Education and Science tedmem

Vol 45 (2020) No 202 113-139

The Effect of Item Exposure Control Methods on Measurement Precision and Test Security under Different Measurement Conditions in Computerized Adaptive Testing *

Recep Gür [1], H. Deniz Gülleroğlu [2]

Abstract

In this research, it was aimed to investigate the effect of item exposure control methods on measurement precision and test security for dichotomous items within different conditions (different ability distributions, test lengths, sample sizes) in computerized adaptive testing applications. Current research is designed as a Monte Carlo simulation research. Within the scope of the research, short test was designed to have 25 items while long test was designed to have 50 items. In addition to this, small sample was simulated to have 250 observations whereas large sample was simulated to have 1000 observations. Regarding the test-takers' ability parameters, computerized adaptive testing samples which have normal, right-skewed, left-skewed and uniform distribution for each sample size condition in which ability parameters were within -3 and +3 range were generated. The condition in which item exposure was not controlled for was taken as the reference condition and Sympson Hetter and Fade Away methods were compared with the reference condition. When the variables and conditions within the research were crossed, 48 simulation conditions were established and 4800 data files were generated by doing 100 replications for each condition. Measurement precision and test security indices computed for each condition were compared to one another. The analyses conducted indicate that measurement precision indices corresponding to the different item exposure control conditions did not create substantial differences. It was found out that when Fade Away method was used, item exposure was distributed in a balanced way and higher test security was obtained. As a consequence of this, it was concluded that item pool became more effective and sustainable without revelation (disclosure) of the items. Accordingly, it was discovered that in computerized adaptive testing applications in different conditions, instead of Sympson Hetter method which is frequently used before computerized adaptive testing applications and in

Keywords

Computerized Adaptive Test
Item Exposure Control Method
Sympson Hetter Method
Fade Away Method

Article Info

Received: 10.31.2018
Accepted: 11.20.2019
Online Published: 04.04.2020

DOI: 10.15390/EB.2020.8256

* This article is derived from Recep Gür's PhD dissertation entitled "The effect of item exposure control methods on measurement precision and test security within different conditions in adaptive testing", conducted under the supervision of H. Deniz Gülleroğlu.

[1] Anadolu University, Faculty of Education, Dept. of Educational Sciences, Turkey, math.recepgur@gmail.com
[2] Ankara University, Faculty of Educational Sciences, Dept. of Educational Sciences, Turkey, denizgulleroglu@yahoo.com

which item exposure control parameters are estimated by iterative simulations, the use of Fade Away method which ensures the item exposure control parameters to be estimated simultaneously resulted in higher test security without reducing the measurement precision.

Introduction

The results obtained from the data collection tools used to measure the psychological characteristics and behaviors of individuals have an important place in individuals' lives. In order to the measurement results to be able to used in important decisions related to education and professional areas, the scores obtained from the measurement tools have to be valid and reliable. Since many implicit features in education and psychology cannot be observed directly, two measurement theories have been developed by the researchers to ensure the measurement of the relevant variables to be valid and reliable. These are Classical Test Theory (CTT) and Item Response Theory (IRT). In the test development phase, these theories are used while estimating the item and test statistics (Lord, 1980).

One of the advantages of IRT in practical application is that it allows the development of adaptive tests. The Binet intelligence test which was developed by Binet and Simon in 1905 is considered to be the first adaptive test application along with showing typical characteristics of the adaptive test (Weiss, 1988). Computerized Adaptive Test (CAT) applications, by using the immutability feature of the IRT, works with an algorithm that enables the selection of items that can be qualified for each individual to be selected from the item pool and the presentation of it to the individual (Embretson & Reise, 2000; Way, 2005).

In CAT applications, items are selected to match the individual's ability (θ: theta) levels. This can be processed differently with various methods. Generally, in the first step, the individual is expected to respond to the moderately difficult item which is selected from the items in the item pool. After the estimation of θ level of the individual is obtained according to the response given, which item in the item pool will provide the maximum information for the θ estimation of the individual is decided. The basic logic of the item selection in CAT applications is to make the individual encounter more difficult items after the item they answer correctly and easier items after the item they answered incorrectly. After the individuals' response to each item, the θ levels are recalculated and a new θ estimation is made. The test ends when different termination rules are performed (Bulut & Kan, 2012; Lord & Stocking, 1988).

The scores obtained from CAT applications can lead important decisions about individuals. Therefore, implementing CAT applications in a valid and a reliable way affects the accuracy of the decisions taken. With the developments in the field of computer, software and psychometry, since CAT applications become widespread, test security and measurement precision among the factors that affect the validity and security of CAT applications become a current issue (Boyd, Dodd, & Fitzpatrick, 2013; Weiss, 2004).

In CAT applications as to the situations that affect the test security and the measurement precision negatively, it is possible to restrict the item exposure rate with CAT algorithm (Han, 2009). The control of the item exposure is emphasized to be one of the main components of CAT, as well as the calibration of the item pool, starting and termination rules, item selection and theta estimation methods, content balancing (Boyd, 2003; Segall, 2004; Magis & Raîche, 2012).

In CAT applications, individuals can be allowed to take the test as many and often as they wish, especially in large-scale tests, since the individuals encounter the most appropriate items for their theta levels. In such applications, when an individual who takes the exam more than once is over exposed to the same items very often, the disclosure of the items leads to an artificial decrease in the psychometric properties of items (Segall, 2004; Revuelta & Ponsoda, 1998). This situation has a negative effect on the validity of test results and test security (Lee & Dodd, 2012).

There is a negative relationship between item information function and the standard error of the theta estimation. As the item information function is taken as the criteria for item selection in CAT applications, the selection of items with high information values increases the measurement precision by decreasing the standard error of theta estimation (Kalender, 2009). However, when the items that maximize the measurement precision are selected, some items in the item pool are applied very often, while some items are not applied to any individuals because the item exposure will not be distributed evenly (Pastor, Dodd, & Chang, 2002). This situation causes an unbalanced use of the item pool and leads to the use of only a limited number of some items (Han, 2009). For this reason, it is aimed to ensure the test security by using the item pool in a more balanced without decreasing the measurement precision by means of the item exposure control methods (Boyd, 2003; Davis & Dodd, 2005; Pastor et al., 2002).

Taking individual differences into consideration, there is a need for a large item pool for CAT applications carried out by selecting the most appropriate items for the individual's theta level (Embretson & Reise, 2000; Magis & Raîche, 2012). CAT applications with the item pools which have a high degree of distinctiveness and which consist of the items with difficulty levels that can address each skill level give better results (Veldkamp & Van Der Linden, 2010; Weiss, 2004). Therefore, the use of only a limited number of items, which does not show a balanced distribution of the item pool, causes the time and the labor force which are spent not to be evaluated well (Aytuğ Koşan, 2013). For these reasons, item exposure control methods have been developed to ensure the security of the test, to make the use of the item pool more fertile and to ensure the functional continuity of the item pool (Boyd, 2003; Davis, 2002; Revuelta & Ponsoda, 1998).

One of the first methods developed for the item exposure control problem is the strategy of 5-4-3-2-1. The 5-4-3-2-1 strategy of the random selection strategies aims to exposure control of application of a randomly selected item as being selected among 5 items in the first step and 4 items in the second step in CAT applications. Kingsbury and Zara (1989) and Thomasson (1998) developed different random selection methods to reduce the item exposure of all items (as cited in Veldkamp, Vershoor, & Eggen, 2010). The Randomesque method, in which one of the ten predetermined items that provide the majority of information is randomly selected; Within .10 logit method (Lunz & Stahl, 1998) in which random items are selected from the .10 logit range of the targeted item difficulty level and the Progressive method which aims to significantly increase the importance of item information by decreasing the effect of an accidental component on item selection as the test proceeds are among the random selection methods. The common purpose in random selection methods is to exposure control of application of a randomly selected item among the items in the range in which the maximum information is most ideal (Georgiadou, Triantafillou, & Economide, 2007).

In conditional selection methods, the item exposure rate is controlled by using the item exposure control parameter determined by repetitive simulations before applying CAT There are conditional selection methods such as the Davey and Parshall method, the Stocking and Lewis multi-nominal method, the targeted exposure control method and the Restricted Maximum Information-[RMI] method. While the Davey and Parshall method (1995) does not only restrict the excessive use of individual items but also prevents individuals from encountering the same sets of items; in the Stocking and Lewis multi-nominal method, the item exposure control parameter for each item as many as the number of each theta level (Stocking & Lewis, 1995). The targeted item exposure control method is intended to increase the possibilities of using unused items rather than focusing on controlling the overuse of the items (Thompson, 2002). In the RMI method which is one of the conditional selection methods, an item is not permitted to be used more than the predetermined item exposure rate (Revuelta & Ponsoda, 1998).

In the rotating pool method (Ariel, Veldkamp, & Van Der Linden, 2004; Way, 1998), the items are used in different tests using a priori distribution on the condition that the item pool intended to reduce the item exposure shows a similar distribution in terms of content and item parameters. In addition to these methods, the stratification methods which aims to use the item pool by stratifying it have been developed Among these, are there a- stratified strategy (Chang & Ying, 1999), a-stratification in which b is blocked (Chang, Qian, & Ying, 2001), the CAT design with content blocked and a-stratified (Yi & Chang, 2001), 0-1 stratified strategy (Chang & Van Der Linden, 2003) the methods combining different methods Forward-delimited Strategy (Revuelta & Ponsoda, 1998); the combination of a stratification and Sympson Hetter (SH) strategy (Leung, Chang & Hau, 2003), a combination of content blockade (Yi, 2002) a stratification and SH strategy(as cited in Georgiadou et al., 2007). Although there are different methods, the most commonly used item exposure control method in CAT applications is the Sympson Hetter (SH) method which is one of the conditional selection methods (Veldkamp et al., 2010).

