

Development of a Computerized Adaptive Version of the Turkish Driving Licence Exam

Nukhet Cikrikci^{1,*}, Seher Yalcin^{2,*}, Ilker Kalender³, Emrah Gul⁴,
Cansu Ayan², Gizem Uyumaz⁵, Merve Sahin-Kursad², Omer Kamis²

¹Faculty of Education, İstanbul Aydın University, İstanbul, Turkey

²Faculty of Education, Ankara University, Ankara, Turkey

³Faculty of Education, Bilkent University, Ankara, Turkey

⁴Faculty of Education, Hakkari University, Hakkari, Turkey

⁵Faculty of Education, Giresun University, Giresun, Turkey

ARTICLE HISTORY

Received: Apr. 09, 2020

Revised: Sep. 3, 2020

Accepted: Sep. 19, 2020

KEYWORDS

Computerized adaptive testing,

Driving licence exam,

Hybrid simulation

Abstract: This study tested the applicability of the theoretical Examination for Candidates of Driving License (ECODL) in Turkey as a computerized adaptive test (CAT). Firstly, various simulation conditions were tested for the live CAT through an item response theory-based calibrated item bank. The application of the simulated CAT was based on data from e-exams administered by the Ministry of National Education (MoNE). Results of the first stage of the study were used to determine the rules for starting, continuing, and terminating the live CAT exam for ECODL. Secondly, the live CAT exam was applied according to the results of the simulation. Candidate drivers (n = 280) who had taken the ECODL as an e-test participated in the second stage. Thirdly, the opinions of the individuals who took the computer-based test towards the computer-based testing application were mapped. In the termination rule of the CAT-based ECODL, testing with a fixed number of questions yielded the smallest estimated measurement error. We also found that when ECODL was implemented as CAT, it could reliably differentiate among testers in terms of competence of theoretical knowledge of driving and provide basis for accurate decisions regarding their proficiency. According to the findings obtained on the candidates' opinions on the computer-based testing application, it was seen that they considered computer-based application more practical an easier in terms of testing.

1. INTRODUCTION

In today's world, training of drivers is of crucial importance because traffic safety depends primarily on driver's skills besides several other factors. Driving a vehicle is a complex process in which cognitive, affective and psychomotor skills are involved (Young, Regan & Hammer, 2007). Drivers who have not adequate skills may cause traffic accidents with death and or severe injuries or damage to public properties. Furthermore, such traffic accidents affect not

*CONTACT: Nükhet ÇIKRIKÇI ✉ nukhetdemirtasli@aydin.edu.tr 📧 İstanbul Aydın University, Faculty of Education, İstanbul, Turkey and Seher YALÇIN, yalcins@ankara.edu.tr, Ankara University, Faculty of Education, Ankara, Turkey.

only parties involving the accidents; it also affects pedestrians, families, social life, economy of the country (Mulkat-Ajibola, 2015).

This is why being an adult is a legal prerequisite for being eligible to drive a motor vehicle (Miser, 1999). Mental health condition and physical abilities are also considered. Any individual who meets these criteria may apply for driving license examination. Passing this examination provides individuals with an official certificate which entitles people to drive.

In general, driving license examinations involve assessment of two traits: (i) theoretical aspects regarding driving (i.e. traffic rules, traffic signs, first-aid, etc.) and (ii) practice (i.e. driving a vehicle). If different examples in the world are considered, this common pattern can be observed. For example, in England, first phase of driving license examination has two sections (Bozkurt, 2003). One of the sections include multiple-choice items about traffic signs, rules and regulations, etc., whereas the other section is a hazard perception test. This test measures examinees' reactions for dangerous situation that can be faced while driving. Multiple-choices test is given as computerized, while perception test is administered by watching some scenarios. Germany gives examinees a computerized test including items about traffic signs, rules, etc. Examinees are also given several items specially designed for the type of vehicle they want to drive (AngloInfo the Global Expat Network Berlin, 2015). Item are randomly selected from an item pool and administered in a computerized medium. Australia uses a similar testing procedure. Items are selected from a pool (AngloInfo the Global Expat Network Berlin, 2015). Finland and Belgium have also similar procedures (Ministry of Transport and Communications, 2015). In all of these countries, completing training programs of is a prerequisite to take the driving license examination.

1.1. Turkey Case

In Turkey, causes of traffic accidents can be grouped into three: problems regarding (i) individuals (as pedestrians, passengers, and drivers), (ii) road and (iii) vehicles. Figures announced by the state agencies indicate that 86.2% of the traffic accidents were caused by drivers. The rest is shared by passengers (0.47%), pedestrians (9.38%), roads (0.95%), and vehicles (0.58%) (Turkish Statistical Institute, 2018).

As figures point out, drivers are the leading factor in traffic accidents. Turkish governments have been implementing some policies regarding traffic accidents. These policies can be group into three: (i) security measures including penalties, (ii) training to increase security level of pedestrians and drivers, and (iii) selection of individuals who are fully eligible to drive.

Although there are several rules and regulations regarding the first two groups of policies implemented by Turkish bodies, the last item (selection procedures) seems to be neglected. For example, a common finding indicated by the literature is that content, length, and objectives of the training programs of learner drivers are not enough and appropriate (Balkız, 1999; Çakır, 2006; İnal, 2001; Kuyumcu, 2001; Tanrikulu, 2002; Türkoğlu, 2002; Tütüncü, 2001; Vursavaş, 2004). Another finding reported in the literature is that training of learner drivers provided a significant increase in drivers' knowledge of driving (Bozkurt, 2003; Çakır, 2006; Güleçen, 1998; Kaçmaz-Omak, 2012; Türkoğlu, 2002). Elander, West and French (1993) found that risk-taking behaviour of drivers have a direct relationship between traffic accidents. Despite the large body of research in the literature regarding driving licenses, there is no study examining the selection procedure of drivers in Turkey. However, certification process of individuals who are eligible to drive is as important as other components such as rules, regulations, and quality of training programs.

