

Investigating Item Exposure Control Methods in Computerized Adaptive Testing*

Nagihan Boztunc Ozturk^a

Hacettepe University

Nuri Dogan^b

Hacettepe University

Abstract

This study aims to investigate the effects of item exposure control methods on measurement precision and on test security under various item selection methods and item pool characteristics. In this study, the Randomesque (with item group sizes of 5 and 10), Sympton-Hetter, and Fade-Away methods were used as item exposure control methods. Moreover, in order to establish a comparison baseline for other methods, the no-exposure control condition was also included in the research. As item selection methods, Maximum Fisher Information, a-Stratification, and Gradual Maximum Information Ratio were employed. While a parameters were generated from a uniform distribution ranging between 0.50 and 2.00 and c parameters were generated from a uniform distribution ranging between 0.05 and 0.20 for both item pools, b parameters were generated from a uniform distribution ranging between -3.00 and +3.00 for medium difficulty item pool, and from a standard normal $N(2, 1.5)$ distribution for high difficulty item pool. Based on the research findings, there were no great differences between the values of indicators used in determining measurement precision when the item exposure control methods were used. On the other hand, in terms of test security, it was found that in general, the Fade-Away Method for controlling item exposure yielded better results than did the other methods in reducing the skewness of item pool utilization and the test overlap.

Keywords: Computerized adaptive testing • Item exposure control methods • Randomesque method • Sympton-Hetter method • Fade-Away method • Measurement precision • Test security

* This study is based on a brief summary of the doctoral dissertation entitled "Investigation of Item Exposure Control Methods in Computerized Adaptive Testing" prepared in the Educational Measurement and Evaluation Program, Hacettepe University.

a Corresponding author

Nagihan Boztunc Ozturk, Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey

Research areas: Computerized adaptive testing

Email: nagihanboztunc@hacettepe.edu.tr

b Nuri Dogan, Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey

Email: nurid@hacettepe.edu.tr

The administering of Computerized Adaptive Testing (CAT) has increased, and continue to do so, in line with concurrent developments in both the fields of computer technology and psychometrics. In lieu of traditional paper-and-pencil tests, CAT style tests have become increasingly attractive because they are easy to apply and to score, there are items in the tests corresponding to examinees' ability levels, such tests can be administered whenever desired, and the tests are shorter when compared to traditional paper-and-pencil tests (Grist, Rudner & Wise, 1989; Meijer & Nering, 1999; Rudner, 1998; Weiss & Kingsbury, 1984). CAT applications are used in the administration of the GMAT (Graduate Management Admission Test) and the GRE (Graduate Record Examination). As in traditional paper-and-pencil tests, these applications, also need to be reliable and valid since they are used to make critical decisions regarding the futures of examinees. As the use of CAT have become increasingly widespread, so have a number of problems with the potential of endangering validity appeared. Accordingly, the issue of such tests' security increases in importance (Chang & Twu, 1998; Davey & Nering, 2002 as cited in Barrada, Olea, Ponsoda, & Abad, 2019; French & Thompson, 2003; Georgiadou, Triantafyllou, & Economide, 2007; Lee & Dodd, 2012).

Although CAT applications require item pools containing a large number of items, certain items are used more frequently than others in specific situations. In such situations, the probability of examinees' attempting simply to memorize the answers to frequently used items is higher. If such items are memorized and then shared, the test's validity becomes jeopardized (Georgiadou et al., 2007; Lee & Dodd, 2012).

Since developing an ideal item pool is the result of a long and laborious process, it is undesirable for test developers to use just a certain percentage of the item pool, instead it would be better to use the entire pool efficiently (Revuelta & Ponsoda, 1998). For the stated reasons, a number of methods were developed so as to assure test security and to make use of the item pool more efficiently. Such methods are called "item exposure control methods," which have gradually been included in the fundamental components of CAT due to the problems encountered in real-life applications (Boyd, 2003; Davis, 2002).

Newly developed item exposure control methods were added to the existing ones so as to prevent such issues as examinees' being repeatedly exposed to

the same items, the same items being administered to examinees of similar proficiency levels, a low percentage of the item pool being used, and the inclusion of frequently used items of the pool in diverse applications after being revealed.

The results of the newly developed methods should be compared with existing ones. The results to be yielded by the item exposure control methods (selection method, the psychometric properties of the items, the size of the item pool, the termination rule, and sampling distribution) in changeable factors during the administration of CAT remain an issue requiring research.

The most important aspect in CAT and in item exposure control is the item pool because whether a successful test may be conducted or not depends on the properties of the items within the pool. With this in mind, one of the factors that influences item exposure is the psychometric properties of the items in the item pool (Revuelta & Ponsoda, 1998). From a psychometric perspective, however, items are assessed based on the item information corresponding to their item parameters and to the ability levels. The desired distribution of the items along the ability scale varies according to the goal of the test. While the item pool for achievement tests should contain a variety of items ranging from "very easy" to "very difficult," the ideal distribution of the b parameter should be uniform distribution. For the criterion-referenced tests aiming to distinguish between examinees above and below a certain cut-off point on the other hand, most of the items in the item pool should be at the necessary difficulty level so as to provide the most information at roughly the cut-off point (Boyd, 2003). In that case, it is necessary to investigate the results of item exposure control methods used for those item pools having diverse psychometric properties.

While examining the working principles of the item exposure control methods, it has been found that although they work in unison with item selection methods, they also impose a number of restrictions on those methods. For instance, when item exposure is not controlled, an item may be used on the test whereas if a control method is used, that same item may not be selected for use. The probability of selecting that particular item may change according to the properties of the control method used. Afterwards, a different item is selected for use on the test. This process continues until the appropriate item to be administered to the examinee has been determined. In brief, when a method for controlling item exposure is used in

combination with different item selection methods, item pools will yield different results based on how they are used depending on the algorithm used in that particular item selection method (Han, 2012). In this case, variations in the item selection methods used for various CAT administrations has caused researchers to investigate under what conditions and to what extent item exposure is influenced.

While focusing on controlling item exposure, the measurement precision should not be ignored. Item exposure control methods must balance the test security with the measurement precision (Boyd, 2003; Boyd, Dodd, & Fitzpatrick, 2013). For these reasons, the precision of ability level estimations and the utilization of item pool should be considered together when comparing the methods for controlling item exposure.