When the maximum item exposure rate is determined as the target value before CAT application and the item is selected, the fact that an individual encounters that item depends on the item exposure control parameter (Davis & Dodd, 2005). For example, if this parameter is defined as .25 for all items, when an item is selected, this relevant item can be applied in each of approximately four computerized adaptive tests (Weiss & Guyer, 2012).

The fact that the item exposure rate of ith. item (P_i (S)) which is selected the most in CAT applications is bigger than the item exposure control parameter (P_i (A|S)) which is determined before the CAT application that is (P_i (A|S))<(P_i(S)) restricts the probability of application P_i (A)) when the i. item is selected. On the other hand, the item exposure rate for the less used items increases the probability of the applications of the relevant items when it is smaller than the item exposure control parameter (P_i (S)< P_i (A|S)) (Segall, 2004; Stocking & Lewis, 2002). Consequently, that the conditional probability of the item to be used P_i(A|S) is defined and used in P_i(A) = P_i(A|S) $*$ P_i(S) equality, allows achieving the targeted exposure rate, controlling whether the selected items are applied or not using the items selected for the individuals participating in CAT application temporarily (Pastor et al., 2002; Veldkampet al., 2010).

Han (2009) emphasizes that there is no need for repeated simulation to be used before CAT application to determine the item exposure control parameters and that item exposure control parameters can be determined simultaneously in CAT implementation. In the item selection process, the item selection criteria for each item which is appropriate to be selected from the pool is weighted reversely ($I_i\ [\hat{\theta}_{m\text{-}1}]\ \frac{e-r_i}{e}$) with the rate between the target exposure rate (e) and the current observed exposure rate (ri). So if e> r_i, the Maximum Fisher Information (MFI) which is an item selection criteria (MFI) $I_i\ [\hat{\theta}_{m\text{-}1}]$ is positive while if r_i>e, $I_i\ [\hat{\theta}_{m\text{-}1}]$ becomes negative (Han, 2012). This method is called as Fade-Away Method by Han since the selection rate of the more used items will be decreased gradually as less used items will be used more often in this case.

In studies conducted on CAT applications, it has been examined to see whether or not there is a significant relationship between the ability levels that are obtained via CAT applications done with various strategies (different starting and termination rules, item selection and abilitiy estimation methods) and the traditional paper-pencil test (Bulut & Kan, 2012; Cömert, 2008; Eroğlu, 2013; Gökçe, 2012; İşeri, 2002; Kalender, 2011; Kaptan, 1993; Kaskatı, 2011; Kezer, 2013; McDonald, 2002; Özbaşı, 2014; Öztuna, 2008; Scullard, 2007; Smits, Cuijper, & Straten, 2011; Sulak, 2013; Wang, Kuo, Tsai, & Laio, 2012; Zitny, Halama, Jelinek, & Kveton, 2012). In the related literature, as a result of the CAT application, it has been concluded that theta estimations similar to paper-pencil tests are done and that similar results can be obtained when abilities are estimated by using various CAT strategies. In addition, it has been concluded that CAT applications increased the measurement precision compared to paper pencil tests and provided significant economic efficiency in terms of both number of items and required time for theta estimation.

When the studies carried out on the item exposure control method which is one of the bases of CAT are examined, it is seen that the item exposure control methods are compared and contrasted according to polytomous IRT models (Burt, Kim, Davis, & Dodd, 2003; Davis, 2002, 2004; Davis & Dodd, 2005), different item selection methods (Boztunç Öztürk, 2014; Han, 2009, 2012), item pools that have different distributions of item discrimination (Revuelta & Ponsoda, 1998), item pools that have various difficulty levels (Boztunç Öztürk, 2014; Lee & Dodd, 2012), different sizes of item pools (Chang & Twu, 1998; Pastor et al., 2002; Revuelta & Ponsoda, 1998); different test termination rules (French & Thompson, 2003; Revuelta & Ponsoda, 1998) and different theta estimations in testlet- based CAT applications (Boyd, 2003; Davis & Dodd, 2003). In the relevant studies, it has been concluded that item exposure methods are affected by item pool sizes, that the number of the unused items in the item pools consisting of medium difficulty items is less and their measurement precision is better and that in the samples which have a normal distribution as difficulty level of item pool increases the number of the unused items increases too. It has also been concluded that Sympson Hetter strategy among the conditional item selection methods in the dichotomous items is more effective compared to the random selection methods and that test security increases when a-stratifed method among the stratification methods and Fade Away item exposure control method are applied together.

Rudman (1987) argues that CAT applications can be considered as the measurement method of the 21st century and draws attention to the importance of the studies to be conducted on CAT strategies. From a psychometric point of view, CAT applications have two important advantages. One of them is to increase the measurement precision and the other is to provide a safe test environment (Weiss, 2004).

In CAT applications, individuals can be allowed to take the test as often and many as they want especially in large-scale tests since individuals encounter the most appropriate items for their theta levels. In this way, although the continuity of CAT applications has advantages in terms of planning the test schedule, providing adequate space and computers in schools and test centers for test applications, it has also disadvantages. These practices raise the risk that an individual taking part in an exam more than once can remember the items that he has previously encountered, and that he shares the items that are exposure with his friends via the internet on social media. They also bring some problems together with them such as matching the individuals' relevant theta levels and the items with a high information value degree in the computer environment, using only a limited number of items by not using the item pool in a balanced way, decreasing the psychometric features of the item exposure and losing the functionality of the item pool in CAT applications. Such factors lead to discussions about test security and measurement precision among the factors that affect the validity and security of CAT applications. In this framework, item exposure control methods take its place among the basic elements of CAT.

In CAT applications, the fact that an individual who takes an exam more than once will do practice in case that he encounters the same items, according to Yen (1993), is among the factors that lead to local item dependence. In CAT applications based on IRT, the violation of IRT's local item independency assumption overchanges the security, affects standard errors in estimating theta and item parameters and therefore leads to incorrect estimations of theta and item parameters (Demars, 2006). The studies which are going to be done on the item exposure control methods are thought to find a solution to these problems.

The test length, the sample size and the ability distribution have an important effect on the application and interpretation of IRT models. To identify which item exposure control method is more functional in various testing environments (for example; small sample, short test and left skewed ability distribution) contributes to both the increase in test security and theta estimations of the individuals with less errors. In this context, to examine how indices such as the measurement precision and the test security change when the item exposure control methods are compared and contrasted according to different sample sizes, test lengths and ability distributions is the problem of this research.

Method

This chapter contains information about the study model, the data generation, the CAT conditions and the data analysis.

Research Model

In studies, where empirical data is difficult to find, simulation researches are studied for understanding the relations between methods. These simulation researches make important contributions in the development of the theory (Davis, Eisenhardt, & Bingham, 2007). This study can be considered as a pure research since it aims to examine how the measurement precision and the test security indices change when the item exposure control methods are compared according to different testing measurement conditions in dichotomous items scaled as 1-0 in Monte-Carlo simulation CAT applications.

Data Generation

In the CAT application, in addition to the large item pool consisting of items scaled according to the IRT, a large number of participants' response patterns are needed in order to obtain the least errors in estimations. In addition, since it was practically impossible to reach the datasets that could provide different measurement conditions for the purpose of the study, simulative data were used. Monte- Carlo simulation studies offer an effective and fast comparison of CAT applications by diversifying data sets in different strategies (Weiss & Guyer, 2012). SimulCAT (Han, 2011) software was used to produce the data.

Test length or sample size has a significant effect in the application and interpretation of IRT models. According to Şahin and Anıl (2017), while interpreting one dimensional dichotomous (1-0) IRT models, examining the test length and the sample size together is said to have a significant effect in terms of item parameter estimations. Although there are differences in the distribution of test length and sample size in the literature, there are usually 25 items for a short test (Demars, 2006; Guyer & Thompson, 2011; Harwell, Stone, Hsu, & Kirişçi, 1996; Weiss & Von Minden, 2012; Yoes, 1995), 50 items for a long test (Demars, 2006; Glas, 2002; Weiss & Von Minden, 2012); 250 individuals in a small sample (Çetin, 2009; Goldman & Raju, 1986; Harwell & Janosky, 1991; Speron, 2009; Vaughn & Wang, 2010; Yoes, 1995), and 1000 individuals in a large sample (Çetin, 2009; Glas, 2002; Goldman & Raju, 1986; Guyer & Thompson, 2012; Hulin, Lissak, & Drasgow, 1982; Köse, 2010; Lord, 1968; Patsula & Gessaroli, 1995; Speron, 2009; Tang, Way, & Carey, 1993; Thissen & Wainer, 1982; Vaughn & Wang, 2010; Yen, 1987; Yoes, 1995; Weiss & Von Minden, 2012). In line with the information in the relevant literature, a short test consisted of 25 items, a long test consisted of 50 items; the small sample was determined as 250 individuals and the large sample was determined as 1000 individuals.

CAT samples were generated by producing theta parameters for the individuals who take the test with (θ) value between -3 and +3 having; i) Normal distribution θ~N(0, 1), according to Weiss and Guyer (2012), ii) Right skewed distribution β (1, 4) iii) Left skewed distribution β (4, 1), iv) Uniform distribution β (1, 1) for each sample size condition. In addition to this, according to Agresti and Coull (1998), for mean and standard deviation values in the uniform distribution, we get CAT samples as $\bar{x} \cong .00$, $SS = 6/\sqrt{12} \cong 1.732$.

In the SH strategy that is one of the item exposure control methods, the item exposure control parameters should be determined by repeated simulations before the CAT application (Veldkampet al., 2010). For this purpose, the sample size of the Sympson Hetter sample was taken as 6000 and the replication number was taken as 5, and the item exposure control parameters were calculated with the distribution of the individuals in each CAT sample (Gu & Reckase, 2007). If the mean selection rate (P_i (S)) of each item after five replications is greater than the maximum target exposure rate (r=.20), the item exposure control parameter is calculated as $k_i=r/P_i(S)$ and if (P_i(S))<r, the application probability of the relevant items is increased by considering $k_i=1$ (Eggen, 2001; Veldkamp & Van Der Linden, 2008).