1.2. The Turkish Examination for Candidates of Driving License (ECODL)

In Turkey, all candidates should complete a training program including both theoretical and practical aspects of driving before applying for the driving license exam. These training

programs are offered by private institutions certified by Turkish authorities. The training program includes sixty-three hours for theory and 20 hours of practice: 42% traffic and environment, 15% first-aid, 24% driving, 19% motor and vehicle techniques and driving technique (Çakır, 2006). Those who complete the training program successful become eligible to take the driving license examination.

The Examination for Candidates of Driving License (ECODL), the standard driving exam in Turkey, has also been offered as a computer-based exam, or e-exam, since 2011 in addition to paper-and-pencil tests. The first part of the driving license examination includes tests measuring different theoretical aspects of the driving. ECODL, driving licence exam has three sections: (i) first aid, (ii) traffic and environment, and (iii) motor and vehicle technique. The items in the bank were included in multiple-choice questions, each with four options.

The ECODL theoretical examination in Turkey currently administered by MoNE has some problems, such as test items being used without a prior pilot study, an excessive number of items testing knowledge of Bloom's taxonomy, limitations with respect to measurement techniques, and the questions being leaked to the public beforehand. Nearly identical questions are used in successive ECODL test sessions. Furthermore, although offered as e-examinations, these computer-based tests are restricted in terms exploiting the benefits of the technology (for example, item assignment suitable to the level of capability of the candidate).

1.3. Development of Computerized Adaptive Version of ECODL

By considering these issues, this study develops a valid and reliable ECODL test that can reliably reflect candidates' proficiency in a computerized environment by using the advantages of this technological development. Accordingly, it examines the applicability of the theoretical examination of ECODL as a computerized adaptive test (CAT). In the present study, development of computerized adaptive version The Turkish ECODL was examined. Some examples of e-exams in various disciplines across the world are the GMAT (Graduate Management Admission Test), GRE (Graduate Record Examination), MAP (Measures of Academic Progress), NCLEX (National Council Licensure Examinations), ASVAB (Armed Services Vocational Aptitude Test Battery), and STAR (Math, Reading, and Early Literacy) (Economides & Roupas, 2007).

The aim of computerized adaptive tests (CATs) is to effectively predict the candidate's proficiency using shorter tests than paper-and-pencil exams (Weiss, 2004). CAT-based tests are more reliable and useful as they match the difficulty of questions with the ability/competence of the test taker. Items in the test are chosen to provide the most information about the candidate's ability (Orcutt, 2002; Weiss & Kingsbury, 1984). A significant benefit of this is that it allows the use of an item bank with estimated parameter values according to the IRT (Item Response Theory), which defines capability parameters and item parameters on the same scale (theta scale). Therefore, it can predict the level of difficulty of questions that a respondent can correctly answer based on his/her ability predicted using the answers to previous questions (Weiss, 2004).

The greatest advantage of CAT is that it provides functional testing through uncompromising adherence to the validity and reliability of measurement. In other words, it provides a difficulty-based question display for candidates depending on their previous responses. In any CAT application, if a candidate correctly answers a question with a mean difficulty, which is shown on the screen, the computer displays a more challenging question next. If the candidate responds incorrectly, the system chooses an easier item next. Thus, respondents with high proficiency levels are assigned more difficult items whereas those with lower levels are presented easier items (Davey, 2011; DeMars, 2010; Weiss, 1982). As a result, the test ensures the optimal question assignment that can best reveal the candidate's proficiency according to performance.

This case allows respondents to answer shorter tests in less time, without the need to assign to them questions that are too difficult, and thus above the proficiency, or too easy items or below their proficiency (Linden & Glas, 2002; Wainer & Mislevy, 1990; Weiss & Kingsbury, 1984). This enhances the usefulness of CAT.

Another advantage of CAT is the low likelihood of cheating on it. That is, the probability of cheating is eliminated because every respondent is given a test consisting of different questions in accordance with his/her proficiency, which enhances the security of CAT (Thompson & Weiss, 2011). Moreover, this ensures that the test results are generated more quickly (Linacre, 2006; Orcutt, 2002; Rudner, 1998; Weiss, 2004; Weiss & Kingsbury, 1984). CAT also has some limitations: the necessity of a large item bank; the respondent's inability to go back to a previous question, check it, and change answers; and test bias due to familiarity with computers (Bugbee & Bernt, 1990; Economides & Roupas, 2007; Sutton, 1991).

Apart from the well-accepted advantages of CAT application, it still has some discussable disadvantages. For instance, it is claimed that reading a text from a computer screen takes longer than reading from a printed material. Moreover, solving mathematics questions is possible to be problematic since the test takers will not be able to underline the text, which is a strategy while answering an item in the test (Bugbee & Bernt, 1990). Not being able to go back to the questions and to check, to edit or to change them during the implementation is also discussed to be a limitation (Linacre, 2000). Another highlighted problem is the concern for a possible increase in bias in tests. It is stated that computer familiarity (hence some variables such as socio-economic status, or gender) may affect the test results as a source of bias. It is thought that this may be an extra source of anxiety for the ones who are not familiar with computers (Linacre, 2000; Sutton, 1991). On the other hand, while this situation would be more problematic in the past, it is expected to decrease today since the computer and smart phone technologies have entered into our lives at such a pace.

The first set of studies on CAT applications in Turkey started with the assessment of academic achievement in science and mathematics (Kaptan, 1993; Köklü, 1990; Yaşar, 1999). The findings showed that there was no significant difference in the academic achievement of students as reflected in CAT and traditional paper-and-pencil tests. In numerous subsequent studies, the applicability of CAT to academic tests for transition to secondary and higher-education examinations was explored (İşeri, 2002; Kalender, 2011). From the findings, it was clear that the examinations can be applied as CATs on the condition that the item bank was enriched with a larger number of specific questions.

