stratification with large bandwidth, Kernel chain equating with optimal bandwidth, and Kernel chain equating with large bandwidth. Kernel equating methods used in the NEC design are Kernel equating with optimal bandwidth using gender as a covariate, Kernel equating with optimal bandwidth using computer/tablet use as a covariate, Kernel equating with large bandwidth using gender as a covariate, Kernel equating with large bandwidth using computer/tablet use as a covariate, Kernel equating with optimal bandwidth using both gender and computer/tablet use as covariates, and lastly Kernel equating with large bandwidth using both gender and computer/tablet use as covariates. In the present study, data obtained from two booklets (i.e., booklet 1 and booklet 14) of TIMSS 2019 8th grade science test were used. Under these conditions, which Kernel test equating method/s gave less incorrect results was examined by comparing the SEEs.

3.1. Preliminary Analysis Results

Table 2 shows the means, standard deviations, and reliability scores of two test forms. When these results are examined, it can be seen that these two groups have similar means and similar standard deviations. Furthermore, it is also seen that test forms used in this study have high-reliability coefficients.

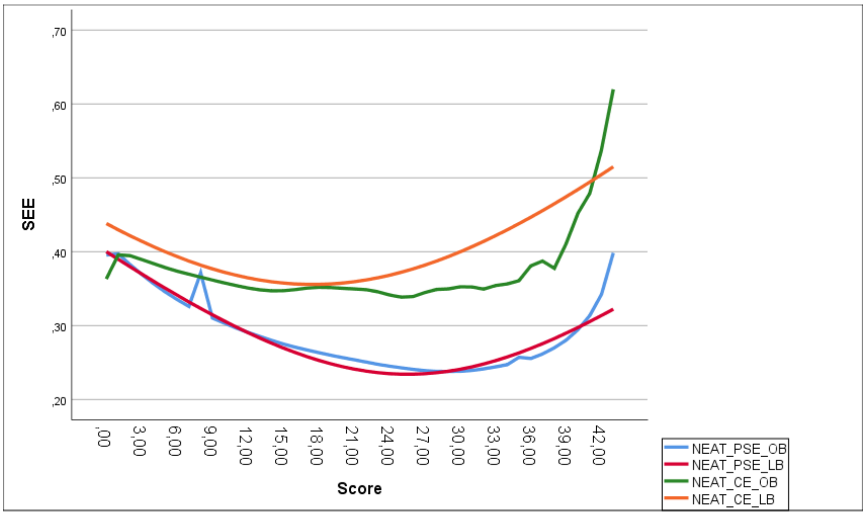
Table 2. *Descriptive statistics and reliability results of test forms.*

	Group 1 (n=284)		Group 2 (n=293)	
	Form A	Anchor items	Form B	Anchor items
Mean	24.25	10.04	22.06	9.48
St. Deviation	9.47	3.68	7.57	3.99
Skewness	-0.16	-0.28	-0.15	-0.23
Kurtosis	-1.00	-0.54	-0.84	-0.88
KR-20	.91	.76	.87	.84

3.2. Comparison of Kernel Test Equating Methods in the NEAT Design

When the standard error of the equating obtained as a result of the equating in the NEAT design is examined in Figure 4, it can be seen that the equating errors are similar for the low scores obtained from the tests. In general, regardless of bandwidth selection, Kernel post-stratification equation methods have been found to give lower error than that of chain equating methods. Kernel post-stratification equating methods show a similar distribution in terms of bandwidth. Although Kernel chain equating methods initially show a similar distribution, they differ in high scores. Specifically, Kernel chain equating with large bandwidth gives the highest SEEs for high scores. On the other hand, equalized scores obtained as a result of equating with the NEAT design are given in Table 5 in the appendices.

Figure 4. *Comparison of Kernel post-stratification and chain equating methods under NEAT design.*



Note. PSE_OB = Post-stratification method using optimal bandwidth, PSE_LB = Post-stratification method using large bandwidth, CE_OB = Chain equating method using optimal bandwidth, CE_LB = Chain equating method using large bandwidth.