In CAT applications, it provides better results when using with an item pool which consists of high discrimination items and has a sufficient number of items suitable for each theta level. In line with this information, according to Weiss and Von Minden (2012) a simulative item pool with 1000 items, which is within the ranged as a (.25 ile 1.75), b (-3 ile +3), c (.20 ile .30), whose a and c parameters show uniform distribution and whose b parameter shows a normal distribution, has been generated. For the multiple choice one dimensional dichotomous items, the estimations were made according to the 3PL model since it covers the c parameter in it (Crocker & Algina, 1986; Hambleton, Swaminathan, & Rogers, 1991). The response function figure for the item pool is given below.

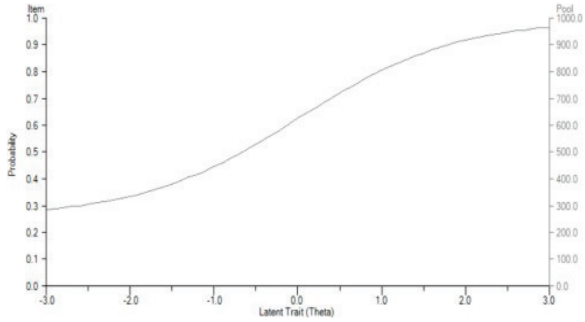

Figure 1. The Response Function for the Item Pool

CAT Conditions

Within the study, Maximum Fisher Information method which is preferred the most among item selection methods and -.50<b<.50 range strategy among starting strategies were used for the simulative CAT application (Weiss, 1988). Since there may be individuals who have all answers correct or all answers wrong, Expected A Posteriori (EAP) method that makes estimations by using the priori distribution instead of the maximum likelihood method for estimation as a theta estimation method (Embretson & Reise, 2000).

Within the study, SH Strategy which is used in CAT applications the most among the item exposure control methods (Veldkamp et al., 2010) and Fade Away Strategy (Han, 2009) which do not require repeated simulations before the CAT application and in which the item exposure control parameters are determined simultaneously during the CAT application. A fixed test length (short test: 25 items; long test: 50 items) is defined as the termination rule. Thus, 2*4*3*2=48 simulation conditions was determined when the relevant variables and conditions were crossed. In addition, 4800 data files were created by using 100 replications (Han, 2009) in each condition in order to avoid sample bias (Harwell 1996, as cited in Evans, 2010).

Data Analysis

Within the frame of the study, measurement precision and test security indices were calculated and compared to each other.

Data Analysis Methods for Measurement Precision in Different Measurement Conditions

Since as the measurement precision increases error values will decrease, in order to determine the measurement precision in the study, fidelity, bias and Root Mean Squared Error (RMSE) coefficients were calculated in each case (French & Thompson, 2003).

The coefficient of fidelity is the correlation coefficient between the estimated theta parameters and the true theta parameters with the CAT application. Pearson Correlation Coefficient was used to calculate the coefficient of fidelity. RMSE values for the absolute difference and bias for the mean difference were calculated between the estimated theta level and true theta level in the CAT application (Gu & Reckase, 2007; Leroux, Lopez, Hembry, & Dodd, 2013; Wang & Vispoel, 1998;

Zheng & Chang, 2014). Whether there is a significant difference between the fidelity values of the measurement precision in different measurement conditions was examined by using Fisher's z test which allows to compare the difference between the two correlation coefficients (Howell, 2010; Şencan, 2005). The Cohen q coefficient (Cohen q = $r'_1 - r'_2$) was used to calculate the effect size for the difference between two correlation coefficients. While the Cohen q coefficient that is less than .10 is interpreted as it has no effect, it is interpreted as a small effect size it is between .10 and .30, a moderate size between .30 and .50, and a great size when it more than .50 (Cohen, 1988).

Mahalanobis distances can be used to compare patterns with known results and to test extreme observations, since it allows grouping of variations (the bias values of the item exposure methods under different measurement conditions) and calculating the distance between the clusters (Hair, Anderson, Tatham & Black, 1998; Pallant, 2010). Therefore, Mahalanobis distances were calculated in order to examine whether there was a significant difference between the bias / RMSE values of measurement precision in different measurement conditions.

RMSE is an absolute fit index developed according to the degree of errors between the covariance matrix for the true theta levels and the covariance matrix for the estimated theta levels (Sümer, 2000). RMSE shows how concordant are the theta parameters estimated from the model and the universe covariance (true theta parameters) and it is valued between .00 and 1.00 (Byrne, 1998). As the difference between the covariances of the true theta parameters and covariances of the estimated theta parameters from the model is close to zero, it can be said that the model is fit (Byrne, 1998; Kline, 2005). That RMSE value is smaller than .05 is considered as the perfect fit (Brown, 2006; Jöreskog & Sörbom, 1993). When it is between .05 and .08, it is considered as the acceptable fit (Hooper, Coughlan, & Mullen, 2008; Howell, 2010; Tabachnick & Fidell, 2007) and finally that it is smaller than .10 is thought as the weak fit (Kelloway, 1998). If the sample is small and the number of the parameters calculated in the model is high, RMSE value can be higher than .10 (Şimşek, 2007).

Data Analysis Methods for Test Security in Different Measurement Conditions

In each condition, test security can be speculated by evaluating item exposure rate, pool utilization and test overlap indices together (Davis & Dodd, 2005). The percentage of the items that were not used during CAT application is meant by saying 'pool utilization' (Leroux et al., 2013). *The item exposure rates in each condition* are determined by calculating the distribution of the item exposure rates ($\boldsymbol{r_i}$: the ratio of the number of item used to the sample size), the standard deviation of item exposure and the maximum item exposure rate. In addition, *the pool utilization index* obtained by calculating the number and the rate of the non-used items allows the determination of the degree of the item pool use.

The test overlap index takes the number of overlapping items and the number of the same items encountered for two individuals who were selected randomly into account. The test overlap index was calculated with the help of the following equations (Chen, Ankenmann, & Spray, 1999):

$$\sum_{i=1}^{n} \binom{m_i}{2}$$

During CAT application, on the condition that m_i: the number of time i. item is used and *n:* the size of the item pool, if $m_i < 2^2$, $\binom{m_i}{}$= 0. In other words, if any item is used only once, this does not affect the test overlap index in CAT applications. In general, the mean of the overlap index among the tests was calculated by the following formula.

$$\bar{T} = \frac{\sum_{i=1}^{n} \binom{m_i}{2}}{k \binom{N}{2}}$$

N shows the number of the participants in CAT application and k shows the constant test length in CAT application. That the test overlap index is larger means that test security is damaged (Huang, Chen, & Wang, 2012).

In order to determine the most ideal item exposure method in accordance with the limitations of the research, the result of F statistics was used (Chang et al., 2001; Grubbs, 1973). Chang & Ying (1999, p.215) state that "When item exposure rates are compared, $\chi2_{method1}$ and $\chi2_{method2}$ are needed to be compared by calculating χ^2 statistics".

($F_{\chi2_{\text{method1}},\chi2_{\text{method2}}} = \frac{\chi2_{\text{method1}}}{\chi2_{\text{method2}}}$) was defined as a measure of the comparison. If $F_{\chi2_{\text{method1}},\chi2_{\text{method2}}} < 1$, the first method is considered to be better than the second method in terms of balancing the item exposure rates in general. χ^2 statistics was calculated with the help of the Equation 1 below (Tay, 2015):

$$x^2 = \frac{\sum_{i=1}^{n}(r_i - uni(r_i))^2}{uni(r_i)} \text{ ...... (Equation 1)}$$

The sum of the squares of the difference between the item exposure rates (r_i) and the uniform item exposure rate that is considered for all items (the ratio of test length to the size of the item pool ($uni(r_i) = k/n$)) was calculated by dividing it to the uniform item exposure rate.

Whether there is a significant difference between the standard deviation values of the item exposure rate related to the test security in different measurement conditions was examined with "*Hypothesis testing for two variants*". Whether there was a significant difference among the maximum item exposure rate; the item pool utilization and the test overlap indexes related to the item exposure control methods in different measurement conditions was examined with "*the hypothesis test of the difference between two ratios*".

The fidelity, bias and RMSE coefficients for measurement precision by taking the averages of the values obtained in 100 repeats in each condition (Leroux et al., 2013; Ross, 2013; Zheng & Chang, 2014) into consideration; the maximum item exposure rate, the standard deviation of the item exposure rate, the item pool utilization and the test overlap index for the test security; χ^2 statistics to compare the item exposure rates and the values related to the F statistic to determine the most ideal item exposure method were calculated in Excel program separately. In order to calculate the Mahalanobis distances of the bias and the RMSE values, the "mahalanobis" command in the package {stats} that is in the R software program was used (R Core Team, 2017).

Results

In this chapter, there are findings and comments about the research questions that are studied within the framework of the purpose of the research.

The Effect of Item Exposure Control Methods on Measurement Precision in Different Measurement Conditions in CAT Applications

In order to examine the effect of item exposure control methods on measurement precision in different measurement conditions in CAT applications, below are the findings and comments related to the coefficient of fidelity, the bias values and the RMSE values, respectively. In this respect, the effect of item exposure control methods on measurement precision in different measurement conditions is presented in Table 1.