Subsequent studies on CAT examined the applicability of different proficiency tests that were not administered by the federal government but were taken by large groups (academic achievement, foreign language proficiency, computer literacy) according to different CAT strategies (Gökçe, 2012; Kezer & Koç, 2014; Özbaşı & Demirtaşlı, 2015). Some studies were conducted in this vein in medicine (Öztuna, 2012), and that long surveys given to patients for accurate diagnoses can be employed as CATs. These studies encouraged the development of CAT software (Kalender, 2011). In general, the above-mentioned common findings show that a variety of tests in different domains can be offered as CATs. A sufficiently large item bank is also needed such that it can ensure the reliability of tests, consisting of specific questions while addressing a wide range of capabilities. Regarding such a popular testing application whose both advantages and disadvantages are discussed, it is considered to be important to get the opinions of the test takers right after a live CAT application in order to be able to research if it is possible to switch into this application, to make suggestions for a possible implementation and to develop methods regarding the precautions to be taken. Powers and O'Neill (1993) stated that the opinions of the study group changed positively and greatly after taking the real application and mentioned the importance of taking opinions after the implementation.

1.4. Research Questions

- (1) What are the methods of estimating capability, and what is the termination rule with the minimum error according to the simulation for a live CAT application?
- (2) As a result of the application of the live CAT under the selected ideal conditions,
 - a. What was the testing time per person?
 - b. What were the distributions of standard error of the test information values of estimation, proficiency level, and estimated proficiency level?
- (3) What are the opinions of the individuals who took the ECODL exam on the computer-based application?

2. METHOD

As stated earlier, this study describes the development of computerized adaptive version of The Turkish ECODL. The development of a CAT-based ECODL starting by creating an item bank for first aid, motor and vehicle technique, and traffic and environment. Items were pre-tested before including into the item bank. Then several CAT simulations were run in order to determine optimal CAT design before live CAT. Then, a group of learner drivers were participated to live CAT version of ECODL. Participants were also asked to report their opinions about the live CAT stage. The rest of this section details the methodology of the study.

2.1. Development of the Item Bank

In the item writing phase, academician who have expertise in item development, trainers of learner drivers, and staff from Ministry of National Education worked in collaboration. A teacher who teaches traffic gave a training for item developers. Separate test plans for each subject were prepared prior to the item writing process to specify the basic and advanced driver proficiencies at three cognitive levels (knowledge, comprehension, and problem solving). The number and distribution of the items employed for measurement were also specified in the plan. The items were written by the authors using a checklist. Each item writer sent his/her items to field experts who independently checked them (item fit to measurement, grammar, spelling, punctuation, and scientific accuracy). Scientific accuracy of items inspected by teachers who traffic, teach first-aid and health classes. In addition, a drawer was employed to draw visuals in the items. Items were revised more than once based on suggestions made by the parties involved. Item development phase completed between November 2017 and March 2017. At the end, a total of 913 multiple-choice items with 4 alternatives were compiled to be included in the pilot study.

2.2. Pilot Study for the Items

The item bank was pilot-studied on a sample of 1787 university students from different cities of Turkey. Also, between June and August in 2017, items were pilot-studies on learner drivers when they took computerized (not adaptive) ECODL. Test-takers were given regular items as well as items developed for the CAT version. At that stage, 90% of the item bank were tested by around 250 learner drivers.

Another phase of close inspection was conducted. Items were once again revised scientific accuracy, measurement quality, grammar, clarity. Some items which included large visuals could not the displayed on the computer screen properly. These items were excluded from the item bank. Finally, a total of 892 items (417 for traffic and environment, 270 for first-aid and 192 for motor and vehicle technique) were kept in the item bank.

2.3. Item Calibration

In a period of 8 weeks, the 892 items were used for learner drivers who took actual driver license examination. Some of items were not responded by sufficient number of learner drivers

(at least 200 responses for each item). As a result, 235 items for traffic and environment, 201 for first-aid and, 192 items for motor and vehicle techniques were left in the item bank.

Analyses revealed that 3PL model had the best item-model fit. Some of the items in the item bank indicated poor item fit and excluded from the study. 193 items for traffic and environment, 139 for first-aid and, 167 items for motor and vehicle techniques were kept. [Table 1](#) presents the items parameters for each domain.

Table 1. *Descriptive of the domains in the item bank.*

	Traffic and Environment			First-Aid			Motor and Vehicle Techniques		
	a	b	c	a	b	c	a	b	c
Min	0.30	-4.00	0.10	0.30	-4.00	0.18	0.30	-4.00	0.17
Max	1.08	4.00	0.38	1.18	4.00	0.30	1.88	4.00	0.43
Mean	0.51	0.90	0.24	0.48	1.35	0.25	0.57	0.35	0.25
Median	0.45	1.00	0.25	0.45	2.28	0.20	0.54	0.32	0.25

2.4. Simulations Prior to Live CAT

Prior to live CAT administration, as suggested in the literature (Thompson & Weiss, 2011), a series of post-hoc simulations were conducted to define optimal live CAT design. At that phase of the project, different CAT strategies were employed. Starting rule and item selection procedure were kept constant. At the beginning of each simulation, examinees were assigned an ability score of 0. Items were selected using Fisher’s method (Weiss, 1982). On the other hand, two ability estimation methods were used: Bayesian Expected a Posteriori and Maximum Likelihood Estimation. Also, two different test termination methods were defined: a fixed number of items and SE-threshold. For SE-based test termination, SEs of 0.30 and 0.40 were used as threshold. Fixed number of items were defined as 50, number of items given in ECODL.