This research seeks answers to the question: "How does the use of methods for controlling item exposure in computerized adaptive testing influence measurement precision and test security based on differing item selection methods and on the characteristics of diverse item pools?"

Based on the literature review, it was observed that diverse conditions were brought together for item exposure control methods (Burt, Kim, Davis, & Dodd, 2003; Chang & Twu, 1998; Chajewski & Lewis, 2009; Davis, 2002; Davis & Dodd, 2005; French & Thompson, 2003; Han, 2009, 2012; Lee & Dodd, 2012; Leroux, Lopez, Hembry, & Dodd 2013; Pastor, Dodd, & Chang, 2002; Revuelta & Ponsoda, 1998; Sanchez 2008). It was also seen in research studies that such conditions as b parameter being close to zero while selecting the first item, the use of Maximum Information method as the method of item selection, the use of Maximum Likelihood Estimation in ability estimation, and the inclusion of content balancing in the studies due to the fact that the parameters obtained from real data were used were frequently preferred.

In research studies aiming to test the item exposure control methods by modifying the properties of item pools (Chajewski & Lewis, 2009; Chang & Twu, 1998; Lee & Dodd, 2012; Leroux et al., 2013; Pastor et al., 2002; Sanchez, 2008), item pools of different sizes and characteristics were used. In the studies where item pools of differing sizes were used, the item pools were composed of dichotomously scored items (Chang & Twu, 1998; Leroux et al., 2013), and of polytomously scored items (Chajewski & Lewis, 2009; Pastor et al., 2002; Sanchez, 2008). In a study conducted by Chang and Twu (1998), it was found that item exposure

control methods were influenced by the size of item pools in different ways. Pastor et al. (2002) pointed out that item pool size was one of the two factors that should be considered when deciding on item exposure control methods. In a study by Lee and Dodd (2012), on the other hand, an item pool of polytomously scored items was employed. They found that by raising or lowering the item pool's difficulty level, the number of unused items in the Maximum Information Method and in the Randomesque method with an item group size of six had increased.

When item exposure control methods were reviewed, a number of conclusions were made: (1) random selection methods were generally less successful in dichotomously scored item pools than were other methods, (2) the Sympon-Hetter method and its derivations required laborious processes despite their success in the use of item pools, and (3) the a-Stratification item selection method yielded more successful results in controlling item exposure than did the other item selection methods.

When the related research studies were inspected, it was found (1) that the Randomesque method was generally used with item group sizes of 3 and 6, (2) that the methods for item exposure were not analyzed in light of item pools' different characteristics, with these pools being scored dichotomously, and (3) that the Fade-Away Method developed by Han (2009) had not been studied by other researchers apart from Han. Moreover, it was also found that in those studies investigating the Fade-Away Method, two factors' level of influence had not been investigated. These factors being (1) one test time slot per day and (2) controlling the item exposure in real time conditions. In addition, it was found that only Han (2009) had investigated to what extent using Gradual Maximum Information Ratio as the method of item selection influenced item exposure.

The number of studies conducted on CAT administrations in Turkey is quite limited, with existing studies focusing on whether or not there were any significant differences between the scores received on traditional paper-and-pencil tests and on the CAT administrations. They also compared the different starting rules, item selection methods, ability estimation methods, and test termination rules in terms of measurement precision (Eroglu, 2013; Kalender, 2011; Kaptan, 1993; Kezer, 2013; Sulak, 2013). A review of the research studies conducted in other countries, however, has shown

that the focus is more on the development of new methods for CAT components. Accordingly, new methods are recommended by researchers in relation to item selection and controlling item exposure (Finkelman, Nering, & Roussos, 2009; Han, 2009, 2012). In terms of giving information related to their use in CAT administrations, it is considered important to test those methods with types of items scored differently, in item pools having different psychometric properties, and in groups with different ability distributions. By investigating new methods, the current researchers therefore seek to perform CAT administrations with higher reliability and validity.

Item exposure control, has become one of the current issues discussed in relation to CAT according to International Association for Computerized Adaptive Testing (IACAT) (IACAT, n.d.) and has, as such, been recently included among its basic components. The sustainability of the of item pool utilization without causing significant differences in measurement precision is also among the goals of item exposure control. The development of item exposure control methods continues so as to regulate the utilization of item pools, to reduce item overlaps between tests, and to render CAT administrations more valid.

Considering the advantages of CAT administrations, the opportunities provided by advanced technology, and ever-increasing importance of time in human life, it may be said that the day when traditional paper-and-pencil tests will be replaced by CAT administrations is not very far in Turkey. Such a possibility brings to light the need for conducting studies on the components of CAT administrations in Turkey. The current research aims to analyze the effect of diverse item exposure control methods on measurement precision and on test security under various circumstances.

Method

This study aims to investigate the various item exposure control methods in CAT under the dichotomous IRT model by using different item pool characteristics and item selection methods and their effect on measurement precision and test security. Since this study aims to add innovative knowledge to the current theoretical knowledge, it carries the properties of a basic research study (Karasar, 2009).

Data Generation

In this study the SimulCAT simulation software tool (Han, 2011) was used both to generate data and for CAT simulation.

Sample Groups: Two different sample groups were used in this research. One was the sample used in the simulation of the CAT whereas the other was used to calculate the item exposure control parameters in the Sympon-Hetter item exposure control method.

CAT Sample: The ability distributions for the simulees in the sample group of the CAT were derived from the normal distribution ($N(0, 1)$), with a mean of 0 and a standard deviation of 1. The size of the sample was determined as 1,000.

Sympson-Hetter Sample: The necessary care was taken to ensure that the sample in the iterational simulation, in which the parameters for the item exposure control were obtained, had the same distribution as the sample used in CAT. Therefore, the ability distributions of the simulees in the sample group were derived from the normal distribution ($N(0, 1)$), with a mean of 0 and a standard deviation of 1. The size of the sample was determined as 10,000.