3.3. Comparison of Kernel Test Equating Methods in The NEC Design

3.3.1. *Gender as a covariate*

Similar results were observed in the Kernel equating methods using optimal and large bandwidths and gender as a covariate under the NEC design (see Figure 5). As can be seen in Figure 5, Kernel equating method using large bandwidth gives higher SEEs at the scores at the bottom and top of the test. On the other hand, Kernel equating method using optimal bandwidth gives the lowest SEEs in the scores at the upper and lower parts of the scale. Both Kernel test equating methods have similar error values in the middle parts of the scale. Additionally, equalized scores obtained as a result of equating with gender variable as a covariate under the NEC design are given in Table 6 in the appendices.

Figure 5. *Comparison of Kernel using gender as a covariate under NEC design.*

Note. OB_G= Equating method using optimal bandwidth by gender covariate, LB_G= Equating method using large bandwidth by gender covariate.

3.3.2. *Computer/tablet use as a covariate*

Figure 6 shows that Kernel equating method using large bandwidth and computer/tablet use as a covariate gives higher SEEs in the top and bottom of the scale, while Kernel equating method using optimal bandwidth and computer/tablet use as a covariate gives lower SEEs in the top and bottom of the scale. Both Kernel test equating methods have similar SEEs in the middle part of the scale. On the other hand, equalized scores obtained as a result of equating with computer/tablet use variable as a covariate under the NEC design are given in Table 7 in the appendices.

Figure 6. *Comparison of Kernel using computer/tablet use as a covariate under NEC design.*

Note. OB_G= Equating method using optimal bandwidth by the use of computer/tablet covariate, LB_G= Equating method using large bandwidth by the use of computer/tablet covariate.

3.3.3. *Gender and computer/tablet use as covariates together*

Similar to the results related to previous variables, in the condition in which gender and computer/tablet are used as covariates together, Kernel equating method using large bandwidth gives higher SEEs in the scores at the bottom and top of the scale. On the other hand, Kernel equating using optimal bandwidth gives the lower error in the scores in the upper and lower parts of the scale. Both Kernel test equating methods have similar SEEs in the middle parts of the scale (see Figure 7). Besides these similar results, it can be seen that Kernel equating method under NEC design in which gender and computer/tablet variables are included together have higher SEE values than the SEEs obtained from Kernel test equating methods in which these variables were considered separately. Additionally, equalized scores obtained as a result of equating with gender and computer/tablet use variables as covariates under the NEC design are given in Table 8 in the appendices.

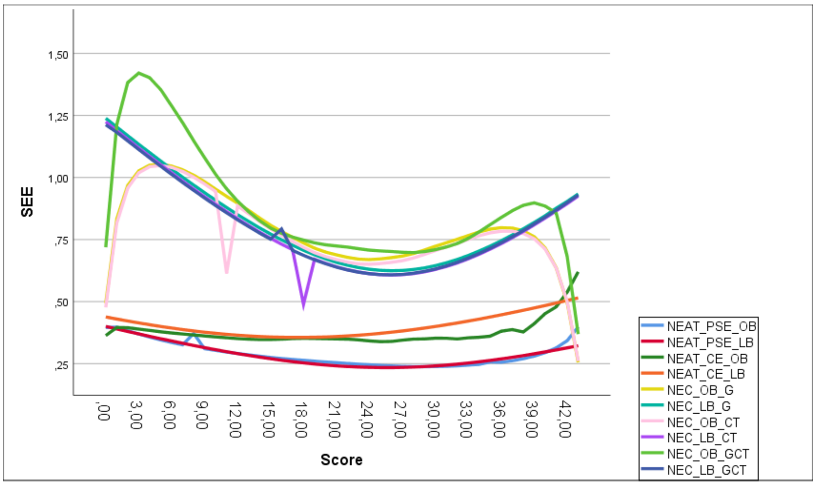
Figure 7. *Comparison of Kernel using gender and computer/tablet use as covariates together.*

Note. OB_G= Equating method using optimal bandwidth by gender and the use of computer/tablet covariates, LB_G= Equating method using large bandwidth by gender and the use of computer/tablet covariates.