At that stage, it was noted that since not every item in the item bank developed for CAT was administered during the pilot study, there were missing values in the item-person matrix (Weiss & Guyer, 2012). As a solution, the hybrid simulation method was adopted (Nydick & Weiss, 2009). In the hybrid simulation approach, the available item set answered by every respondent was used to predict the ability of each respondent on all questions that were not answered by him/her. Estimation was then employed to assign responses to the omitted and unanswered items using a suitable IRT model and the Monte Carlo simulation. An item response matrix with all data for every respondent that could be employed in a post-hoc simulation was thus obtained (Weiss & Guyer, 2012). Thus, six different CAT (three test termination rules x two ability estimation methods) strategies were simulated.

2.5. Live CAT

The live CAT test was performed with an item bank of a total of 628 items. A computer software developed by Kalender (2011) was used to conduct live CAT. The software was installed 11 laptop computers. The live CAT administrations, which were applied between September and November 2017, were conducted just after the learner drivers took ECODL in testing centers. Learner drivers were invited to voluntarily participate to the live CAT sessions. A small gift (a first-aid kit) was given to each participant. After the purpose of the study was explained and the consent were received, the participants were administrated a live CAT session.

A total of 280 learner drivers (age mean: 21 years, s.d.: 5.19) were participated in this study. The descriptive statistics of the participants are given in [Table 2](#). As can be seen, most of the participants are at least high school graduates. More than 85% of the participants have experience with computers more than 5 years. Almost 90% of the participants preferred ECODL exam in computerized format instead of paper-and-pencil format.

Table 2. Demographic information of live CAT participants.

Variable	Value	N	%	Total
Gender	Female	121	43.2	280
	Male	159	56.8	
Educational background	Primary school	5	1.8	280
	Secondary school	8	2.9	
	High school	89	31.8	
	Undergraduate	168	60.0	
	Postgraduate	10	3.6	
Any experience with computers	None	5	1.8	280
	1-2 Years	11	3.9	
	3-5 Years	22	7.9	
	More	242	86.4	
Any experience with computerized tests	Yes	118	42.1	279
	No	161	57.5	
Preferred test format	Computerized	250	89.3	279
	Paper-and-pencil	29	10.4	

2.6. Opinions after the Live CAT

Just after the live CAT administrations, participants were also given a questionnaire to report their opinions. The questionnaire consisted of two parts. The first part included some questions on participants' demographic information (gender, education level) and their experience with computers (computer usage experience, computer-based testing experience). The second part included 18 items inquiring their opinions towards the comparison of computerized tests and pen-and-paper tests and the techniques and usefulness of computerized tests. Each item was presented as an opinion statement and the participants were asked to mark their agreement level among the suitable reflective categories which ranged from *Completely agree* (1) to *Completely disagree* (4).

2.7. Data Analysis

Data analysis was carried out in three phases. In the first, IRT-based analysis of each item included in the pilot study was performed to determine ones that were in the item bank and predict their related parameters. Before IRT analyzes, the unidimensionality assumption was examined separately for each booklet on the basis of every subject area. In this context, exploratory factor analysis based on the tetrachoric correlation matrix was done with the R programming. It has been observed that there was one dominant dimension in all subject areas. The provision of the unidimensionality assumption shows that the local independence assumption is also met (Hambleton & Swaminathan, 1985). Also, the test completion time of the participants shows that the test was not a speed test. The responses of the candidate drivers were analyzed according to the IRT scaling with best model–data fit. As a result of the item analyses carried out using the 3PL model that showed the best model–data fit, the difficulty of the questions/items varied in the range (-4.00 and +4.00) and item discrimination in the range (0.30 and 1.88). The probability of correct response by those with low capability was estimated at 0.25. Questions with the expected level of measurement quality were selected and included in the item bank according to the test subjects.

For sub-questions a and b, in accordance with the second research sub-question of the study, descriptive statistics of the participants' response times, the final proficiency estimation, standard error of estimation, and test information were calculated. The final proficiency and the

standard error in measurement obtained from the CAT application were analyzed to find the distribution of proficiency estimation and values representing the reliability of the estimation. Test information for each respondent obtained from the overall ECODL test and the total test information value for all participants in the live CAT application for each ECODL were calculated. In IRT-based test development, reliability is the amount of test information at a certain theta/proficiency level (Hambleton, Swaminathan & Rogers, 1991). For the third sub-question, the opinions of the group were made suitable to interpret on the item level by taking the frequencies and percentages in the reflective categories for each item.

3. RESULT

According to the results of the simulative CAT of ECODL analyses in accordance with the first research sub-question, the estimated ability of all test subjects yielded the smallest error when the EAP method was employed. With regard to the rule for test termination, the application with fixed questions was found to yield the smallest estimated error. The initial level of theta was set to zero in all conditions. The results of an analysis of the simulation conditions of the CAT are presented in Table 3.

Table 3. Analysis of tested conditions in CAT simulation on three ECODL subjects.

Test Subject	Condition No.	Estimated Ability on Method	Test Termination Rule	Item Selection Method	Mean Measurement Error of the Item Bank	Mean Measurement Error of Proficiency Estimations	Mean Number of Items (Minimum and Maximum Numbers of Questions)
Traffic and Environment	1	2	Se (0.30)	1	0.887	0.887	190 (190–190)
	2	3	Se (0.30)	1	0.396	0.396	187.967 (44–190)
	3	2	Fixed question (23)	1	0.885	1.253	23 (23–23)
	4	3	Fixed number of items 23)	1	0.396	0.487	23 (23–23)
	5	2	Se (0.40)	1	0.891	0.891	190 (190–190)
	6	3	Se (0.40)	1	0.396	0.410	123.932 (28–190)
Motor and Vehicle Technique	1	2	Se (0.30)	1	1.010	1.010	139 (139–139)
	2	3	Se (0.30)	1	0.477	0.477	139 (139–139)
	3	2	Fixed question (12)	1	1.017	1.592	12 (12–12)
	4	3	Fixed number of items (12)	1	0.477	0.606	12 (12–12)
	5	2	Se (0.40)	1	1.018	1.018	139 (139–139)
	6	3	Se (0.40)	1	0.477	0.477	138 (37–139)
First Aid	1	2	Se (0.30)	1	0.745	0.745	170 (170. 170)
	2	3	Se (0.30)	1	0.328	0.339	130.703 (31. 170)
	3	2	Fixed question (9)	1	0.735	1.488	9 (9. 9)