Item Pool Characteristics: Two item pools were used in this research. While the a and the c parameters of these item pools had the same distributions, the b parameter distributions differed. The first item pool was the pool with medium level of difficulty in which the b parameter had a uniform distribution so as to represent the achievement tests, whereas the second pool was the pool with high level of difficulty in which the b parameter was of normal distribution so as to represent the criterion-referenced tests.

Item Pool with Medium Difficulty Level: All item parameters used in the first item pool were determined as the a parameter in the $[0.50, 2.00]$ interval, the b parameter in the $[-3.00, 3.00]$ interval, and the c parameter in the $[0.05, 0.20]$ interval with uniform distribution. Five hundred (500) items were included in the item pool.

Item Pool with High Difficulty Level: While the b parameter of the items used in the second item pool was determined in such a manner so as to give it a mean of 2 and a standard deviation of 1.5 with normal distribution ($N(2, 1.5)$), the a parameter was determined in the $[0.50, 2.00]$ interval and the c parameter in the $[0.05, 0.20]$ interval. Both the a and c parameters had uniform distributions. Five hundred (500) items were included in the item pool.

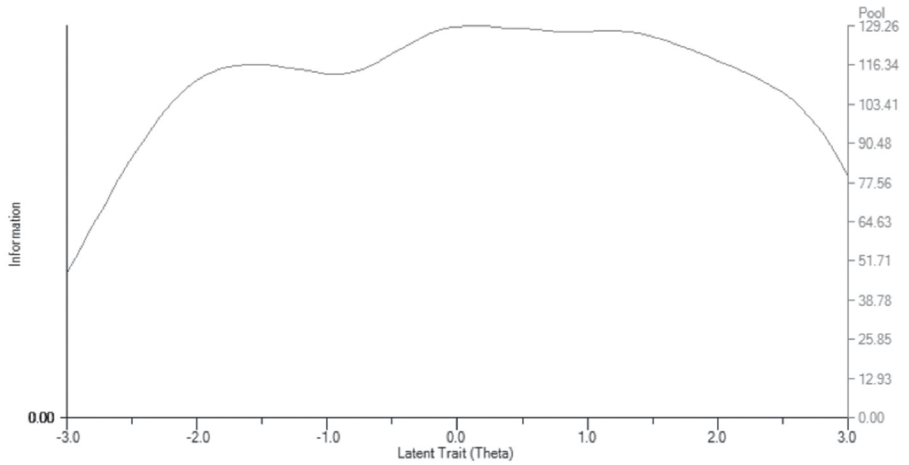


Figure 1: Information function of the item pool with medium difficulty level.

CAT Simulations

The CAT administration based on dichotomously scored items was simulated in this research. Therefore, the 3PLM — contend to be the most appropriate model for CAT as a model of IRT — was employed. For the selection of the first item, the simulees' initial theta estimate was set to zero. To estimate ability, the estimation method of Expected A Posteriori (EAP) was used. The fixed test length was chosen as a termination rule, and the length was set to 25. The simulation was designed to simulate 1,000 examinees attending a single CAT session during a single test time slot. The methods of Maximum Fisher Information (MFI), a-Stratification, and Gradual Maximum Information Ratio (GMIR) were used as item

selection methods. As item exposure control methods, the Randomesque method (with item group sizes of 5 and 10), the Sympson-Hetter method (with a target exposure rate of 0.20), and the Fade-Away method (with a target exposure rate of 0.20) were chosen. Harwell et al. (1996 as cited in Evans, 2010) recommends the use of at least 25 replications in order to reduce potential sample bias in simulations. Thus, 25 replications were used in the research. The mean was calculated for the results obtained and the analyses performed.

Item Exposure Control

Controlling item exposure prevents the same item from being over-used, and thus aids in preserving the

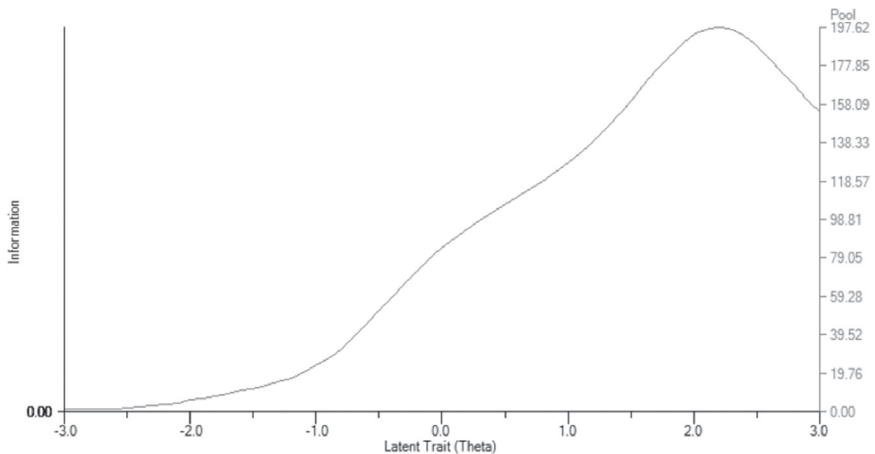


Figure 2: Information function of the item pool with high difficulty level.

integrity of the item pool and the confidentiality of the items within the item pool (Davis & Dodd, 2005).

In CAT procedures, in which different items are administered to examinees of different ability levels, items can be used at different frequencies depending on the relations between examinees' ability levels and the information structure within the item pool (IACAT, n.d.). Using item selection methods maximizes measurement precision due to the inverse relationship existing between information and the standard error of estimation. Yet, when the items are selected to maximize measurement precision, it is also observed that some items are administered to all examinees whereas others are not been administered at all, a situation which leads to uneven item exposure (Pastor et al., 2002). The results of the study conducted by Hulin, Drasgow, and Parsons in 1983 are a noteworthy example of this situation. In their study, Hulin et al. (1983 as cited in Revuelta & Ponsoda, 1998) found that 141 of the 260 items were not administered at all when the Maximum Information Method was the preferred item selection method.

According to Stocking and Lewis (2000), the relation between item exposure and measurement precision resembles a balloon. A balloon, when squeezed on one side, swells up on the other. Therefore, while controlling for item exposure, precision in ability estimation decreases. The purpose of using an efficient item exposure control method is two-fold: the first being to prevent this very relation and the second, to ensure that the item pool is used in a more balanced way without reducing measurement precision (Pastor et al., 2002).