3.4. Comparison of the Role of Bandwidth Selection in SEEs in Both NEAT and NEC Design

In the current study, optimal and large bandwidths were determined by kequate R package. The bandwidths for the Kernel post-stratification equating method are h(X) = 0.49 and h(Y) = 0.65. For the Kernel post-stratification equating method, the large bandwidths are h(X) = 12694.88 and h(Y) = 4190.42 (see Table 3). The results show that Kernel post-stratification methods using optimal and large bandwidth under the NEAT design demonstrate similar results. Furthermore, these methods demonstrate the lowest SEEs in the NEAT design (see Figure 8).

Figure 8. *Comparison of ten Kernel test equating methods in terms of bandwidth selection.*



Note. PSE_OB = Post-stratification method using optimal bandwidth, PSE_LB = Post-stratification method using large bandwidth, CE_OB = Chain equating method using optimal bandwidth, CE_LB = Chain equating method using large bandwidth.

In the Kernel chained equating, two linking functions, from X to A on P (group answering form X) and from A to Y on Q (group answering form Y) were used. Therefore, four distributions were to be continuized (von Davier et al., 2006). The optimal bandwidths are h(X) = 0.49, h(AP)= 0.43 and h(Y) = 0.68, h(AQ) = 0.43. The large bandwidths of the same equating method are h(X) = 12662.04, h(AP)= 3524.42 and h(Y) = 4183.10, h(AQ) = 5481.53 (see Table 3). Although Kernel chain equating methods with both optimal and large bandwidths in NEAT design gave higher error than that of Kernel post-stratification equating methods under NEAT design, they resulted in fewer errors than all Kernel equating methods under NEC design. Although Kernel chain equating methods using optimal and large bandwidths initially showed a similar distribution, they differed in equating high scores. Specifically, Kernel chain equating using large bandwidth gave the highest SEEs for high scores. Although Kernel equating methods using optimal bandwidth gave fewer SEEs in scores at the top and bottom of the scale, in general Kernel equating methods using large bandwidth demonstrated fewer SEEs (see Figure 8).

Table 3. *Bandwidths (h parameters) for NEAT design.*

	PSE-OB	PSE-LB	CE-OB	CE-LB
h_x	0.49	12694.88	0.49	12662.04
h_y	0.65	4190.42	0.68	4183.10
h_{aP}			0.43	3524.42
h_{aQ}			0.45	5481.53

Note. PSE-OB = Post-stratification method using optimal bandwidth, PSE-LB = Post-stratification method using large bandwidth, CE-OB = Chain equating method using optimal bandwidth, CE-LB = Chain equating method using large bandwidth.

In NEC design, the optimal bandwidths for the Kernel equating method in which gender was used as covariate are h(X) = 0.56 and h(Y) = 0.60. For the same equating method, the large bandwidths are h(X) = 9773.86 and h(Y) = 7563.13. The optimal bandwidths for the Kernel equating method in which computer tablet was used as covariate are h(X) = 0.56 and h(Y) = 0.60. For the same equating method, the large bandwidths are h(X) = 9474.77 and h(Y) = 7559.21. The optimal bandwidths for the Kernel equating method in which gender and computer/tablet use were applied as covariate are h(X) = 0.56 and h(Y) = 0.60. For the same equating method, the large bandwidths are h(X) = 9522.01 and h(Y) = 7633.00 (see Table 4). In the NEC design for all covariate options, although they gave higher SEEs for the scores at the top and bottom of the scale, the methods using large bandwidth gave less error regardless of covariate selection.