(Continued)

Table 3. Continued

Test Subject	Condition No.	Estimated Ability on Method	Test Termination Rule	Item Selection Method	Mean Measurement Error of the Item Bank	Mean Measurement Error of Proficiency Estimations	Mean Number of Items (Minimum and Maximum Numbers of Questions)
First Aid	4	3	Fixed number of items (9)	1	0.328	0.553	9 (9. 9)
	5	2	Se (0.40)	1	0.744	0.744	170 (170. 170)
	6	3	Se (0.40)	1	0.329	0.398	37.403 (18. 170)

Note: 1: Item selection method based on maximum information, 2: Maximum likelihood estimation, proficiency estimation method, 3: Expected a posteriori estimation, proficiency estimation method

The CAT results in [Table 3](#) show that the best strategy in the item bank for all three test subjects was condition 4. Accordingly, when questions in live CAT were given, the test needed to be terminated with a fixed number of questions. It was thus decided that a fixed number of questions, parallel to questions used by MoNE, would be best: 23 for traffic and environment, 12 for first aid, and nine for motor and vehicle technique. When the EAP method was used for proficiency estimation, the result was reliable. In this case, the item bank of all three test subjects was arranged according to these strategies to be employed in live CAT. Descriptive statistics of the participants' response times in the live CAT performed under ideal conditions, defined in question "a" of the second sub-question of the research, are presented in [Table 4](#).

Table 4. Descriptive statistics of participants' response times in live CAT application of ECODL.

Test subject	Response Time (min)	
	Mean	Median
Traffic and Environment	7.97	7.85
First Aid	4.97	4.87
Motor and Vehicle Technique	2.11	2.05

[Table 4](#) shows that the mean response time ranged from two to eight minutes according to test subject. Although the ECODL as a Computerized Adaptive Test (ECODL_CAT) application had a fixed number of questions, it was practical in terms of time taken for application. The distributions of the predicted proficiency levels, test information levels, and standard errors according to the live CAT application performed under the ideal conditions defined in question "b" of the second sub-question of the research were examined. At the end of the ECODL tests, the proficiency level of each participant (θ) and standard error of measurement (Sem) values of estimated proficiency were obtained. As a result of the IRT-based CAT application of ECODL, test information values calculated in light of the participants' predicted proficiency levels (values) and the measured standard errors in estimated proficiency were considered the criteria of reliability (Embretson & Reise, 2000; Hambleton et al., 1991), and were obtained for each proficiency level. The levels of θ of the participants' estimated proficiency, and the related estimated errors and distribution of test information are summarized in [Table 5](#).

Table 5. Descriptive statistics of distributions of test information values, estimated proficiency, and errors in estimated proficiency of participants in live CAT application of ECODL.

Statistics	Distribution of estimated proficiency (θ) of the participants	Distribution of test information values ($I_i(\theta)$) of estimated proficiency of participants	Distribution of measured standard errors (Sem) in estimated proficiency of participants
Mean	0.38	6.31	0.40
Median	0.52	6.47	0.39
Minimum and Maximum	-3.00–2.96	3.16–8.21	0.35–0.56
Standard Deviation	1.28	1.00	0.03

Table 5 shows that individual proficiency levels in the ECODL_CAT application were moderate in terms of theta, in the range of $-4.00 < \Theta < +4.00$ with a standard deviation of 1.00 and an IRT theta scale mean of zero. The value of theta of participants was scattered over a wide range around and at the moderate level of ability (mean: 0.38; range: -3.00 and 2.96). Theta was distributed in two extremes and over a wide range. Respondents who showed a more homogeneous distribution of scores (min: 56, max: 98, mean: 81.65, median: 82, sd: 7.34) in the ECODL e-examination conducted by MoNE, displayed a more heterogeneous distribution of ability in the CAT application. This might have obtained because the item bank used in the ECODL_CAT application consisted of items at different levels of difficulty and a sufficient number of parameter values to discriminate among them. Items in the bank had good psychometric quality, and were employed after being tested on the candidates for the driver’s exam. Thus, an ECODL_CAT application based on these items provides a valid and reliable measurement of individual differences. The measured distribution of standard error (Sem) of the participants’ estimated proficiency ranged from 0.35 to 0.56. The calculated test information values ranged from 3.16 to 8.21. In the scope of the third sub-question of the research, after grouping the opinions given on the questionnaire items according to their common points, the results were summarized in Table 6.

Table 6. Frequency and percentages about questionnaire items.

Items	Statistic	Response Categories				Total	
		I Completely agree	I Agree	I don’t agree	I never agree		
Opinions about comparison of CAT and paper-pencil applications	i1	Frequency	174	76	23	6	279
		%	62.4	27.2	8.2	2.2	100.0
i2	Frequency	157	75	36	10	278	
	%	56.5	27.0	12.9	3.6	100.0	
i5	Frequency	161	82	18	17	278	
	%	57.9	29.5	6.5	6.1	100.0	
i9	Frequency	46	43	87	90	266	
	%	17.3	16.2	32.7	33.8	100.0	
i10	Frequency	143	112	15	4	274	
	%	52.2	40.9	5.5	1.5	100.0	
i16	Frequency	199	56	14	5	274	
	%	72.6	20.4	5.1	1.8	100.0	

(continued)