The Item Exposure Control Methods: Although there are a variety of methods to control for item exposure, this research makes use of the Randomesque Method, the Sympton-Hetter Method, and the Fade-Away Method in its analyses.

The Randomesque Method (RA): In order to avoid the over-exposure of the items which gives maximum information in item selection, Kingsbury and Zara (1989) recommended to use the Randomesque Method. The Randomesque Method is based on randomly selecting items from a group of items (ranging from 2 items to 10 items) which are composed of most information items rather than selecting the single most informative for a simulee's current theta estimate (Davis & Dodd, 2005; Han, 2011; Kingsbury & Zara, 1989; Lee & Dodd, 2012). In this method, an item is randomly selected for administering from the group of items and then sends the remaining items back to the pool. The

process continues until the CAT is completed (Davis & Dodd, 2005; Macken-Ruiz, 2008).

The Sympton-Hetter (SH) Method: Developed by Sympton and Hetter (1985), in this method item selection algorithm differentiates the probability $P(A)$ that an item will be administered from the probability $P(S)$ that the item will be selected. In order to maintain the $P(A)$ at the desired target level, the $P(A/S)$ is obtained in the target $P(A)$ which is derived from the iterative simulations (Han, 2011). This method involves a 2-phase implementation. In the first phase control parameters for item exposure are obtained, and in the second phase 2 the actual CAT procedure starts. The parameters for controlling item exposure obtained in phase 1 are used in phase 2 in order to control the item exposure (Davis & Dodd, 2005). In the SH method, computed item exposure parameters are pool-specific, meaning that control parameters should be re-computed whenever a change occurs in the item pool (Han, 2011).

The Fade-Away Method (FAM): Developed by Han (2011), this method, weights the item selection criterion for each item eligible in the item pool against the proportion between that item's updated and target exposure. With the implementation of this method, rarely used items are expected to be used more often, and excessively used items are expected to "fade away" (Han, 2011).

Data Analysis

After the CAT simulation, the data were analyzed so as to evaluate both measurement precision and test security. Therefore, in order to evaluate measurement precision, the fidelity coefficient, RMSE, bias, and average absolute difference were calculated for each condition. Furthermore, in order to evaluate test security, the following values were calculated: item exposure rates, means for item exposure, standard deviations for item exposure, maximum item exposure rates, the percentages of items with a zero-exposure rate, scaled chi-square statistics, and test overlap rates.

All the calculations other than the calculation of test overlap were done on Excel, a software built for calculation tables. Test overlap rates were calculated using MATLAB programming language.

Evaluation of Measurement Precision: The fidelity, RMSE, bias, and average absolute difference values were calculated separately for 25 replications. The arithmetic means for the results obtained are shown in the results section.

Fidelity: The correlations of simulated estimates with generated values, in simulation studies are often reported as “fidelity” (Urry, 1970; Vale & Weiss, 1975 as cited in Leung, Chang, & Hau, 2002). A high fidelity coefficient is used both to make more reliable decisions and to determine better item selection methods (Leung et al., 2002). Pearson’s Product Moments Correlation was employed in calculating the fidelity coefficient.

$$RMSE: RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}}$$

$$Bias: Bias = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)}{n}$$

Average absolute difference (AAD):

$$AAD = \frac{\sum_{i=1}^n |\hat{\theta}_i - \theta_i|}{n}$$

This formula calculates the RMSE, bias, and average absolute difference values, in which:

$\hat{\theta}_i$ = represents the estimated level of ability for person *i*,

θ_i = represents the known level of ability for person *i*, and

n = represents the size of the sample (Boyd et al., 2013; Davis, 2002).

Evaluation of Test Security:

Item Exposure Rate: The rate of exposure for an item is found by dividing the number of times that the item is administered to examinees by the total number of examinees taking the test. When *m* represents the number of examinees, the observed exposure rate (*ko_j*) is calculated as in the following equation for item *j*:

$$ko_j = \frac{\text{number of times the } j\text{th item is used}}{m}$$

The exposure rates for each of the 25 replications for every item in the item pool were calculated, and the arithmetic means for the values were calculated. Both the distribution of the mean item exposure and the number of items used not even once during all 25 replications represented by *ko* = 0, are presented in the results section. Items whose rate of mean exposure was other than 0 were reported in categorical groups of 0.10. The maximum exposure rate was determined by selecting the highest item exposure rate from mean item exposure rates.

Item exposure rates’ means and standard deviations for each of the 25 replications were calculated. The arithmetic means were found for these indicators.

Scaled Chi-square Statistics: In order to determine the equality of the item exposure rates, a target distribution is necessary. Having the same rate of exposure and a uniform distribution is ideal for all items. Thus, the desirable uniform rate for all items is calculated as $\overline{ko_j} = L/N$, where *N* represents the size of item pool and *L* represents the length of test. Mimicking Pearson’s *x*² statistics, the *x*² was regulated as in the following to measure the similarity of the observed and the expected exposure rates:

$$x^2 = \sum_{j=1}^N \frac{(ko_j - \overline{ko_j})^2}{\overline{ko_j}}$$

The equation shows differences between observed and expected item exposure rates. One of the primary goals of an item exposure control method is to make that all the items in the item pool are used in the best possible way. If a method has a low *x*² value, it may be said that most of the items in an item pool have been used (Chang & Ying, 1999).

The *x*² values for each of the 25 replications were calculated in this study. The arithmetic mean for those values is shown in the results section.

Test Overlap Rates: Another important brief index for controlling item exposure is test overlap rate. The index is found by dividing the expected number of overlapping items for two randomly selected examinees (*m*) by the test length (*L*). Overlap rate is calculated by following the steps: (1) the number of common items for each of the *m(m-1)/2* pairs of examinees is counted, (2) each *m(m-1)/2* calculation is totaled, and (3) the total number is divided by *Lm(m-1)/2*.