Table 4. *Bandwidth (h parameters) for NEC design.*

	G-OB	G-LB	CTU-OB	CTU-LB	G&CTU-OB	G&CTU-OB
h_x	0.56	9473.86	0.56	9474.77	0.56	9522.01
h_y	0.60	7563.13	0.60	7559.21	0.60	7633.00

Note. G-OB = Gender-optimal bandwidth, G-LB = Gender-large bandwidth, CTU-OB = Computer/tablet use-optimal bandwidth, CTU-LB = Computer/tablet use-large bandwidth, G&CTU-OB = Gender and computer/tablet use-optimal bandwidth, G&CTU-LB = Gender and computer/tablet use-large bandwidth.

In sum, NEAT design demonstrated lowest SEE values compared to those of NEC design, regardless of the covariate variable/s and bandwidth used. Kernel post-stratification equation methods showed a similar distribution in terms of bandwidth. Although Kernel chain equating methods initially showed a similar distribution, they differed in high scores. Specifically, Kernel chain equating with large bandwidth gave the highest SEEs for high scores. Kernel post-stratification equating methods resulted in less SEE than that of Kernel chain equating methods. NEC design was the design with the highest SEE values overall, regardless of the covariate variable/s and bandwidth used. When the methods based on NEC design are evaluated based on bandwidth in themselves, for all covariate options, the methods using large bandwidth gave less error. When the methods based on the NEC design were evaluated in terms of covariate selection, it was seen that the test equating methods in which gender and computer/tablet variables were handled separately resulted in less SEE than those in which these variables were considered together.

4. DISCUSSION and CONCLUSION

In this study, Kernel test equating methods were compared under NEAT and NEC designs. In NEAT design, taking into account optimal and large bandwidths, Kernel post-stratification and chain equating methods were compared. In the NEC design, gender and/or computer/tablet use was considered as a covariate, and by using these covariates Kernel test equating methods were performed. In these comparisons, bandwidths were considered as well.

In research that compares performance or errors of different methods, the main question asked is which methods should be preferred. In line with previous research (e.g., Akın Arıkan, 2019), the current study has shown that the Kernel post-stratification equating method provides fewer SEEs than those of the Kernel chain equating method in NEAT design. However, some studies show that the post-stratification method was more biased than the chain equating method (e.g., Livingston et al., 1990). It should be noted that these studies emphasize that the chain equating method may be preferable to post-stratification equating methods when the groups differ widely on the anchor test. In the current study, the reason why post-stratification methods show less error compared to chain equating may be that the two groups in this study have similar achievements.

The general finding of the current study is that Kernel equating methods in NEAT design resulted in fewer errors than those of Kernel equating methods in NEC design. The current study is consistent with previous studies (e.g., Akın Arıkan 2020; Wiberg & Branberg, 2015) which showed that NEAT design provides more accurate results in comparison to NEC design. Given that the performance of NEC design in equating depends on how well the covariates predict test scores and how well background variables explain differences in test scores (Wiberg & Branberg, 2015), one of the possible explanations for these results can be the selection of covariates. Gender and computer tablet use may not be eligible covariates for this group or this discipline (i.e., science). On the other hand, the second possible explanation may be sample size. In this study, Kernel equating was performed under the NEC design using a small sample; however, use of a small sample size can have caused the risk of having sparse data in some cells to be increased. Indeed, related studies that take into account the sample size in the NEC design (e.g., Branberg & Wiberg, 2011; Gonzales et al., 2015) show that the equating errors of the equating using the covariate under the NEC design are higher in the small sample. These explanations are also valid for results that show that conditions in which gender and computer/tablet use variables were handled together resulted in more SEEs in comparison to conditions in which these variables were handled separately in NEC design. As one adds more covariates, one obtains a rapid increase in the number of categories. Given that adding more covariates increases the risk of having sparse data in some cells (Wiberg & Branberg, 2015), it can be understood why conditions in which gender and computer/tablet use variables were handled together resulted in more SEEs in comparison to conditions in which these variables were handled separately in NEC design. Although the results of this study show that Kernel test equating methods in NEAT design give fewer SEEs, compared to SEEs of NEC design in which covariates (i.e., gender and computer/tablet use) were used, they still provide useful information to the literature on test equating. If even more suitable covariates for the distribution of the groups could be found and research could be replicated in large samples, equating performance could be closer to the results obtained with the NEAT design. At this point, more research is needed on NEC design.