Tablo 1. Effect of Control Methods on Measurement Precision in Different Measurement Conditions in CAT Applications

Test Length	Sample Size	Ability Distribution	Item Exposure Control Methods								
			No Method			SH			AD		
			Fidelity	*Bias*	*RMSE*	*Fidelity*	*Bias*	*RMSE*	*Fidelity*	*Bias*	*RMSE*
Short (25 items)	**Small (n=250)**	**Right Skewed**	.9990	.1236	.1523	.9988	.1249	.1584	.9954	.2111	.2715
		Normal	.9997	-.0025	.0428	.9997	-.0035	.0434	.9997	-.0044	.0497
		Uniform	.9997	.0111	.1014	.9998	.0094	.0987	.9997	.1277	.1112
		Left Skewed	.9992	-.0922	.1155	.9991	-.0963	.1201	.9977	-.1420	.1822
	Large (N=1000)	**Right Skewed**	.9990	.1113	.0045	.9989	.1135	.0046	.9962	.1844	.0076
		Normal	.9997	.0009	.0015	.9997	-.0013	.0015	.9997	-.0016	.0017
		Uniform	.9998	.0095	.0032	.9998	.0097	.0032	.9998	.0125	.0036
		Left Skewed	.9993	-.0881	.0036	.9993	-.0895	.0036	.9985	-.1263	.0051
Long (50 items)	**Small (n=250)**	**Right Skewed**	.9994	.0668	.0863	.9994	.0689	.089	.9964	.1526	.2043
		Normal	.9999	-.0004	.0241	.9999	-.0003	.0242	.9999	-.0025	.0285
		Uniform	.9999	.0056	.0546	.9999	.0044	.0544	.9999	.0076	.0638
		Left Skewed	.9996	-.0522	.0681	.9996	-.0524	.0677	.9985	-.0985	.1321
	Large (N=1000)	**Right Skewed**	.9995	.0587	.0025	.9996	.0597	.0025	.9973	.1275	.0055
		Normal	.9998	-.0005	.0008	.9998	-.0001	.0009	.9998	-.0002	.0009
		Uniform	.9998	.0059	.0018	.9998	.0051	.0018	.9998	.0074	.0020
		Left Skewed	.9997	-.0465	.0020	.9997	-.0467	.0020	.9989	-.0816	.0035

In CAT applications, it is seen that the fidelity coefficients in different measurement conditions ranged between .9954 and .9999. In this case, it can be said that there is a high level of fidelity between the estimated theta parameters and true theta parameters estimated in different measurement conditions. In CAT applications, Cohen q values for fidelity coefficients in different measurement conditions vary between .00 and .90. There were no significant differences between the fidelity coefficients when the Sympson Hetter Strategy was used in different measurement conditions (SH) and no item exposure control method was used (p> .05). In addition, it was concluded that there were no significant differences between the calculated fidelity coefficients of item exposure methods in normal and uniform ability distributions (p> .05).

In terms of effect size, a large significant difference was found in favor of the fidelity coefficients calculated when the Sympson Hetter Strategy is used and the item exposure control method was not used compared to the fidelity coefficients calculated when Fade Away Strategy in the right skewed distributions (p <.05). In the left skewed distributions on the other hand, there was a moderate significant difference when there was a short test between the fidelity coefficients calculated when Fade Away Strategy and Sympson Hetter Strategy were used, while a large significant difference was found in long tests (p <.05). In addition to this, except for the short test-large sample condition in the left skewed distributions, while there was a large significant difference between the fidelity coefficients calculated when Fade Away Strategy was used and the fidelity coefficients when the item exposure control methods were not applied; it was found that there was a moderate significant difference, in the short test-large sample condition (Cohen q = .38) it was concluded that there was a moderate significant difference.

When the fidelity coefficients of item exposure control methods in different measurement conditions were evaluated in general, it was concluded that there was a high level of fidelity between the estimated theta parameters and true theta parameters in each condition. Moreover, when other measurement conditions were kept constant, as the skew coefficient of the ability distributions came closer to zero, although the fidelity coefficients related to the item exposure control methods increased, no significant difference between them was found. It was concluded that there was a significant difference in favor of the fidelity coefficients calculated when the item exposure control method was not used and Sympson Hetter Strategy was used compared to the fidelity coefficients calculated when Fade Away Strategy was used in the right and left skewed distributions. In other measurement conditions, no significant difference was found between the fidelity coefficients calculated for the item exposure control methods.

While there was a moderate or pre-moderate effect size (*Cohen q=.53*) in short tests with left skewed distributions among the fidelity coefficients calculated when Fade Away strategy was used and the fidelity coefficients calculated when Sympson Hetter strategy was used but the item exposure control method was not used; there was found to be a large effect size in short tests with right skewed distributions. In addition, there was a large effect size between the fidelity coefficients calculated when Fade Away Strategy was used and the fidelity coefficients calculated when the item exposure control method was not used but Sympson Hetter Strategy was used in both the right and left skewed distributions in long tests. Therefore, it can be stated that in case of selecting Fade Away Strategy from the item exposure control methods in skewed distributions, the fidelity coefficient between the estimated theta level and the true theta level according to the other control methods of item exposure is significantly reduced. In other words, it can be said that the error will increase as the fidelity coefficient between the estimated theta level and the true theta level will be greatly reduced.

When the bias values are examined, it is seen that it changes between -.1420 and .2111. Since the long test, the large sample, the normal ability distribution and the bias value (-.0001) are the close to zero in the Sympson Hetter Strategy condition, the average difference between the estimated theta level and the true theta level is the lowest. Mahalanobis distances were examined by calculating whether there was a significant difference between the bias values of item exposure control methods in different measurement conditions.

The Mahalanobis distance values of the bias coefficients calculated for the item exposure control methods ranged from .19 to 14.04 (p> .001). In this respect, it can be interpreted that there is no significant change in the mean differences between the estimated theta level on the item exposure control methods and the true theta level in the different measurement conditions.

In the long test-large sample and normal ability distribution, the calculated bias value when the item exposure control method is not used is further from zero compared to the calculated bias values when the Sympson Hetter Strategy and Fade Away Strategy are used. In all conditions except this condition, the calculated bias values when using Fade Away Strategy in terms of the item exposure control methods are observed to be further from zero compared to the calculated bias values when the item exposure control method is not applied and when using the Sympson Hetter Strategy. Therefore, provided that other measurement conditions are kept constant, the mean difference between the estimated theta level and the true theta level when Fade Away Strategy among the item exposure control methods is used is higher than that of the other methods.

Since the bias values are close to zero under all conditions, it can be stated that the mean difference between the estimated theta parameters and the true theta parameters in each condition is low. In addition, there are no significant differences in the findings obtained regarding the bias values in each measurement condition. When the other measurement conditions are kept constant, generally, as the sample size and test length increased and the skew coefficient of the ability distributions approached to zero, it was found that the mean difference between the estimated theta level and the true theta level decreased when Fade Away Strategy among the item exposure control methods was not used.

When the RMSE values are examined, provided that the test is long, the sample is large and it shows a normal ability distribution and the item exposure control method is not applied, the lowest RMSE value is found to be (.0008); however, when the test is short, the sample is small and it shows a right skewed distribution with Fade Away Strategy, the highest RMSE value is found to be (.2715).

Whether there was a significant difference between the RMSE values of item exposure control methods in different measurement conditions was examined by calculating the Mahalanobis distances. The Mahalanobis distance values of the calculated RMSE coefficients as to the item exposure control methods ranged between .61 and 7.32 (p> .001). In this respect, it has been concluded that there is no significant change between the estimated theta levels as to the item exposure control methods and the true theta levels in terms of absolute differences in different measurement conditions.

In the long test-large samples, since the RMSE values were found to be less than .05, there was found to be an perfect fit (Brown, 2006; Jöreskog & Sörbom, 1993). In addition, according to the RMSE values regarding the item exposure control methods in different measurement conditions, it is listed from the highest to the lowest as following; Fade Away Strategy, the Sympson Hetter Strategy and the RMSE values calculated when the item exposure control method is not applied. Accordingly, it was found that the difference between the covariance regarding the estimated theta parameters and the covariance regarding the true theta parameters was getting closer to zero when the item exposure control method was not applied in different measurement conditions compared to the conditions when the item exposure control method was applied. Therefore, since the RMSE values calculated when the item exposure control method is not applied under different measurement conditions are smaller than the RMSE values calculated when using the Sympson Hetter Strategy and Fade Away Strategy, it can be concluded that the absolute differences between the estimated theta levels and the true theta levels can be reduced when the item exposure control method is not applied.

Keeping that there are no significant differences in RMSE values in different measurement conditions in mind and when other measurement conditions were kept constant, in general, as the sample size and the test length increased, as the skew coefficient of the ability distributions approached zero and when Fade Away Strategy was not used, the absolute difference between the estimated theta level and the true theta level decreased.

In order to examine the effect of the item exposure control methods on measurement precision in different measurement conditions in CAT applications, since as the measurement precision increases, the error values will decrease, the concordation, the bias and the RMSE coefficients which are the error indicators in each condition should be evaluated together (Gu & Reckase, 2007; Wang & Vispoel, 1998; Zheng & Chang, 2014). In line with this information, when the findings about the measurement precision are evaluated, it is concluded that the measurement precision increases in general as the skew coefficient of the ability distributions is near zero and as the sample size and test length increase. In addition, it was concluded that there were no significant differences among the item exposure control methods in terms of the bias and the RMSE values, although the bias closest to zero and the lowest RMSE value were reached when no item exposure control method was applied.

In terms of the fidelity coefficients on the other hand, while there are no significant differences between the fidelity coefficients regarding the cases when the item exposure control method is not applied but the Sympson Hetter Strategy is used; there is a significant difference in favor of the fidelity coefficients calculated when the item exposure control method is not but the Sympson Hetter Strategy is used compared to the fidelity coefficients calculated when Fade Away Strategy is used in right and left skewed distributions.