Table 6. Continued

Items	Statistic	Response Categories				Total	
		I Completely agree	I Agree	I don't agree	I never agree		
Opinions about technical application properties of CAT application	i3	Frequency	186	78	10	4	278
		%	66.9	28.1	3.6	1.4	100.0
	i4	Frequency	107	95	44	31	277
		%	38.6	34.3	15.9	11.2	100.0
	i12	Frequency	169	98	5	2	274
		%	61.7	35.8	1.8	.7	100.0
	i13	Frequency	23	9	71	171	274
		%	8.4	3.3	25.9	62.4	100.0
	i14	Frequency	109	100	38	24	271
		%	40.2	36.9	14.0	8.9	100.0
Opinions about visual properties	i6	Frequency	200	55	10	7	272
		%	73.5	20.2	3.7	2.6	100.0
	i8	Frequency	138	97	32	6	273
		%	50.5	35.5	11.7	2.2	100.0
	i18	Frequency	12	31	105	125	273
		%	4.4	11.4	38.5	45.8	100.0
Opinions about content of the items used in CAT application	i7	Frequency	98	108	61	7	274
		%	35.8	39.4	22.3	2.6	100.0
	i11	Frequency	152	104	14	2	272
		%	55.9	38.2	5.1	.7	100.0
	i15	Frequency	48	82	98	41	269
		%	17.8	30.5	36.4	15.2	100.0
	i17	Frequency	59	102	89	23	273
		%	21.6	37.4	32.6	8.4	100.0

As it is seen in Table 6, more than half of the participants marked 'completely agree' on the majority of the items. In line with this, the participants preferred computer-based application at a high rate in terms of marking the answers both faster and more easily for the items that compared the computer-based application they experienced within the project and pen-and-paper tests. In frame of the items which inquired the opinions towards the technical properties of CAT application, most of the participants demanded the possibility to go back and check their previous answers. Apart from this, they stated that they understood the instructions at beginning of the test and that they did not have any problems reading using computers.

The participants have positive opinions towards the visual properties of the computer-based application in the project. According to this, seeing only one question on the screen, the visual quality, the screen brightness and the resolution were evaluated positively in terms of its relation with the participants' performance on answering. Regarding the content of the items used in CAT application, almost half of them considered it to be easy as they could answer the questions showing up on the screen while almost half of them found the content difficult. Similarly, half of the participants thought the questions were easier than those of MoNE. In addition, the majority of the participants (almost 80%) had in the opinion that the questions were clear and understandable.

4. DISCUSSION and CONCLUSION

The purpose of this study was to develop a valid and reliable application for electronic examinations for driver's licenses. Accordingly, the applicability of CAT to the ECOLL theoretical examination was tested and verified. In this context, the ideal conditions for live

CAT were defined through a simulation. This study can serve as basis for further research to improve and upgrade examinations for driver's license.

The results of simulations showed that when the number of questions in the MoNE ECODL was considered fixed with a termination rule, the CAT application yielded estimation proficiency with a smaller error. When the EAP method was used for proficiency estimation, reliable results were obtained. Some studies that have employed CATs have also concluded that the EAP method gives better results than other techniques (Eroğlu, 2013; Kezer, 2013).

According to the findings, the total time taken by respondents to answer the entire test was approximately 15 minutes on average. In this respect, although the ECODL_CAT had a fixed number of questions in this study, it was practical in terms of time taken to complete the test. Because ECODL_CAT considered the level of difficulty of questions in terms of individual proficiency, candidates did not have to deal with questions above or below their levels of proficiency. This is an indication of the practicability of live CAT, a major advantage of adaptive tests (Kimura, 2017; Thissen & Mislevy, 2000; Thompson & Weiss, 2011).

The mean and median values of the standard measured error for the ECODL_CAT application was within the limits (0.50–0.30) suggested and accepted in the literature, even though the CAT application with a fixed number of questions was used (Embretson & Reise, 2000; Linacre, 2006). The standard error in measurement in this case was expected to range from 0.50 to 0.40, provided that 15–30 questions had been posed, to obtain the probability of correct responses (achievement). An error of 0.50–0.70 was obtained. The results show that the ECODL_CAT theoretical test measures differences in drivers' proficiency in theoretical knowledge more reliably than the ECODL e-test application of MoNE.

The bank consisting of items with a good model–data fit according to IRT provided the questions used in the CAT application and the predicted proficiency parameters (obtaining fixed items and proficiency estimations) (Hambleton et al., 1991). Moreover, given that, it is compatible with CAT applications; IRT-based item bank development enhances the reliability and practicability of the ECODL_CAT (Thissen & Mislevy, 2000; Thompson & Weiss, 2011). However, written based only on expert views and test subjects, items in the MoNE e-examination application have been used without any pilot study. No objective proof of the difficulty of the questions or their ability to discriminate among candidates' proficiency has thus been provided.

The results here have shown that ECODL can reliably measure (by ensuring adequate test information) candidate drivers' proficiency even though the CAT application had a fixed number of questions. They also show that when ECODL was applied as a CAT, it reliably predicted individual ability levels. Most of the participants stated that they considered the computer-based application more practical and easier in terms of testing. These findings are consistent with the findings of the research done by Schmidt, Urry & Gugel (1978), in which the opinions of 163 participants who took part in the adaptive test pilot study in American Civil Service Commission were taken. On the other hand, the participants stated that they did not experience any technical difficulties. This case contradicts the finding in the study done by Schmidt et al. (1978). One of the topics that the participants complained about in their study was technical details such as screen light; however, this is thought to stem from the fact that the study was a very old one. The possibility to experience such a problem with today's technological opportunities is low. Georgiadou, Triantafillou & Economides (2006), in their study in which they determined the standards for CAT application, stated technical properties and visual design to be one of the dimensions they cared. They said that one of the important points regarding CAT application was designing the screen so clear that the test-takers could understand it on their own and they identified the situation as "Using CAT should be so simple,

clear and easy that the test-taker can focus on the question's correct answer rather than how the system works" (p. 268).