Based on the results of the current study, five main conclusions can be drawn as follows:

1) In the NEAT design, Kernel chain equating methods (for both optimal and large bandwidths) exhibited higher error than the post-stratification equating methods did (for both optimal and large bandwidths).

2) The lowest error in the NEC design was obtained from the Kernel equating method with large bandwidth through the computer/tablet variable.

3) Kernel test equating results based on the NEC design, which considers gender and computer tablet use variables as a covariate separately, showed lower SEE than that of the NEC pattern, which takes these variables together as covariates.

4) In terms of bandwidth, when all methods are compared within the pattern used (i.e., NEAT and NEC), it has been seen that generally Kernel test equating results with large bandwidth result in fewer errors than the Kernel test equating results with optimal bandwidth.

5) When the NEAT and NEC designs are compared generally, the NEAT design has a lower SEE than that of the NEC design.

The results of the current study showed that in the NEAT design, Kernel post-stratification equating methods give fewer SEEs compared to those of Kernel chain equating methods. Given that relatively small samples were used in the present study, it can be recommended that in studies to be conducted on small samples, post-stratification equating methods be preferred to Kernel chain equating methods. The present study results also showed that the Kernel test equating results with large bandwidth result in fewer errors than the Kernel test equating results with optimal bandwidth. In test equating studies on small samples, large bandwidth can be preferred.

In the present study, Kernel test equating results based on the NEC design, which considers gender and computer tablet use variables as a covariate separately, showed lower SEE than that of the NEC design, which takes these variables together as covariates. In practice, using more than one covariate could be a reason for the inflation of SEEs because of increase in category number. Therefore, if covariates are going to be used in test equating studies that do not use anchor tests for equating, it should be noted that covariate number can be a reason for the increase in SEEs.

Lastly and most importantly, the results of the study showed that Kernel equating methods using anchor tests give fewer SEEs compared to those using covariate/s. In practice, it could be recommended that researchers or educators prefer to apply anchor tests instead of covariates. However, the present study did not address using anchor tests or students' demographic variables together as covariates. In further studies, those effects of usage of anchor tests as well as demographic variables as a covariate in NEC design can be examined.

Acknowledgments

The preliminary results of this study have been presented in 7th International Measurement and Evaluation in Education and Psychology Congress.

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 authors.

Authorship Contribution Statement

Seyma Nur Ozsoy: Conception, Design, Fundings, Materials, Data Collection and/or Processing, Analysis and/or Interpretation, Literature Review, Writing, Critical Review. **Sevilay Kilmen**: Supervision, Critical Review.

Orcid

Seyma Nur Ozsoy https://orcid.org/0000-0002-8306-6114
Sevilay Kilmen https://orcid.org/0000-0002-5432-7338

REFERENCES

Akın Arıkan, Ç. (2020). The impact of covariate variables on kernel equating under the non-equivalent groups. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme, 11*(4), 362-373.

Albano, A.D., & Wiberg, M. (2019). Linking with external covariates: examining accuracy by anchor type, test length, ability difference, and sample size. *Applied Psychological Measurement*, *43*(8), 597-610.

Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software*, *55*(6), 1-25.

Bränberg, K. (2010). *Observed score equating with covariates* [Unpublished Doctoral dissertation]. Department of Statistics, Umeå University.

Branberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement, 48*(4), 419-440.

Braun, H.I., & Holland, P.W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P.W. Holland & D.B. Rubin (Eds.) *Test equating* (9-49). Academic Press.

Choi, S.I. (2009). *A comparison of kernel equating and traditional equipercentile equating methods and the parametric bootstrap methods for estimating standard errors in* *equipercentile equating* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.

Godfrey, K.E. (2007). *A comparison of kernel equating and IRT true score equating methods* [Unpublished doctoral dissertation]. The Faculty of the Graduate School at the University of North Carolina at Greensboro.