Ideally, the number of items shared in the tests taken by two randomly chosen examinees should be at a minimum (Chang & Ying, 1999). Higher rates of item overlap indicate that item exposure rates are excessively skewed. If each item in the pool has an equal probability of being selected, the number of common items seen by two randomly selected examinees will be at a minimum (Chang & Zhang, 2002). The lowest expected limit of test overlap rate for a test of fixed length is indicated by the proportion *L/N*, where *L* is the length of the test and *N* is the size of the item pool (Chang & Zhang, 2002).

The overlap rates for each method in the 25 replications were calculated in this study. The arithmetic means for the values are presented in the results section.

Results

Research findings are shown below according to each item pool's characteristics.

Results of the Item Pool with Medium Difficulty Level

The mean values for the fidelity coefficient, RMSE, bias, and average absolute difference were used to evaluate measurement precision. They are shown in Table 1 for the no-exposure control condition and for each method of controlling item exposure.

According to Table 1, the values for the fidelity coefficient range between 0.9755 and 0.9854. The lowest fidelity coefficient was obtained when the a-Stratification method and the Randomesque method with an item group size of ten were used together. Accordingly, the highest RMSE value observed was 0.2250 when the a-Stratification and the Randomesque-10 were used together, and the lowest value was 0.1744 when the Gradual Maximum Information Ratio was used with no exposure control. Although the bias were quite close to zero in every condition, the highest value was obtained when using of Maximum Fisher Information and the Sympson-Hetter method in combination. When the average absolute difference index was investigated, it was found that the lowest average absolute difference for between the actual and estimated ability levels was 0.1346 with no exposure control through the use of the Gradual Maximum Information Ratio. The highest index was 0.1748 when the a-Stratification was used to select items in tandem with using the Fade-Away method to control for item exposure.

The distributions of mean item exposure rates, mean standard deviations for those rates, the maximum mean rate of item exposure, the percentages of items not administered in any of the replications, the mean χ^2 value for the usage of items, and the mean rate of test overlaps were used to evaluate test security. These values are shown in Table 2 for the cases in which no exposure control condition and for each method used to control item exposure.

As it is seen from Table 2, the number of the items not used in any replications of CAT procedure ranges between 0 (when using the a-Stratification method and the Fade-Away method) and 293 (when the Maximum Fisher Information method was used with no exposure control). Accordingly, the value is 58.6% of the size of the item pool when no exposure control through the Maximum Fisher Information method is used. With the combined use of the a-Stratification method and the Fade-Away method, however, each item was used at least once in at least one replication of the CAT procedure. Considering the fact that the target rate of exposure was set to 0.20 with the Sympson-Hetter and the Fade-Away methods, it can be seen that by using the Maximum Fisher Information and the Sympson-Hetter methods together on one hand and the a-Stratification method along with the Fade-Away method on the other, the target exposure rate was exceeded.

When the standard deviations for the purposes of checking the item exposure rates were examined, it was found that the highest value of standard deviation (0.110) appeared when item exposure was not controlled (when Maximum Fisher Information method was used) whereas the lowest value (0.049)

Table 1
Indicators used to Evaluate the Measurement Precision in the Item Pool of Medium Difficulty Level

Item Selection Method	Exposure Control Conditions	Fidelity	RMSE	Bias	AAD
Maximum Fisher Information	No-exposure control	0.9853	0.1748	-0.0026	0.1348
	Randomesque - 5	0.9851	0.1758	-0.0028	0.1356
	Randomesque - 10	0.9845	0.1793	-0.0025	0.1380
	Sympson-Hetter	0.9826	0.1900	-0.0037	0.1469
	Fade-Away	0.9823	0.1918	-0.0013	0.1486
a-Stratification	No-exposure control	0.9769	0.2186	-0.0022	0.1696
	Randomesque - 5	0.9765	0.2209	-0.0029	0.1712
	Randomesque - 10	0.9755	0.2250	-0.0019	0.1740
	Sympson-Hetter	0.9769	0.2188	-0.0012	0.1698
	Fade-Away	0.9756	0.2248	-0.0016	0.1748
Gradual Maximum Information Ratio	No-exposure control	0.9854	0.1744	-0.0024	0.1346
	Randomesque - 5	0.9851	0.1763	-0.0034	0.1354
	Randomesque - 10	0.9845	0.1797	-0.0028	0.1381
	Sympson-Hetter	0.9830	0.1880	-0.0023	0.1453
	Fade-Away	0.9823	0.1919	-0.0029	0.1486

*All of the values were obtained by calculating the mean for the data resulting from the 25 replications.

Table 2
Indicators used to Evaluate Test Security in the Item Pool with Medium Difficulty Level

Exposure Rate (ko)	Maximum Fisher Information					a-Stratification					Gradual Maximum Information Ratio				
	NEC	RA5	RA10	SH	FAM	NEC	RA5	RA10	SH	FAM	NEC	RA5	RA10	SH	FAM
0.9 < ko ≤ 1.0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
0.8 < ko ≤ 0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.7 < ko ≤ 0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6 < ko ≤ 0.7	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
0.5 < ko ≤ 0.6	1	5	0	0	0	0	0	0	0	0	1	1	0	0	0
0.4 < ko ≤ 0.5	7	6	8	0	0	2	0	0	0	0	7	8	3	0	0
0.3 < ko ≤ 0.4	14	15	16	0	0	1	2	0	0	0	16	18	20	0	0
0.2 < ko ≤ 0.3	25	24	21	1	0	6	3	9	0	3	23	22	27	0	0
0.1 < ko ≤ 0.2	41	36	44	121	134	83	109	108	106	87	40	39	44	122	133
0.0 < ko ≤ 0.1	117	144	169	128	176	300	338	341	326	410	119	142	165	135	199
ko = 0	293	270	242	250	190	106	48	42	68	0	292	270	241	243	168
SD _{ko}	0.110	0.105	0.099	0.076	0.060	0.080	0.055	0.053	0.059	0.049	0.109	0.102	0.094	0.075	0.059
maximum ko	1.0	0.565	0.469	0.201	0.173	1.0	0.304	0.290	0.197	0.295	1.0	0.509	0.438	0.199	0.174
% (ko = 0)	58.6	54.0	48.4	50.0	38.0	21.2	9.6	8.4	13.6	0	58.4	54.0	48.2	48.6	33.6
χ ²	120.569	110.246	98.229	57.647	35.814	63.698	30.619	27.947	34.285	23.865	119.777	104.533	88.720	56.913	35.193
Test overlap rate	0.290	0.270	0.246	0.164	0.121	0.177	0.110	0.105	0.118	0.097	0.289	0.258	0.227	0.163	0.120

* All of the values were obtained by calculating the mean for the data resulting from the 25 replications.
 ** The means of item exposure rates were found to be 0.05 in all conditions
 ***NEC: No-exposure control, RA-5: Randomesque with an item group size of five, RA-10: Randomesque with an item group size of ten, SH: Sympton-Hetter, FAM: Fade-Away

was observed when the a-Stratification method was used in combination with the Fade-Away method. Considering that χ^2 analysis results provide general information on item pool use, it was found that by using the a-Stratification method in combination with the Fade-Away method, the item pool ($\chi_2 = 23.865$) was used the closest to the ideal when compared to all other conditions.