When the item exposure control method was not applied, the measurement precision was found to be generally higher than that of the conditions when the item exposure control method was applied. This finding is in line with the findings of the study of Boyd (2003) and Davis (2004) in which the case when the item exposure is not controlled is taken as a reference. As the item information function is taken into consideration in the item selection in CAT applications, the fact that the standard error of the theta estimation is decreased by selecting the items with high information value can be cited since the item selection is not restricted when the item exposure control is not used. In other words, the fact that the standard error of the theta estimation is decreased when the item with high information value is selected can be said to be the reason for that the measurement precision when the item exposure is not controlled is generally higher than that of the conditions when the item exposure is controlled.

When selecting the items that will maximize the measurement precision, since the item exposure ratio will not be distributed evenly (Pastor et al., 2002) in order not to use the item pool in an unbalanced manner and not to use only a limited number of items (Hulin, Drasgow, & Parson 1983, as cited in Revuelta & Ponsoda, 1998) by means of the item exposure control methods, it is aimed to provide the test security by using the item pool in a more balanced way without reducing the measurement precision (Pastor et al., 2002). In this case, when the relevant variables are considered together, there are generally no significant differences between the measurement precision under the different item exposure controlled conditions. As a result of this finding, it can be stated that the measurement precision will not be reduced when the related item exposure control methods are applied.

The effect of the item exposure control methods on test security in different measurement conditions in CAT applications

In order to investigate the effect of item exposure control methods in different measurement conditions on test security in CAT applications, the results and the discussion on standard deviations of item exposure ratios, the distribution of maximum item exposure ratio, indexes of the item pool utilization, the skew coefficients of the item exposure ratios (χ^2values), test overlap indexes are given below, respectively. Table 2 represents the data obtained from the effect of item exposure control methods at different measurement conditions on test security in CAT applications.

Table 2. The effect of the item exposure control methods on test security in different measurement conditions in CAT applications

Test Length	**Sample Size**	**Ability distribution**	***Item Exposure Control Methods***														
			No method					***SH***					***FA***				
			SD	*Max*	*Benefit*	χ^2	*Overlap*	*SD*	*Max*	*Benefit*	χ^2	*Overlap*	*SD*	*Max*	*Benefit*	χ^2	*Overlap*
Short (25 items)	**Small (n=250)**	**Right skewed**	.097	.834	.1259	375.06	.396	.091	.622	.1256	330.99	.353	.048	.193	.0980	90.78	.112
		Normal	.074	.750	.1254	218.66	.239	.071	.486	.1253	201.84	.223	.037	.117	.0939	54.50	.076
		Uniform	.068	.677	.1242	183.84	.205	.064	.448	.1242	165.06	.187	.038	.152	.0952	58.49	.079
		Left skewed	.101	.922	.1272	406.71	.429	.092	.642	.1264	341.64	.363	.044	.184	.0928	77.97	.099
	Large (N=1000)	**Right skewed**	.095	.830	.0308	362.35	.386	.090	.594	.0308	321.22	.346	.047	.189	.0243	89.39	.114
		Normal	.075	.739	.0309	224.52	.249	.072	.492	.0309	208.19	.232	.037	.117	.0232	54.17	.078
		Uniform	.068	.683	.0310	182.74	.207	.064	.454	.0309	164.07	.188	.038	.151	.0238	58.23	.082
		Left skewed	.099	.915	.0312	387.70	.412	.091	.623	.0308	328.23	.352	.044	.181	.0239	76.78	.101
Long (50 items)	**Small (n=250)**	**Right skewed**	.142	.840	.0534	403.50	.452	.138	.727	.0528	381.86	.429	.063	.200	.0312	79.54	.126
		Normal	.098	.761	.053	193.19	.239	.096	.508	.0536	185.85	.232	.054	.151	.0322	58.51	.105
		Uniform	.089	.686	.051	158.61	.205	.087	.463	.0504	150.22	.197	.055	.183	.0330	60.25	.107
		Left skewed	.143	.9294	.0533	406.23	.454	.137	.733	.0530	375.25	.422	.060	.196	.0288	72.14	.119
	Large (N=1000)	**Right skewed**	.139	.836	.0127	389.23	.439	.136	.721	.0126	369.19	.419	.063	.197	.0075	78.38	.128
		Normal	.099	.749	.0127	197.60	.247	.098	.523	.0126	191.56	.241	.054	.151	.0078	57.01	.106
		Uniform	.089	.689	.0126	157.27	.206	.086	.461	.0126	148.79	.198	.055	.181	.0082	59.82	.109
		Left skewed	.139	.921	.0128	387.87	.437	.134	.714	.0127	359.85	.409	.060	.192	.0069	71.50	.121

When table 2 is examined, the standard deviation values of the item exposure rates are found to be (.143) and (.037) when the highest value long test-small sample-left skewed distribution and the item exposure control method are not applied and the lowest value short test-small sample-normal distribution and Fade Away Strategy are applied, respectively. When the results of the hypothesis test that there is a significant difference between the standard deviation values of item exposure among the control methods in different measurement conditions, while F_{NM-FA} and F_{SH-FA} values calculated in line with the rates of the binary sample variances take place in the hypothesis rejection zone in all measurement conditions; only F_{NM-SH} takes place in the admission zone except the condition in which the test is short, the sample is large and the distribution is left-skewed. Consequently, when the item exposure control method was not applied and Sympson Hetter Strategy was applied (except short test-large sample- left skewed ability distribution condition) there was no significant difference between the standard deviation values of the item exposure rates ($p>.05$). Furthermore, there was a significant difference between the standard deviation of the item exposure rates obtained from Fade Away Strategy and the standard deviation obtained from the test item exposure control method were not applied ($F_{no\ method-FA}$) and Sympson Hetter Strategy (F_{SH-FA}) was applied ($p<.05$).

Although there was a significant difference between the standard deviations of the item exposure rates under short test-large sample- left skewed ability distribution conditions when the item exposure control method was not applied and Sympson Hetter Strategy was applied, F value was very close to the critical value ($F_{no\ method-SH}$ = 1.18>$F_{999,999,.025}$ = 1.13). As a reason for this finding, besides the difficulties experienced at selecting the items in the item pool that is able to appeal to the theta levels (in large sample) of the individuals (super talented) in the left skewed ability distributions, the fact that the test termination rule is a short test can be said to cause an increase in the skew among the item exposure rates according to whether a limitation is made in the item selection method by using the prior distribution or not.

When the findings related to the standard deviations of the item exposure rates obtained under different measurement conditions are examined in general, while there is no significant difference between SH and FA, there is a significant difference between NoMethod and FA. The standart devations of the item exposure rates when Fade Away strategy is applied are significant smaller than those of other two methods. Consequently, it can be said when Fade Away Strategy is used, the item exposure rates distribute more homogenously than the cases when the other strategies are used.

When the other measurement conditions are kept constant, in general, as the test length decreases, the skew coefficient of the ability distributions approaches to the zero, the sample size increases in general and since the calculated standard deviations of item exposure ratios decreases when Fade Away Strategy is used among the item exposure control methods, it is seen that the item exposure rates show a more homogenous distribution. In other words, as the item exposure ratios show a more homogenous distribution under corresponded conditions, the item exposure shows a balanced distribution and the item pool is used more efficiently.

Regarding the maximum item exposure rate, when the highest value long test-small sample-left skewed distribution and the item exposure control method were not applied, it was found to be (.929); when the lowest value short test-both small and large sample-normal distribution and Fade Away Strategy were applied, it was found to be (.117) When the results of the test regarding the item exposure control methods under different measurement conditions on the difference between two rates are examined in order to understand whether there is a significant difference between the item exposure ratios, z values which are regarding differences between the maximum item exposure ratios were found to be ranging from 5.19 to 48.82. Therefore, it can be said that the differences between two ratios are significant (p<.05).

According to the maximum item exposure ratios calculated related to the item exposure control methods at different conditions, the order from the lowest to the highest is Fade Away Strategy, Sympson Hetter Strategy and the provision in which the item exposure was not applied. Since as item exposure rate increases, the posibilities of items' being disclosed increases, under related conditions when Fade Away Strategy is used, it is found that the risk of the items' losing psychometric properties by being disclosed decreases more significantly compared to the other conditions.

In order to determine degree of the item pool use under each condition, when the pool utilization index is investigated, the degrees of the pool use were found to be (.1259) when the highest value short test-small sample- right skewed distribution and the item exposure control method are not applied; and (.0069) when the lowest value long test-large sample-left skewed distribution and Fade Away Strategy are applied. Therefore, the degree of the item pool use ranges from *(1-.1259=%87.41) to (1-.0069=%99.31)* under different conditions.

Regarding the item exposure control methods under different measurement conditions, in order to determine whether there is a significant difference between the item pool utilization rates or not, when the results of the variation between two ratios hypothesis test are investigated, it is found that z values of the difference between the item pool unutilization ratios are found to be ranging from -.03 to 1.38. Thereby, it can be said that there is no significant difference between two rates ($p>.05$).

There is no significant difference between the item pool utilization rates under different measurement conditions. In addition to that, when other measurement conditions are kept constant, as the sample size and the test length increase, the item pool utilization rates decrease when Fade Away Strategy among the item exposure control methods is used. Consequently, it can be said as the sample size and the test length increase and Fade Away Strategy among item exposure control methods is used, the item pool use degree increases.

When the skew coefficient (χ^2 values) of the item exposure rates is investigated in order to determine the distance between the item exposure rates regarding the item exposure methods and the intended item exposure ratio, in other words, to determine how balanced is the distribution that the item pool use shows, when the highest χ^2 value is found to be χ^2 = 406.71 when the short test-small sample- left skewed distribution and the item exposure control method are not applied; the lowest χ^2 value is found to be 54.17 when the short test-large sample-normal distribution and Fade Away Strategy are applied.