Gonzales, J., Barrientos, A.F., & Quintana, F.A. (2015). Bayesian nonparametric estimation of test equating functions with covariates. *Computational Statistics and Data Analysis, 89*, 222-244.

Gonzales, J., & Wiberg, M. (2017). *Applying test equating methods*. Springer.

Kolen, M.J. (1988). Traditional equating methodology. *Educational measurement: Issues and Practice*, *7*(4), 29-37.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practises*. Springer

Kolen, M.J., & Brennan, R.L. (2004). *Test equating, scaling and linking: Methods and practises.* Springer

Liang, T., & von Davier, A.A. (2014). Cross-validation: An alternative bandwidth-selection method in Kernel equating. *Applied Psychological Measurement, 38*(4), 281-295.

Livingston, S.A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement, 30*(1), 23-39.

Livingston, S.A. (2014). Equating test scores (without IRT). *Educational testing service.*

Livingston, S.A., Dorans, N.J., & Wright, N.K. (1990). What combination of sampling equating methods works best?. *Applied Measurement in Education, 3*, 73–95.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

R Core Team. (2013). *R: A language and environment for statistical computing.* (Versiyon 4.0.3) [Computer software]. R Foundation for Statistical Computing.

von Davier, A.A. (2013). Observed-score equatıng: an overvıew. *Psychometrika, 78(*4), 605-623.

von Davier, A.A., Holland, P.W., Livingston, S.A., Casabianca, J., Grant, M.C., & Martin, K. (2006). An evaluation of the kernel equating method: A special study with pseudo tests constructed from real test data. *ETS Research Report Series*, *2006*(1), i-31.

von Davier, A.A., Holland, P.W., & Thayer, D.T. (2004a). The chain and post-stratification methods for observed-score equating: their relationship to population invariance. *Journal of Educational Measurement, 41*(1), 15-32.

von Davier, A.A., Holland, P.W., & Thayer, D.T. (2004b). *The kernel method of test equating*. Springer.

von Davier, A.A., & Chen, H. (2013). The Kernel levine equipercentile observed-score equating function. *ETS Research Report Series*, *2013*(2), i-27.

Wallin G., Häggström J., & Wiberg M. (2018) *How to select the bandwidth in kernel equating-An evaluation of five different methods*. In Wiberg M., Culpepper S.,Janssen R., González J., & Molenaar D. (Ed.), *Quantitative Psychology*. IMPS 2017. Springer Proceedings in Mathematics & Statistics, vol 233. Springer. https://doi.org/10.1007/978-3-319-77249-3_8

Wiberg, M. (2015). A note on equating test scores with covariates. In Ellinor Fackle- Fornius (Ed), *Festschrift in Honor of Hans Nyquist on The Occasion of His 65th Birthday.* Stockholm University.

Wiberg, M., & Branberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement, 39*(5), 349-361.

Wiberg, M., & von Davier, A.A. (2017). Examining the impact of covariates on anchor tests to ascertain quality over time in a college admissions test. *International Journal of Testing, 17*(2), 105-126.

Yurtçu, M. (2018). *The comparison of test equating with covariates using Bayesian nonparametric method* [Unpublished master thesis]. Hacettepe University.