In terms of the highest item exposure rate, it was found that the item exposure was 1.0 for all item selection methods when item exposure was not controlled. This indicator obtained its lowest value of 0.173 when the Maximum Fisher Information method was used in combination with the Fade-Away method.

An examination of test overlap rates demonstrated that the highest rate of overlap (0.290) was observed when item exposure was not controlled (when the Maximum Fisher Information method was used) whereas the lowest rate of overlap was observed when the a-Stratification method was used in combination with the Fade-Away method (0.097). The expected rate of overlap in the research was 0.05 (25/500). Yet, none of the methods of controlling item exposure attained this value.

Results of the Item Pool with High Difficulty Level

The mean values for the fidelity coefficient, RMSE, bias, and the average absolute difference, are used in the evaluation of measurement precision. They are shown in Table 3 for the no-exposure control condition and for each method of controlling item exposure.

As is clear from Table 3, the values for the fidelity coefficient range between 0.9584 and 0.9775. The lowest fidelity coefficient was obtained when the a-Stratification method and the Fade-Away method were used together. The highest RMSE value was 0.2919 (when the a-Stratification and the Fade-Away methods were used together) while its lowest valued was 0.2161 (when the Gradual Maximum Information Ratio was used with no exposure control). Although the bias were quite close to zero in every condition, the highest value (-0.0044) was obtained when the a-Stratification method was used in combinations with the Sympton-Hetter method. On investigating the average absolute difference index, it is clear that the lowest average absolute difference for the actual and the estimated ability levels was 0.1623 when Gradual Maximum Information Ratio was used with no exposure control whereas the highest index was 0.2164 when the a-Stratification method was used to select items in combination with the Fade-Away method being used to control item exposure.

The distributions of mean item exposure rates, mean standard deviations for those rates, the maximum mean rate of item exposure, the percentages of items not administered in any of the replications, the mean χ^2 value for the usage of items, and the mean rate of test overlaps were used to evaluate test security These values are shown in Table 4 for the cases in which no exposure control condition and for each method used to control item exposure.

According to Table 4, the number of the items not used in any replications of the CAT procedure

Table 3
Indicators used to Evaluate the Measurement Precision in the Item Pool with High Difficulty Level

Item Selection Method	Exposure Control Conditions	Fidelity	RMSE	Bias	AAD
Maximum Fisher Information	No-exposure control	0.9773	0.2168	-0.0008	0.1627
	Randomesque - 5	0.9769	0.2185	-0.0020	0.1642
	Randomesque - 10	0.9756	0.2247	-0.0010	0.1679
	Sympson-Hetter	0.9637	0.2734	0.0007	0.2040
	Fade-Away	0.9693	0.2518	-0.0009	0.1902
a-Stratification	No-exposure control	0.9713	0.2436	-0.0010	0.1862
	Randomesque - 5	0.9686	0.2546	-0.0019	0.1930
	Randomesque - 10	0.9645	0.2704	-0.0028	0.2043
	Sympson-Hetter	0.9614	0.2816	-0.0044	0.2128
	Fade-Away	0.9584	0.2919	-0.0019	0.2164
Gradual Maximum Information Ratio	No-exposure control	0.9775	0.2161	-0.0013	0.1623
	Randomesque - 5	0.9766	0.2202	-0.0028	0.1653
	Randomesque - 10	0.9757	0.2243	-0.0019	0.1684
	Sympson-Hetter	0.9649	0.2688	-0.0005	0.2006
	Fade-Away	0.9689	0.2533	-0.0018	0.1904

* All of the values were obtained by calculating the mean for the data resulting from the 25 replications.

ranges between 0 (when the a-Stratification method was used in combination with the Fade-Away method) and 321 (when the Gradual Maximum Information Ratio was used with no exposure control). Accordingly, when the Gradual Maximum Information Ratio was used with no exposure control, the value was found to be 64.2% of the size of the item pool. When the a-Stratification method was used in combination with the Fade-Away method, however, each item was used at least once in at least one replication of the CAT procedure. Considering that the target rate of exposure was set to 0.20 with the Sympson-Hetter and the Fade-Away methods, it is clear that using the

a-Stratification method along with the Fade-Away method, the target exposure rate was exceeded.

Considering the standard deviations for the purposes of checking item exposure rates, it was found that the highest standard deviation value (0.123) was observed when item exposure was not controlled when the Maximum Fisher Information was used whereas the lowest value (0.062) appeared when the a-Stratification method was used in combination with the Fade-Away method. In terms of the χ^2 analysis results, which give general information on item pool usage, it was found that by using the a-Stratification method in combination with the Fade-Away method, the use

Table 4
Indicators used to Evaluate Test Security in the Item Pool with High Difficulty Level