According to the calculated χ^2 values, from the lowest to the highest the order is found to be; Fade Away Strategy, Sympson Hetter Strategy, and the item exposure control method is not applied. Hence, in order to decide which method is the ideal item exposure method, when the item exposure control is not applied or either Sympson Hetter Strategy or Fade Away Strategy among the item exposure control methods is applied, it is found to be $F_{\chi 2_{FA}, \chi 2_{SH}}$<1 and $F_{\chi 2_{NOM}, \chi 2_{SH}}$>1 regarding the comparison criteria F statistics under each condition In this case under each condition, it is seen that not applying an item exposure control method ($F_{\chi 2_{NOM}, \chi 2_{SH}}$>1); is a less useful strategy than Sympson Hetter Strategy in terms of the point that the item exposure rates are balanced in general. It is also seen that Fade Away Strategy($F_{\chi 2_{FA}, \chi 2_{SH}}$<1), on the other hand, is better strategy than Sympson Hetter Strategy. In the direction of the findings and the related conditions, it is found that when Fade Away Strategy among the item exposure control methods is used, the item pool shows a more balanced distribution.

When other measurement conditions are kept constant, the item exposure control method is not applied and Sympson Hetter Strategy is applied and when the *test length increases,* while χ^2 value increases at the right and left skewed distribution (except the condition of small sample-no control method is applied), χ^2 value decreases with normal and uniform distributions. In case Fade Away Strategy is applied, in contrast with the findings obtained under other two strategies' conditions, *when test length increases*, while χ^2 value decreases at the right and left skewed distribution; χ^2 value increases at normal and uniform distribution.

When the sample size increases, χ^2values decrease at right and left skewed and uniform distributions. In case of normal distributions, while χ^2value increases when the item exposure control method is not applied and Sympson Hetter Strategy is applied; χ^2value decreases when Fade Away Strategy is used. As the skew coeffiecient of ability distributions approaches to the zero, χ^2 value decreases. When Fade Away Strategy is applied, χ^2value *decreases* towards *the right and left skewed, uniform and normal distribution* respectively. Moreover, when Fade Away Strategy is applied, while χ^2 value is the smallest at *normal distributions*; χ^2 value is the smallest at *uniform distributions* when the item exposure control method is not applied and Sympson Hetter Strategy is applied.

When the results obtained for the skew coefficients (χ^2 values) of the item exposure rates at different measurement conditions are examined in general, the lowest χ^2 value is obtained in the short test-large sample-normal distribution condition when Fade Away Strategy is applied. When the item exposure control method is not applied and the Sympson Hetter Strategy is applied, the lowest χ^2 value is obtained in the long test-large sample- uniform distribution condition. However, in all conditions, the calculated χ^2 value is found to be smaller than that of the other strategies when Fade Away Strategy is applied. Hence, when Fade Away Strategy among item exposure methods is used, it can be said that the distance between the item exposure rate and the intended exposure ratio (.025 in short tests, .05 in long tests) decreases. In other words, when Fade Away Strategy is applied as the skew coefficient of the item exposure rate approaches to the zero, the item pool use shows a more balanced distribution.

When test overlapping indexes regarding the item exposure control methods in different measurement conditions are investigated, the highest value regarding the test overlap indexes in different measurement conditions is found to be (.454) when long test-small sample-left skewed distribution are applied and the item exposure control method is not applied. On the other hand, the lowest value is found to be (.076) when short test-small sample-normal distribution and Fade Away Strategy are applied.

When the hypothesis test results of the difference between two ratios regarding whether or not there is a significant difference between the test overlap indexes related to the item exposure control methods in different measurement conditions, z_{NM-FA} values range between 3.05 and 17.06, z_{SH-FA} values range between 2.82 and 15.47. Therefore, under related measurement conditions, there was a significant difference between test overlap indexes calculated when Fade Away Strategy was applied and test overlap indexes calculated when item exposure control methods were not applied (z_{NM-FA}) and the ones when Sympson Hetter Strategy was applied (z_{SH-FA}) ($p<.05$). There was no significant difference between the calculated test overlap indexes when item exposure control method was not applied Sympson Hetter Strategy was applied (except short test-large sample-left skewed ability distribution z_{NM-SH} = 2.78) ($p>.05$).

When the findings obtained from the calculated test overlap indexes under different measurement conditions were examined in general, it was found that while there was no significant different between SH and NM in general, there were significant differences between SH and FA and between NM and AD. When Fade Away Strategy was applied, it was found that calculated test overlap indexes were significantly smaller than those of other methods. Therefore, as the test overlap indexes increase, the test security is damaged (Huang et al., 2012), when Fade Away Strategy is used it can be said that the test security is less damaged than those of item exposure control is not applied and Sympson Hetter Strategy is applied.

Under the following conditions in which test length and ability distribution are kept constant, item exposure control method is not applied and Sympson Hetter Strategy is applied, while test overlap indexes decrease when sample size increases in right skewed and left skewed distributions; overlap indexes increase in normal and uniform ability distributions. When Fade Away Strategy is applied, as the sample size increases test overlap indexes also increase.

When the test length and the ability distribution are kept constant and the sample size is increased, under the conditions in which item exposure control method is not applied and Sympson Hetter Strategy is applied, although test overlap indexes decrease in right skewed and left skewed distributions, since test overlap indexes are higher compared to the normal and uniform ability distributions, it can be concluded as test security is affected in a more negative way. Moreover, even if the test overlap indices increase as the sample size increases when Fade Away Strategy is used, it can be reported that test security is affected in a more positive way since test overlap index in each condition are smaller considerably compared to the test overlap indexes calculated in other item exposure control methods.

When other measurement conditions are kept constant and under the conditons in which item exposure control method is not applied and Sympson Hetter Strategy is applied; when the test length increases, it is found that while test overlap indexes increase in right and left skewed distributions, there is no significant difference at normal and uniform distributions. In this case, as the test length increases, it can be said that test security is negatively affected in right and left skewed ability distributions since the width between the calculated test overlap indexes in right and left skewed ability distributions and the calculated test overlap indexes in normal and uniform ability distributions. When Fade Away Strategy was applied, although the test overlap indexes increase as the test length increases, when it was compared with the calculated test overlap indexes in other item exposure control methods, it can be said that test security is affected more positively as it is significantly smaller in each condition.

It was found that as the skew coefficient of ability distributions approached to the zero, test overlap indexes decreased. When Fade Away Strategy is applied, test overlap indexes decrease in right skewed, left skewed, uniform and normal distributions respectively. While the test overlapping index is the smallest in normal distribution when Fade Away Strategy is applied, the test overlapping index is the smallest in uniform distribution when item exposure control method is not applied and Sympson Hetter Strategy is used.

When the findings of test overlap indexes under different measurement conditions were investigated in general, the lowest test overlapping index was obtained when Fade Away Strategy was applied under short test-small sample-normal distribution condition. The smallest test overlapping indice was obtained when item exposure control method was not applied and Sympson Hetter Startegy was applied under short test-small sample-uniform distribution condition. However, the calculated test overlap indexes were found to be significantly smaller than those of other strategies when Fade Away Strategy was applied in all conditions. Therefore, in CAT applications, it can be said that test security is less damaged and item pool is provided to be used for years when Fade Away Strategy is applied among item exposure control methods since the number of same items that any two individuals encounter decreases.

That the value of item exposure rate $r_i = .20$ means, the related item is used in approximately one of five CAT applications (Weiss & Guyer, 2012). Therefore, as the item exposure rate increases, this causes the factitious descent of the psychometric properties of the related item to decline by disclosing it (Segall, 2004; Revuelta & Ponsoda, 1998). Because of that, neither the item exposure nor never using items in the item pool is an intended situation. In line with this information, in order to determine degree of efficient use of item pool, the distribution regarding item pool use in different measurement conditions is given in Table 3.

Table 3. The Distribution Regarding Item Pool Use in Different Measurement Conditions

Test Length	Sample Size	Ability Distrubition	Exposure Control Method	r_i=0	0<r_i≤.2	.2<r_i≤.4	.4<r_i≤.6	.6<r_i≤.8	.8<r_i≤1
Short (25 items)	Small (n=250)	Right Skewed	*NM*	787	168	18	20	6	1
			SH	785	165	28	20	2	0
			FA	613	387	0	0	0	0
		Normal	*NM*	784	177	33	3	3	0
			SH	783	173	36	8	0	0
			FA	587	413	0	0	0	0
		Uniform	*NM*	776	191	28	4	1	0
			SH	776	181	42	1	0	0
			FA	595	405	0	0	0	0
		Left Skewed	*NM*	795	168	13	15	6	3
			SH	790	166	19	23	2	0
			FA	608	392	0	0	0	0
	Large (N=1000)	Right Skewed	*NM*	771	186	19	19	4	1
			SH	770	180	28	22	0	0
			FA	608	392	0	0	0	0
		Normal	*NM*	773	186	32	6	3	0
			SH	772	184	35	9	0	0
			FA	581	419	0	0	0	0
		Uniform	*NM*	774	195	26	4	1	0
			SH	773	185	41	1	0	0
			FA	596	404	0	0	0	0
		Left Skewed	*NM*	780	182	16	15	5	2
			SH	771	184	21	23	1	0
			FA	598	402	0	0	0	0
			SH	633	286	31	32	18	0
			FA	344	656	0	0	0	0
Long (50 items)	Small (n=250)	Right Skewed	*NM*	668	251	28	26	24	3
			SH	660	253	37	26	24	0
			FA	390	610	0	0	0	0
		Normal	*NM*	667	226	95	9	3	0
			SH	670	222	98	10	0	0
			FA	403	597	0	0	0	0
		Uniform	*NM*	634	279	82	4	1	0
			SH	630	274	95	1	0	0
			FA	412	588	0	0	0	0
		Left Skewed	*NM*	666	261	26	16	28	3
			SH	662	259	29	26	24	0
			FA	360	640	0	0	0	0
	Large (N=1000)	Right Skewed	*NM*	635	280	33	31	18	3
			SH	633	279	36	32	20	0
			FA	374	626	0	0	0	0
		Normal	*NM*	633	270	85	9	3	0
			SH	631	272	81	16	0	0
			FA	390	610	0	0	0	0
		Uniform	*NM*	631	283	81	4	1	0
			SH	630	274	95	1	0	0
			FA	412	588	0	0	0	0
		Left Skewed	*NM*	639	287	28	20	23	3
			SH	633	286	31	32	18	0
			FA	344	656	0	0	0	0

When Table 3 is analysed, as a result of 100 replications the number of the items which are not used in any CAT application (r_i= 0), is found to be ranging between 631 and 795 when item exposure control method is not applied, while it is ranging between 630 and 790 at Sympson Hetter Strategy, between 344 and 613 at Fade Away Strategy, respectively. When item exposure control method was not applied and Sympson Hetter Strategy was applied in related conditions, big differences were not found in terms of the number of the items that were not used in any CAT application (r_i=0). When Fade Away Strategy was applied on the other hand, it was found it is smaller than those of other strategies in terms of the number of the items not used in all conditions (r_i=0).