APPENDICES

Table 5. *Equalized scores obtained in the NEAT design.*

	NEAT_PSE_EQ	NEAT_PSE_L	NEAT_CE_EQ	NEAT_CE_L
1	2.26	4.55	1.45	3.44
2	3.92	4.88	2.71	3.65
3	4.89	5.21	3.43	3.86
4	5.60	5.54	3.95	4.08
5	6.19	5.87	4.38	4.29
6	6.70	6.20	4.74	4.50
7	7.15	6.53	5.07	4.71
8	7.57	6.86	5.37	4.93
9	7.94	7.19	5.64	5.13
10	8.28	7.52	5.89	5.35
11	8.60	7.85	6.12	5.56
12	8.91	8.18	6.34	5.78
13	9.21	8.51	6.53	5.99
14	9.50	8.84	6.71	6.20
15	9.78	9.17	6.88	6.41
16	10.04	9.50	7.05	6.63
17	10.29	9.83	7.21	6.84
18	10.51	10.16	7.38	7.05
19	10.74	10.49	7.54	7.26
20	10.97	10.82	7.69	7.48
21	11.20	11.15	7.85	7.69
22	11.43	11.48	8.01	7.90
23	11.68	11.81	8.17	8.11
24	11.93	12.14	8.34	8.32
25	12.19	12.47	8.50	8.54
26	12.45	12.80	8.66	8.75
27	12.71	13.13	8.83	8.96
28	12.97	13.46	9.01	9.17
29	13.23	13.79	9.19	9.38
30	13.51	14.12	9.38	9.60
31	13.80	14.45	9.57	9.81
32	14.10	14.78	9.77	10.02
33	14.41	15.11	9.97	10.24
34	14.72	15.44	10.17	10.45
35	15.04	15.77	10.38	10.66
36	15.35	16.10	10.61	10.87
37	15.69	16.43	10.85	11.09
38	16.06	16.76	11.13	11.30
39	16.47	17.09	11.42	11.51
40	16.95	17.42	11.75	11.72
41	17.53	17.75	12.24	11.94
42	18.26	18.08	12.83	12.15
43	19.27	18.41	13.81	12.36
44	21.09	18.74	15.93	12.57

Table 6. *Equalized scores obtained in the NEC design by the use computer/tablet covariate.*

	NEC_EQ_GK	NEC_L_GK	NEC_EQ_LK	NEC_EQ_UK
1	0.73	2.73	0.83	0.68
2	2.35	3.53	2.36	2.38
3	3.66	4.33	3.66	3.68
4	4.79	5.13	4.79	4.79
5	5.81	5.93	5.81	5.81
6	6.75	6.73	6.75	6.76
7	7.65	7.53	7.65	7.66
8	8.52	8.32	8.52	8.54
9	9.36	9.12	9.36	9.37
10	10.18	9.92	10.18	10.19
11	10.99	10.72	10.99	10.99
12	11.79	11.52	11.79	11.79
13	12.58	12.32	12.58	12.59
14	13.37	13.11	13.37	13.37
15	14.15	13.91	14.14	14.14
16	14.92	14.71	14.92	14.92
17	15.69	15.51	15.69	15.69
18	16.46	16.31	16.46	16.47
19	17.23	17.11	17.23	17.23
20	17.99	17.90	17.99	17.99
21	18.76	18.70	18.76	18.76
22	19.52	19.50	19.52	19.53
23	20.29	20.30	20.29	20.29
24	21.05	21.10	21.05	21.05
25	21.82	21.90	21.82	21.82
26	22.58	22.69	22.58	22.58
27	23.35	23.49	23.35	23.35
28	24.12	24.29	24.12	24.12
29	24.90	25.09	24.90	24.90
30	25.67	25.89	25.67	25.67
31	26.45	26.69	26.45	26.45
32	27.24	27.48	27.24	27.24
33	28.03	28.28	28.03	28.04
34	28.84	29.08	28.84	28.84
35	29.65	29.88	29.65	29.64
36	30.47	30.68	30.48	30.46
37	31.32	31.48	31.32	31.31
38	32.18	32.28	32.18	32.18
39	33.08	33.07	33.08	33.08
40	34.01	33.87	34.02	34.01
41	35.00	34.67	35.01	35.00
42	36.07	35.47	36.07	36.07
43	37.27	36.27	37.26	37.26
44	38.62	37.07	38.57	38.65