Exposure Rate (ko)	Maximum Fisher Information					a-Stratification					Gradual Maximum Information Ratio				
	NEC	RA-5	RA-10	SH	FAM	NEC	RA-5	RA-10	SH	FAM	NEC	RA-5	RA-10	SH	FAM
0.9 < ko ≤ 1.0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
0.8 < ko ≤ 0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.7 < ko ≤ 0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6 < ko ≤ 0.7	1	2	0	0	0	0	0	0	0	0	1	0	0	0	0
0.5 < ko ≤ 0.6	4	8	6	0	0	1	2	0	0	1	3	4	0	0	0
0.4 < ko ≤ 0.5	12	8	10	0	0	10	3	0	0	0	13	15	14	0	0
0.3 < ko ≤ 0.4	17	15	19	0	0	14	17	15	0	2	16	12	23	0	0
0.2 < ko ≤ 0.3	15	16	15	0	0	14	18	35	0	10	19	23	19	0	0
0.1 < ko ≤ 0.2	25	29	29	129	128	45	50	45	122	72	24	25	25	128	124
0.0 < ko ≤ 0.1	106	116	132	103	138	210	264	267	225	415	102	117	131	99	142
ko = 0	319	306	289	268	234	205	146	138	153	0	321	304	288	273	234
SD _{ko}	0.123	0.119	0.115	0.078	0.070	0.107	0.094	0.088	0.076	0.062	0.121	0.115	0.109	0.079	0.070
maximum ko	1.0	0.652	0.565	0.197	0.195	1.0	0.511	0.393	0.199	0.518	1.0	0.565	0.475	0.198	0.196
% (ko = 0)	63.8	61.2	57.8	53.6	46.8	41	29.2	27.6	30.6	0	64.2	60.8	57.6	54.6	46.8
χ^2	150.734	141.826	131.645	61.215	48.810	114.271	88.565	78.081	58.191	38.808	146.733	132.397	117.927	61.647	49.482
Test overlap rate	0.351	0.333	0.313	0.172	0.147	0.278	0.226	0.205	0.166	0.127	0.343	0.314	0.285	0.172	0.148

*All of the values were obtained by calculating the mean for the data resulting from the 25 replications.

** The means of item exposure rates were found to be 0.05 in all conditions

***NEC: No-exposure control, RA-5: Randomesque with an item group size of five, RA-10: Randomesque with an item group size of ten, SH: Sympson-Hetter, FAM: Fade-Away

of the item pool ($\chi^2 = 38.808$) was the closest to the ideal, when compared to the other cases.

When examining the highest item exposure rate, item exposure was found to be 1.0 for all item selection methods in which item exposure was not controlled. This indicator obtained its lowest value of 0.195 when it was used Maximum Fisher Information in combination with the Fade-Away method.

An examination of test overlap rates demonstrated that the highest rate of overlap was observed in cases where the Maximum Fisher Information method was used with no exposure control (0.351), whereas the lowest rate of overlap was observed in cases where the a-Stratification method was used in combination with the Fade-Away method (0.127). Although expected rate of overlap in the research was 0.05 (25/500), none of the methods for controlling item exposure attained this value.

Discussion

A general review of the research findings shows that the performances displayed by the methods for controlling item exposure in item pools of different difficulty levels are very similar. The most important difference between them is that the number of items $k_0 = 0$ in the item pool with high difficulty level is bigger. In addition, it was also found that the values which indicated measurement precision were better in the item pools with medium level of difficulty; which was expected due to the fact that a sample derived from the normal distribution $N(0,1)$ was used in the CAT procedures. Therefore, the item pool with medium level of difficulty better suited the sample. Moreover, it was also found when investigating the item pool with high difficulty level that the unused items were in general those of higher difficulty, the b parameter means of those items ranging between 2.50 and 3.55.

According to the literature review, no studies were encountered which investigated changes in the performance of different methods for controlling item exposure in item pools of differing characteristics which were also scored dichotomously. In a study by Lee & Dodd (2012) where polytomously scored items were available and where the methods for controlling item exposure in item pools with differing levels of difficulty were compared, three different item pools (easy, medium, and hard) and two different ability distributions (negatively skewed and normally distributed) were employed. As a result, it was found that the medium item pool yielded

better results in terms of measurement precision, maximum item exposure rates, and the usage of the item pool regardless of examinees' ability distribution. In addition, when the group that was normally distributed was investigated, it was found that the number of unused items increased when using the Maximum Information method and the Randomesque method with an item group size of six as the difficulty level of the item pool used increased. The current study was conducted on item pools of polytomously scored items. A literature review revealed that the methods for controlling item exposure yielded differing results in terms of item exposure in dichotomously scored and in polytomously scored items (Boyd, 2003; Boyd et al., Dodd, & Choi, 2010 as cited in Lee & Dodd, 2012). Despite this, when the item pool with high difficulty level was used, the results were found to be similar to the results of previous studies, in that the number of unused items increased for some of the methods.

According to the research findings, using the Fade-Away method in combination with each item selection method resulted in the item pool being used more effectively. Han (2009) studied different methods of item selection and item exposure control methods for an item pool in which the information function was the most intensive ($\theta = 1$). As a consequence, he found that the item pool was better benefited from when the Gradual Maximum Information Ratio method was used in combination with the Fade-Away method as opposed to using the Maximum Fisher Information method in combination with the Fade-Away method. In the current study, however, when the MFI and the GMIR methods were used in combination with the Fade-Away method (FAM), similar results were observed, indicating that the findings of the two studies differ.

When used the medium difficulty item pool, the fidelity coefficients of the MFI and the a-Stratification methods were used with no exposure control, the SH and the FAM methods show results similar to those obtained by Han (2012). In terms of item pool usage, while similar results were obtained for test overlap, differing results were obtained for target exposure. In Han's study (2012), in both conditions where the MFI and the a-Stratification were used, the SH method exceeded target exposure but the Fade-Away method did not. In the current study, however, target exposure was exceeded when the Fade-Away method was used in combination with the a-Stratification method. The SH method met

target exposure in all of the conditions except when combined with the MFI.

By comparing the condition in which the item exposure was not controlled with the a-Stratification method, it was found that the a-Stratification method in combination with the Sympon-Hetter method was a better choice for item pools of medium and high levels of difficulty in terms of test security. While the absence of large differences in measurement precision; specifically, it provided a more efficient usage of the item pool, fitted the target exposure, and reduced the test overlaps. This finding is also supported by Leung et al. (2002).

In terms of the medium difficulty item pool level findings similar to those of Leroux et al. (2013) were found; these being, the conditions in which the Maximum Fisher Information method was used with the no exposure control, the RA-5, and the SH method, the changes of the fidelity and the RMSE (the highest RMSE MFI+SH, the lowest RMSE MFI + no exposure control condition) and the change of the number of unused items (the lowest MFI+SH, the highest MFI + no exposure control condition) in the CAT procedure.