The participants stated that the item contents were clear and understandable in terms of visual properties and content. Moreover, the numbers of ones who claimed that they encountered easier and more difficult questions than they could do are quite close to each other, which means there is no apparent agglomeration of a particular view. Although it is stated that CAT provides the individuals to encounter the questions which are consistent with their skills as a working principle in theory, Linacre (2000) states that this situation creates different situations in terms of the perception of the individuals with high performance. Highly skilled individuals are accustomed to a 90% correct response rate in paper-and-pencil applications. However, in CAT application, the more correct answers an individual gives, the more difficult questions he or she encounters and the success of the that person is determined jointly by the person's performance with the item selection algorithm. This may cause the successful individuals to be surprised during the test.

When the study group's experience with computers is examined, it is seen that the majority of the group is a computer user for more than 3-5 years. Computer familiarity is a variable that has been studied in various aspects of CAT applications in the literature. When the literature is examined, it is possible to come across findings that state that there is a relationship between computer familiarity and attitude towards computer (Powers & O'Neill, 1993; Wilder, Mackie, & Cooper, 1985) and the relationship between familiarity and attitude towards computer testing (Burke, Normand & Raju, 1987). Similarly, in this study, it was found that the majority of the study group (86.4%) had a high computer familiarity and likewise, the majority of the group (89.3%) preferred to take the test in the computer environment. Similarly, the majority of the study group (67%) stated that they did not experience any anxiety about computerized testing. In some studies (Kernan & Howard, 1990; Powers & O'Neill, 1993), while there was a relationship between familiarity and computer anxiety, Kim and McLean (1995) stated that individuals' participation in paper-and-pencil form or CAT application did not make a significant difference in terms of anxiety towards testing.

It is possible to say that the most basic point that individuals are not satisfied with is the lack of opportunity to review their answers. The majority of the group (95%) stated that they would prefer to go back and check their answers. This finding is consistent with the study conducted by Schmidt et al. (1978). On the other hand, Vispoel (1998, 2000) worked on two groups in which the students were allowed and were not allowed to return to their answers in the computer-based exam in their studies and examined the answer patterns. As a result of both studies, a very small percentage (3-5%) of the students changed their answers. In addition to this, it was stated that there was no significant difference between the groups who changed their answers and those who did not.

This research has certain limitations. Because the items were randomly assigned to respondents in the e-examination administered by MoNE, the 377 new items that were generated were not responded to by a sufficient number of candidates and thus were not included in the analysis, although the initial number of items in the pilot study was higher (876 items). In future studies on this subject, necessary precautions must be taken to ensure that every candidate driver is given new items at a certain rate in the e-exam. Another limitation is that the items in this study were multiple-choice questions and fitted the 3PL IRT model well. In future work, open-ended items with restricted responses can also be used in ECODL tests in addition to multiple-choice questions. In this case, psychometric results can be tested by sampling these items, and the precision of capability estimation between multiple-choice questions and mixed ECODL CAT test questions can be compared.

In this study, item exposure was not considered in the e-ECODL applied as live CAT. Item exposure represents the number of times a given item is used in all CAT applications (Georgiadou, Triantafillou & Economides, 2007; Revuelta & Ponsoda, 1998). It is associated with the validity of test scores and is intended to prevent changes in level of difficulty of questions in case test takers are familiar with a few. Future work can examine the effects of the incorporated use of item exposure with other termination strategies by considering different values of the reliability of proficiency estimation.

The MoNE carries out ECODL e-exam applications in 31 different halls. According to the findings obtained in the study, it is recommended that the MoNE should increase the number of halls and use other advantages provided by the computerized examination. The most important of these advantages is the inclusion of questions that determine the risk-taking disposition of individuals. It has been concluded that measuring this skill with internet or computer-based tests instead of traditional paper-and-pencil tests reduces both costs and increases the validity (Horswill & Coster, 2001; McKenna & Horswill, 1999). It is recommended to increase the size and quality of the ECODL item pool, to ensure the protection of confidentiality of items, and to carry out more studies on the transition to CAT application. Taking into consideration the opinions of individuals regarding CAT application, it is thought that CAT application can be started in ECODL theoretical exam with the arrangements made based on these research findings.

Acknowledgements

This research (Project # 215k018) was supported by the TUBITAK Project 1001 program in Turkey. The first draft of the paper presented at 6th International Congress on Measurement and Evaluation in Education and Psychology, Prizren, Kosova.

Declaration of Conflicting Interests and Ethics

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

ORCID

Nükhet ÇIKRIKÇI  <https://orcid.org/0000-0001-8853-4733>

Seher YALÇIN  <https://orcid.org/0000-0003-0177-6727>

İlker KALENDER  <https://orcid.org/0000-0003-1282-4149>

Emrah GÜL  <https://orcid.org/0000-0001-8799-3356>

Cansu AYAN  <https://orcid.org/0000-0002-0773-5486>

Gizem UYUMAZ  <https://orcid.org/0000-0003-0792-2289>

Merve Şahin KÜRŞAD  <https://orcid.org/0000-0002-6591-0705>

Ömer KAMIŞ  <https://orcid.org/0000-0003-0605-087X>

5. REFERENCES

- AngloInfo, The Global Expat Network Berlin. (2015). *Taking a German driving test*. Retrieved May 24, 2017, from <http://berlin.angloinfo.com/information/transport/driving-licences/the-driving-test>
- Balkız, C. (1999). *The effects of traffic education on the prevention of traffic accidents*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.
- Bozkurt, M (2003). *Evaluation of the first aid knowledge of driving nominees at the beginning and at the end of driving course in Ankara*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.