Table 7. *Equalized scores obtained in the NEC design by gender covariate.*

	NEC_EQ_GK	NEC_L_GK	NEC_EQ_LK	NEC_EQ_UK
1	0.78	2.79	0.88	0.73
2	2.43	3.59	2.43	2.46
3	3.75	4.39	3.74	3.76
4	4.87	5.19	4.87	4.88
5	5.89	5.99	5.89	5.89
6	6.84	6.78	6.83	6.84
7	7.73	7.58	7.73	7.74
8	8.60	8.38	8.59	8.61
9	9.43	9.18	9.43	9.45
10	10.25	9.98	10.25	10.26
11	11.06	10.78	11.06	11.06
12	11.86	11.57	11.86	11.86
13	12.64	12.37	12.64	12.65
14	13.42	13.17	13.42	13.44
15	14.20	13.97	14.20	14.20
16	14.97	14.78	14.97	14.97
17	15.74	15.57	15.74	15.74
18	16.51	16.36	16.51	16.52
19	17.28	17.16	17.28	17.28
20	18.04	17.96	18.04	18.04
21	18.80	18.76	18.80	18.80
22	19.57	19.56	19.57	19.57
23	20.33	20.36	20.33	20.33
24	21.10	21.15	21.10	21.10
25	21.86	21.95	21.86	21.86
26	22.63	22.75	22.63	22.63
27	23.40	23.55	23.40	23.40
28	24.17	24.35	24.17	24.17
29	24.95	25.15	24.95	24.95
30	25.73	25.94	25.73	25.72
31	26.51	26.74	26.51	26.50
32	27.30	27.54	27.30	27.29
33	28.09	28.34	28.09	28.09
34	28.90	29.14	28.90	28.90
35	29.71	29.94	29.71	29.71
36	30.54	30.73	30.54	30.53
37	31.39	31.53	31.39	31.38
38	32.26	32.33	32.26	32.25
39	33.15	33.13	33.15	33.15
40	34.09	33.93	34.09	34.09
41	35.08	34.73	35.08	35.08
42	36.14	35.52	36.14	36.14
43	37.32	36.32	37.32	37.31
44	38.65	37.12	38.60	38.68

Table 8. *Equalized scores obtained in the NEC design by gender and the use computer/tablet covariate.*

	NEC_EQ_GK	NEC_L_GK	NEC_EQ_LK	NEC_EQ_UK
1	0.66	2.85	0.75	0.62
2	2.23	3.65	2.24	2.24
3	3.54	4.44	3.54	3.56
4	4.67	5.24	4.67	4.68
5	5.70	5.79	5.70	5.71
6	6.67	6.84	6.67	6.68
7	7.59	7.63	7.59	7.61
8	8.48	8.43	8.48	8.50
9	9.34	9.23	9.34	9.35
10	10.19	10.03	10.19	10.19
11	11.02	10.83	11.02	11.02
12	11.84	11.62	11.84	11.84
13	12.65	12.42	12.65	12.66
14	13.45	13.22	13.45	13.46
15	14.24	14.02	14.24	14.25
16	15.03	14.82	15.03	15.03
17	15.82	15.61	15.82	15.82
18	16.60	16.41	16.60	16.60
19	17.38	17.21	17.37	17.38
20	18.15	18.01	18.15	18.15
21	18.92	18.80	18.92	18.92
22	19.69	19.60	19.69	19.69
23	20.45	20.40	20.45	20.46
24	21.22	21.20	21.22	21.22
25	21.98	22.00	21.98	21.98
26	22.75	22.79	22.75	22.75
27	23.51	23.59	23.51	23.51
28	24.27	24.39	24.27	24.27
29	25.04	25.19	25.04	25.04
30	25.80	25.98	25.80	25.80
31	26.57	26.78	26.57	26.57
32	27.35	27.58	27.35	27.34
33	28.13	28.38	28.13	28.13
34	28.91	29.18	28.91	28.91
35	29.71	29.97	29.71	29.70
36	30.52	30.77	30.52	30.50
37	31.34	31.57	31.34	31.34
38	32.19	32.37	32.19	32.19
39	33.07	33.17	33.07	33.07
40	33.99	33.96	33.99	33.99
41	34.97	34.76	34.97	34.97
42	36.03	35.56	36.03	36.03
43	37.22	36.36	37.22	37.21
44	38.60	37.15	38.54	38.62