It can be indicated that the number of the items, which are not used in any CAT applications ((r_i=0), in other words, in terms of the item pool utilization, is smaller than those under other measurement conditions at long test and large samples. Similarly, regarding the degree of the item pool use, it was seen that item pool use degree is higher at long test and large samples.

The intended uniform exposure rates for all items are .025 and .05 for short test and long test, respectively (Tay, 2015). Besides, the targeted exposure rate is defined as .20 in the related literature (Eggen, 2001; Eignor, Stocking, Way, & Steffen, 1993; Veldkamp & Van Der Linden, 2008). In line with this information, when the values in the item exposure ranged as $.00<r_i\leq.20$ were investigated regarding effective use of item pool, it was found that while there were not big differences at uniform distribution conditions when test length and sample size increased, the number of the items used increased at all other measurement conditions.

When Fade Away Strategy is applied, there seems to be no number of the items which is higher than targeted exposure rate (r_i>.20) and is in the range of item exposure. When the distribution of item exposure rates which are higher than targeted exposure rate when Sympson Hetter Strategy was applied was examined, it was seen that since there were too many items higher than targeted item exposure rate in long test-small sample-left skewed distribution, the risk of descent of psychometric properties of these related items was the most by being disclosed (Skewness=-1.86). When Sympson Hetter Strategy was used– in uniform and long tests, in both small and large sampled conditions, as the number of items higher than targeted exposure rate was not too many, it can be indicated that the risk of descent of the psychometric properties of the related items is the lowest by being disclosed (Skewness=2.00).

When the distributions of item exposure rates higher than the targeted exposure rate (r_i=.20) were examined in case item exposure control method was not applied, since the number of items higher than the targeted exposure rate is more in long test-small sample-right skewed distribution, it was seen that the risk of descent of the psychometric properties of the related items was at its maximum by being disclosed (Skewness=-1.09). Besides that, as the number of items higher than the targeted exposure rate is less in long test-uniform distribution, it was seen that the risk of descent of the psychometric properties of related items was at its minimum by being disclosed (Skewness=1.15).

The skewness coefficients of the distribution related to the item exposure rates higher than the targeted exposure rate (r_i=.20) is negative when item exposure control method is not applied in right and left skewed distributions and in the conditions when Sympson Hetter Strategy-long test right and left skewed distributions are applied. Therefore, it can be said that the risk of descent of the psychometric properties increases when item exposure control method is not applied in both short and long tests in right and left skewed distributions, but in only long tests when Sympson Hetter Strategy is applied.

Discussion, Conclusion and Suggestions

When the findings related to measurement precision and test security in different measurement conditions in CAT application process are investigated in general, although the lowest RMSE and the bias value which is the closest to the zero are encountered when item exposure control is not applied, it is concluded that there are no significant differences between item exposure control methods regarding RMSE and bias values. As to the coefficients of fidelity on the other hand, while there are not significant differences between the coefficients of fidelity regarding the conditions in which item exposure control method is not applied and Sympson Hetter Strategy is applied, there is found to be a favourable significant difference between the coefficients of fidelity calculated when item exposure control method is not applied and Sympson Hetter strategy is used when compared to the coefficients calculated when Fade Away strategy is used in right and left skewed distributions.

When the related variables regarding measurement precision were evaluated together, big differences were not found between measurement sensitivities under different item exposure control conditions. In line with this finding, it can be indicated that measurement precision is not reduced when related item exposure control methods are applied.

In order to analyze the effect of item exposure control methods on test security in CAT applications on different measurement conditions, item exposure rate, item pool utilization and the test overlap index are evaluated together (Davis & Dodd, 2005). In this case, when the related variables regarding test security were evaluated together, it was seen that item exposure had a balanced distribution and the item pool use became more efficient by having test security better when Fade Away Strategy was applied in the related conditions (Boyd, 2003; Davis, 2002). In addition to these, when Fade Away Strategy was chosen, since the ratio of overlapping items for individuals (Chen et al., 1999) and the maximum exposure ratio were low, it was found that sustainability would be provided for several applications without disclosing item pool (Revuelta & Ponsoda, 1998).

It is aimed to obtain test security by using the item pool in a more balanced position without descenting the measurement precision with item exposure control methods (Boyd, 2003; Boyd et al., 2013; Davis & Dodd, 2005; Pastor et al., 2002). In line with this, as a result of the conducted research, when Fade Away Strategy is applied instead of Sympson Hetter Strategy which is applied the most common among item exposure control methods under different measurement conditions in CAT applications (Veldkamp et al., 2010), it can be concluded that test security can be obtained by using the item pool in a more balanced position without descenting the measurement precision. This finding is similar to that of Davis (2002) who says although item exposure control parameters are determined with iterative simulations in order to obtain the target exposure rate before the CAT application in Sympson Hetter strategy, this target item exposure rate is not achieved in terms of the item overlapping index and the degree of item pool use. Furthermore, this is also in line with Boztunç Öztürk's (2014) study in which various item exposure control methods are compared according to item selection methods and item pools of various mean difficulty levels, which states test security increases when Fade Away strategy is used instead of Sympson Hetter.

In this study, item exposure control methods were compared to each other in dichotomous items under different measurement conditions in CAT applications. A study can be conducted on the effect of item exposure control methods on test security and measurement precision at different ability distributions, test lengths and sample sizes in the polytomous items. Moreover, in order to avoid sample bias, 100 replications were used for each condition within this study. In order to investigate the effect of number of replications on measurement precision and test security of item exposure control methods, a new study on replicate numbers can be conducted keeping ability distributions, test lengths and sample sizes constant.

References

Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for ınterval estimation of binomial proportions. *The American Statistician, 52*(2), 119-126.

Ariel, A., Veldkamp, B. P., & Van Der Linden, W. J. (2004). Constructing rotating item pools for constrained adaptive testing. *Journal of Educational Measurement, 41*(4), 345-359.

Aytuğ Koşan, A. M. (2013). *Tıp eğitiminde gelişim sınavı soru bankası oluşturulması ve benzetim verileri ile bilgisayar uyarlamalı test uygulaması* (Unpublished doctoral dissertation). Ankara University, Ankara.

Boyd, A. M., Dodd, B., & Fitzpatrick, S. (2013). A comparison of exposure control procedures in CAT systems based on different measurement models for testlets. *Applied Measurement in Education, 26*(2), 113-135.

Boyd, M. A. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* (Unpublished doctoral dissertation). The University of Texas, Austin.

Boztunç Öztürk, N. (2014). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında madde kullanım sıklığı kontrol yöntemlerinin incelenmesi* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, N.J.: Guiford Press.

Bulut, O., & Kan, A. (2012). Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eurasian Journal of Educational Research, 49*, 61-80.

Burt, W. M., Kim, S. J., Davis, L. L., & Dodd, B. G. (2003, April). *A comparison of item exposure control procedures using a CAT system based on the generalized partial credit model*. Paper presented at the Annual meeting of the American Educational Research Association. Chicago, IL.

Byrne, B. M. (1998). *Structural equation modeling with lirsel, prelis, and simplis: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.

Chang, H. H., & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement, 23*, 211-222.

Chang, H. H., Qian, J., & Ying, Z. (2001). A-stratified multistage computerized adaptive testing with b blocking. *Applied Psychological Measurement, 25*, 333-341.

Chang, S. W., & Twu, B. Y. (1998). *A comparative study of item exposure control methods in CAT* (Research Report. 98-3). Iowa City, IA: ACT.

Chen, S. Y., Ankenmann, R. D., & Spay, J. A. (1999). *Exploring the relationship between item exposure rate and test overlap rate in computerised adaptive testing (*ACT Research Report Series. 99-5). Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACT_RR99-05.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Cömert, M. (2008). *Bireye uyarlanmış bilgisayar destekli ölçme ve değerlendirme yazılımı geliştirilmesi* (Unpublished master’s thesis). Bahçeşehir University, İstanbul.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.

Çetin, E. (2009). *Dikey ölçeklemede klasik* test *ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Davey, T., & Parshall, C. G. (1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing.* Paper presented at the Annual meeting of the American Educational Research Association (AERA), San Francisco, CA.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007) Developing theory through simulation methods. *Academy of Management Review, 32*(2), 480-499.

Davis, L. L. (2002). *Strategies for controlling item exposure in computerized adaptive testing with polytomously scored items* (Unpublished doctoral dissertation). The University of Texas, Austin.

Davis, L. L. (2004). Strategies for controlling item exposure in computerized adaptive testing with the generalized partial credit model. *Applied Psychological Measurement, 28*(3), 165185.

Davis, L., & Dodd, B. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement, 27*(3), 335-356.

Davis, L. L., & Dodd, B. G. (2005). *Strategies for controlling item exposure in computerized adaptive testing with partial credit model* (PEM Research Report No. 05-01). Austin, TX: Pearson Educational Measurement.

DeMars, C. E. (2006). Application of the Bi-Factor multidimensional item response theory model to Testlet-Based tests. *Journal of Educational Measurement, 43*(2), 145-168.