In the case where item exposure was not controlled in both item pools, the a-Stratification method yielded the best results for item pool usage and in preventing test overlaps. The Maximum Fisher Information and the Gradual Maximum Information Ratio methods yielded very similar results. In Chang and Ying (1999), the a-Stratification method yielded better results for preventing test overlaps and item pool usage compared to the Maximum Fisher Information method.

Recommendations

According to the research findings, it may be recommended that to ensure test security, item exposure control methods be used in CAT administrations to assure test security and that the a-Stratification method be used as the item selection method in item pools composed of both medium and high difficulty level for test security, if a method for item exposure is not used. It is recommended that the Fade-Away method be used, if a method for controlling item exposure is to be used. It is recommended that the a-Stratification method be used for item selection in combination with the Fade-Away method for controlling item exposure when the item pool's difficulty level is medium or high. It is also recommended that the Fade-Away method be preferred over the Sympon-Hetter method.

In relation to this research, it may be recommended that a similar study be performed using different item selection methods and different item exposure control methods, that a similar study be conducted using item pools by changing the a and c parameters of the item pool psychometric characteristics, that the item exposure control methods be compared to item pools of different sizes, that the item exposure control methods be compared using the data obtained by real-life CAT applications, and that the item exposure control be analyzed in applications in which content balance is included.

References

- Barrada, J., Olea, J., Ponsada, V., & Abad, F. (2009). Test overlap rate and item exposure rate as indicators of test security in CATs. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved April 13, 2013 from www.psych.umn.edu/psylabs/CATCentral/
- 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. doi: 10.1080/08957347.2013.765434
- Boyd, M. A. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No.3110732)
- 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.
- Chajewski, M., & Lewis, C. (2009). Optimizing item exposure control algorithms for polytomous computerized adaptive tests with restricted item banks. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved April 13, 2013 from www.psych.umn.edu/psylabs/CATCentral/
- Chang, H.-H., & Ying, Z. (1999). a-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23, 211–222. doi: 10.1177/01466219922031338
- Chang, H. H., & Zhang, J. (2002). Hypergeometric family and item overlap rates in computerized adaptive testing. *Psychometrika*, 67(3), 387–398.
- 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.
- Davis, L. L. (2002). *Strategies for controlling item exposure in computerized adaptive testing with polytomously scored items* (Doctoral dissertation, The University of Texas at Austin). Retrieved from <http://repositories.lib.utexas.edu/handle/2152/1701?show=full>
- 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.
- Eroğlu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması* (Doctoral dissertation, Hacettepe University, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Evans, J. J. (2010). *Comparability of examinee proficiency scores on computer adaptive tests using real and simulated data* (Doctoral dissertation, The State University of New Jersey). Retrieved from <https://rucore.libraries.rutgers.edu/rutgers-lib/26525/>
- Finkelman, M., Nering, M. L., & Roussos, L. A. (2009). A conditional exposure control method for multidimensional adaptive testing. *Journal of Educational Measurement*, 46(1), 84–103. doi: 10.1111/j.1745-3984.2009.01070.x
- French, B. F., & Thompson, T. D. (2003, April). *The evaluation of exposure control procedures for an operational CAT*. Poster presented at the annual meeting of the American Educational Research Association. Chicago, IL.
- Georgiadou, E., Triantafyllou, 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.
- Grist, S., Rudner, L. M., & Wise, L. L. (1989). *Computerized adaptive tests*. Washington, DC: American Institutes for Research.
- 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 April 24, 2013 from www.psych.umn.edu/psylabs/CATCentral/
- Han, K. T. (2011). *User's manual: SimulCAT*. Retrieved March 01, 2013 from http://www.umass.edu/rempp/software/simcata/simulcat/SimulCAT_Manual.pdf
- Han, K. T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement*, 49(3), 225–246. doi: 10.1111/j.1745-3984.2012.00173.x
- International Association for Computerized Adaptive Testing. (n.d.). *Some current issues in CAT*. Retrieved March 15, 2013 from <http://iacat.org/node/491>
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability* (Doctoral dissertation, Middle East Technical University, The Graduate School of Natural And Applied Sciences, Ankara). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kağıt kalem testi uygulamasının karşılaştırılması* (Doctoral dissertation, Hacettepe University, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Karasar, N. (2009). *Bilimsel araştırma yöntemi*. Ankara: Nobel.
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması* (Doctoral dissertation, Ankara University, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2, 359–375.
- 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. doi: 10.1177/0013164411411296
- Leung, C.-K., Chang, H.-H., & Hau, K.-T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the sympson-hetter algorithm. *Applied Psychological Measurement*, 38(7), 376–392. doi: 10.1177/014662102237795
- 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(5), 857–874. doi: 10.1177/001316441348680
- Macken-Ruiz, C. L. (2008). *A comparison of multi-stage and computerized adaptive tests based on the generalized partial credit model* (Doctoral dissertation, The University of Texas at Austin). Retrieved from <http://repositories.lib.utexas.edu/bitstream/handle/2152/17878/mackenruizc.pdf?sequence=2>
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, 23(3), 187–194. doi: 10.1177/01466219922031310

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. doi: 10.1177/01421602026002003

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

Rudner, L. M. (1998). *An on-line, interactive, computer adaptive testing tutorial*. Retrieved from <http://echo.edres.org:8080/scripts/cat>

Sanchez, E. I. (2008). *A comparison of item exposure control procedures with the generalized partial credit model* (Doctoral dissertation, the University of Texas at Austin). Retrieved from <http://repositories.lib.utexas.edu/bitstream/handle/2152/6818/sancheze37838.pdf?sequence=2>

Stocking, M. L., & Lewis, C. (2000). Methods of controlling the exposure of items in CAT. In W. J. van der Linden, & G. A. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 163-182). Boston: Kluwer.

Sulak, S. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında kullanılan madde seçme yöntemlerinin karşılaştırılması* (Doctoral dissertation, Hacettepe University, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375. doi: 10.1111/j.1745-3984.1984.tb01040.x