Eggen, T. J. H. M. (2001). *Overexposure and underexposure of items in computerized adaptive testing* (Measurement and Research Department Reports. 2001-1). Arnhem, The Netherlands: CITO.

Eignor, D. R., Stocking, M. L., Way, W. D., & Steffen, M. (1993). *Case studies in computer adaptive test design through simulation* (ETS Research Report. RR-93-56). Princeton, NJ: Educational Testing Service.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.

Eroğlu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının* ölçme *kesinliği ve test uzunluğu açısından karşılaştırılması* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Evans, J. J. (2010). *Comparability of examinee proficiency scores on computer adaptive tests using real and simulated data* (Unpublished doctoral dissertation). The State University of New Jersey, New Jersey.

French, B., & Thompson, T. (2003). *The evaluation of exposure control procedures for an operational CAT*. Paper presented at the Annual meeting of the American Educational Research Association (AERA), Chicago, IL.

Georgiadou, E., Triantafillou, E., & Economide, A. A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment, 5*(8), 3-37.

Glas, G. A. W. (2002). Item calibration and parameter drift. In W. J. Van Der Linden & G. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp.183-199). New York: Kluwer Academic Publishers.

Goldman, S. H., & Raju, N. S. (1986). Recover of one and two parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement, 24*, 11-21.

Gökçe, S. (2012). *Comparison of linear and adaptive versions of the Turkish pupil monitoring system (pms) mathematics assessment* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.

Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, *Technometrics, 15*, 53-66.

Gu, L., & Reckase, M. D. (2007). Designing optimal item pools for computerized adaptive tests with Sympson-Hetter exposure control. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerize Adaptive Testing.* Retrieved from http://www.psych.umn.edu/psylabs/CATCentral/

Guyer, R., & Thompson, N. (2011). *Item response theory parameter recovery using Xcalibre 4.1.* St. Paul MN: Assessment Systems Corporation.

Guyer, R., & Thompson, N.A. (2012). *User's manual for Xcalibre item response theory calibration software, version 4.1.6*. St. Paul MN: Assessment Systems Corporation.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate data analysis.* New Jersey: Prentice Hall, Upper Saddle River.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.

Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved from http://www.psych.umn.edu/psylabs/ CATCentral/

Han, K. T. (2011). *User's manual: SimulCAT: Windows software for simulating computerized adaptive test administration.* USA: Graduate Management Admission Council.

Han, K. T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement, 49*(3), 225-246.

Harwell, M. R., & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement, 15*(3), 279-291.

Harwell, M., Stone, C. A., Hsu, T., & Kirişçi, L. (1996). Monte carlo studies in item response theory. *Aplied Psychological Measurement, 20*(2), 101-125.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronical Journal of Business Research Methods, 6*(1), 53-60.

Howell, C. D. (2010). *Statistical methods for psychology*. USA: Cengage Wadsworth.

Huang, H. Y., Chen, P. H., & Wang, W. C. (2012). Computerized adaptive testing using a class of high-order ıtem response theory models. *Applied Psychological Measurement, 36*(8), 689-706.

Hulin, C. L., Lissak, R. L., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study*. Applied Psychological Measurement, 6*(3), 249-260.

İşeri, A. I. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago, IL: Scientific Software.

Kalender, İ. (2009). Başarı ve yetenek kestirimlerinde yeni bir yaklaşım: Bilgisayar ortamında bireyselleştirilmiş testler (Computerized adaptive tests-CAT). *CITO Egitim Kuram ve Uygulama, 5*, 39-48.

Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.

Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Kaskatı, O. T. (2011). *Rasch modelleri kullanarak romatoid artirit hastaları özürlülük değerlendirimi için bilgisayar uyarlamalı test yöntemi geliştirilmesi* (Unpublished doctoral dissertation). Ankara University, Ankara.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide.* United States of America: Sage Publications.

Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması* (Unpublished doctoral dissertation). Ankara University, Ankara.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Publications.

Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* (Unpublished doctoral dissertation). Ankara University, Ankara.

Lee, H. Y., & Dodd, B. G. (2012). Comparison of exposure controls, item pool characteristic, and population distributions for cat using the partial credit model. *Educational and Psychological Measurement, 72*(1), 159- 175.

Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement, 73*, 857-874.

Lord, F., & Stocking, M. (1988). Item reponse theory. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook*. New York: Pergamon Press.

Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement, 28*(4), 989-1020.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lunz, M. E., & Stahl, J. A. (1998). *Patterns of item exposure using a randomized CAT algorithm.* Paper presented at the Annual meeting of the National Council on Measurement in Education, San Diego, CA.

Magis, D., & Raîche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software, 48*(8), 1-31.

McDonald, P. L. (2002). *Computer adaptive test for measuring personality factors using item response theory* (Unpublished doctoral dissertation). The University Western of Ontario, London.

Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma* (Unpublished doctoral dissertation). Ankara University, Ankara.

Öztuna, D. (2008). *Kas-iskelet sistemi sorunlarının özürlülük değerlendiriminde bilgisayar uyarlamalı test yönteminin uygulanması* (Unpublished doctoral dissertation). Ankara University, Ankara.

Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Maidenhead: Open University Press/McGraw-Hill.

Pastor, D. A., Dodd, B. G., & Chang, H. H. (2002). A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement*, *26*(2), 147-163.

Patsula, L. N., & Gessaroli, M. E. (1995, April). *A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different test lengths and sample sizes.* Paper presented at the Annual meeting of the National Council on Measurement in Education, San Francisco.

R Core Team. (2017). R: *A language and environment for statistical computing*. Retrieved from https://www.R-project.org/.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 35*(4), 311-327.

Ross, S. M. (2013). *Simulation.* USA: Elsevier.

Rudman, H. C. (1987). The future of testing is now. *Educational Measurement: Issues and Practice, 4*, 5-11.

Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory* (Unpublished doctoral dissertation). University of Minnesota, USA.

Segall, D. O. (2004). Computerized adaptive testing. In K. Kempf-Leanard (Ed.), *The encyclopedia of social measurement* (pp. 429-438). San Diego, CA: Academic Press.

Smits, N., Cuijpers, P., & Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research, 188*, 145-155.

Speron, E. (2009). *A comparison of metric linking procedures in ıtem response theory* (Unpublished doctoral dissertation) Chicago, Illinois: Illinois Institute of Technology.

Stocking, M. L., & Lewis, C. (1995). *A new method of controlling item exposure in computerized adaptive testing* (Research Report. 95-25). Princeton, NJ: Educational Testing Service.

Stocking, M. L., & Lewis C. (2002). Methods of controlling the exposure of items in CAT. In W. J. Van Der Linden & G. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 163-182). New York: Kluwer Academic Publishers.

Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması* (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk Psikoloji Yazıları, 3*(6) 49-74.

Şahin, A., & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17*, 321-335.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik.* Ankara: Seçkin Yayıncılık.

Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları.* Ankara: Ekinoks Yayınları.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.

Tang, K. L., Way, W. D., & Carey, P. A. (1993). *The effect of small calibration sample sizes on TEOFL IRT-based equating* (TOEFL technical report. TR-7). Princeton, NJ: Educational Testing Service.

Tay, P. H. (2015). *On-the-fly assembled multistage adaptive testing* (Unpublished doctoral dissertation). The University of Texas, Urbana.

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika, 47*(4), 397-412. doi:10.1007/BF02293705

Thompson, T. (2002, April). *Employing new ideas in CAT to a simulated reading test*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Vaughn, B. K., & Wang, Q. (2010). DIF trees: using classifications trees to detect differential item functioning. *Educational and Psychological Measurement, 70*(6), 941-952.

Veldkamp, B. P., & Van Der Linden, W. J. (2008). Implementing Sympson-Hetter item-exposure control in a shadow-test approach to constrained adaptive testing. *International Journal of Testing, 8*(3), 272289.

Veldkamp, B. P., & Van Der Linden, W. J. (2010). Designing item pools for adaptive testing. In W. J. Van Der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing*. New York: Springer.

Veldkamp, B. P., Verschoor, A. J., & Eggen, T. J. H. M. (2010). A multiple objective test assembly approach for exposure control problems in computerized adaptive testing. *Psicológica, 31*(2), 335-355.

Wang, H. P., Kuo, B. C., Tsai, Y. H., & Liao, C. H. (2012). A cefr-based computerized testing system for chinese proficiency. *The Turkish Journal of Educational Technology, 11*(4), 1-12.

Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods computerized adaptive testing. *Journal of Educational Measurement, 35*(2), 109-135.

Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: lssues and Practice, 17*, 17-27.

Way, W. D. (2005). *Practical questions in ıntroducing computerized adaptive testing for k-12 assessments* (PEM Research Report. 05-03). London: Pearson Educational Measurement.

Weiss, D. J. (1988). Adaptive testing. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 881-884). New York: Pergamon Press.

Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*(2), 70-84.

Weiss, D. J., & Guyer, R. (2012). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing.* St. Paul MN: Assessment Systems Corporation.

Weiss, D. J., & Von Minden, S. (2012). *A comparison of item parameter estimates from Xcalibre 4.1 and Bilog-MG*. St. Paul MN: Assessment Systems Corporation.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 52*(2), 275291. doi:10.1007/BF02294241

Yen, W. M. (1993). Scaling performance assessment: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187-213.

Yoes, M. (1995). *An updated comparison of microcomputer-based item parameter estimation procedures used with the 3-parameter IRT model.* Saint Paul, MN: Assessment Systems Corporation.

Zheng, Y., & Chang, H. H. (2014). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement, 39*(2), 105-118.

Zitny, P., Halama, P., Jelinek, M., & Kveton, P. (2012). Validity of cognitive ability tests-comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations*. Studia Psychologica, 54*(3), 181-194.