- Bugbee, A. C. & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982–1988. *Journal of Research on Computing in Education*, 23(1), 87-100. <https://doi.org/10.1080/08886504.1990.10781945>
- Burke, M. J., Normand, J., & Raju, N. S. (1987). Examinee attitudes toward computer-administered ability testing. *Computers in Human Behavior*, 3(2), 95-107. [https://doi.org/10.1016/0747-5632\(87\)90015-X](https://doi.org/10.1016/0747-5632(87)90015-X)
- Çakır, H. (2006). *Effectiveness of web and computer assisted instructions developed concerning dominant intelligence type on traffic education*. (Unpublished doctoral dissertation). Ankara University, Graduate School of Educational Science, Ankara.
- Davey, T. (2011). *A guide to computer adaptive testing systems educational testing service for technical issues in large-scale assessment (TILSA)*. State Collaborative on Assessment and Student Standards (SCASS). Retrieved May 11, 2018, from <https://files.eric.ed.gov/fulltext/ED543317.pdf>
- DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.
- Economides, A. A. & Roupas, C. (2007). Evaluation of computer adaptive testing systems. *International Journal of Web-Based Learning and Teaching Technologies*, 2(1), 70–87. <https://doi.org/10.4018/jwltd.2007010104>
- Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methodology and findings. *Psychological Bulletin*, 113(2), 279-294. Retrieved from https://www.researchgate.net/profile/Davina-French/publication/14748646_Behavioral_correlates_of_individual_differences_in_road-traffic_crash_risk_An_examination_of_methods_and_findings/links/00b49530310a4ec836000000/Behavioral-correlates-of-individual-differences-in-road-traffic-crash-risk-An-examination-of-methods-and-findings.pdf
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Eroğlu, M. G. (2013). *Comparison of different test termination rules in terms of measurement precision and test length in computerized adaptive testing*. (Published doctoral dissertation). Hacettepe University, Graduate School of Educational Sciences, Ankara.
- Georgiadou, E., Triantafillou, E., & Economides, A. A. (2006). Evaluation parameters for computer-adaptive testing. *British Journal of Educational Technology*, 37(2), 261-278. <https://doi.org/10.1111/j.1467-8535.2005.00525.x>
- Georgiadou, E., Triantafillou, E., & Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment*, 5(8), 1-39. Retrieved from <https://files.eric.ed.gov/fulltext/EJ838610.pdf>
- 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, The Graduate School of Natural and Applied Sciences, Ankara.
- Güleçen, M. (1998). *The effect of driving courses in traffic education*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. USA: Sage Publications.
- Horswill, M. S., & Coster, M. E. (2001). User-controlled photographic animations, photograph-based questions, and questionnaires: Three internet-based instruments for measuring drivers' risk-taking behavior. *Behavior Research Methods, Instruments, & Computers*, 33(1), 46-58. <https://doi.org/10.3758/BF03195346>

- İnal, K. (2001). *Traffic safety and assessment of driving education in Turkey*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.
- İşeri, A. I. (2002). *Assessment of students' mathematics achievement through computer adaptive testing procedures*. (Unpublished doctoral dissertation). Middle East Technical University, The Graduate School of Natural and Applied Sciences, Ankara.
- Kaçmaz-Omak, S. (2012). *Awareness levels of primary school 5th-grade students about traffic information and traffic accidents*. (Unpublished master thesis). Istanbul University, Institute of Health Sciences, Istanbul.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability*. (Unpublished doctoral dissertation). Middle East Technical University, The Graduate School of Natural and Applied Sciences, Ankara.
- Kaptan, F. (1993). *Comparison of adaptive (individualised) test application and traditional paper-pencil test application in estimation of ability*. (Unpublished doctoral dissertation). Hacettepe University, Graduate School of Social Sciences, Ankara.
- Kernan, M. C., & Howard, G. S. (1990). Computer anxiety and computer attitudes: An investigation of construct and predictive validity issues. *Educational and psychological measurement*, 50(3), 681–690. <https://doi.org/10.1177/0013164490503026>
- Kezer, F. (2013). *Comparison of computerized adaptive testing*. (Published doctoral dissertation). Ankara University, Graduate School of Educational Science, Ankara.
- Kezer, F. & Koç, N. (2014). A comparison of computerized adaptive testing. *Journal of Educational Sciences Research*, 4(1), 145–174.
- Kim, J., & McLean, J. E. (1995, April). *The influence of examinee test taking motivation in computerized adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Kimura, T. (2017). The impacts of computer adaptive testing from a variety of perspectives. *Journal of Educational Evaluation for Health Professions*, 14(12), 1–5. <https://doi.org/10.3352/jeehp.2017.14.12>
- Köklü, N. (1990). *A comparison of tailored test, which is developed according to classical test theory, and group test*. (Unpublished doctoral dissertation), Hacettepe University, Graduate School of Social Sciences, Ankara.
- Kuyumcu, M. (2001) *The ideas of the personnel and instructors who participate in the in-seervices training activities related to the traffic arranged by Gendarme Schools Command*. (Unpublished master thesis). Hacettepe University, Graduate School of Social Sciences, Ankara.
- Linacre, J. M. (2000). *Computer-adaptive testing: A methodology whose time has come* [MESA Memorandum No. 69]. Retrieved from https://www.cehd.umn.edu/EdPsych/C-Bas-R/Docs/Linacre2000_CAT.pdf
- Linacre, J. M. (2006). Computer adaptive tests (CAT): Standard errors and stopping rules. *Rasch Measurement Transactions*, 20(2), 1062. Retrieved from www.rasch.org/rmt/rmt202f.htm
- Linden, W. & Glas, G. (2002). *Computerized adaptive testing: Theory and practice*. New York: Kluwer Academic Pub.
- McKenna, F. P., & Horswill, M. S. (1999). Hazard perception and its relevance for driver licensing. *Journal of the International Association of Traffic and Safety Sciences*, 23(1), 36–41. Retrieved from <https://trid.trb.org/Results?txtKeywords=McKenna&txtTitle=&xtSerial=%22IATSS%20Research%22&ddlSubject=&txtReportNum=&ddlTrisfile=&xtIndex=&specificTerms=&txtAgency=&txtAuthor=&ddlResultType=&chkFulltextOnly=&recordLanguage=&subjectLogic=or&dateStart=&dateEnd=&rangeType=emptyrange&sortBy=publisheddate&sortOrder=DESC&rpp=25#/View/492227>

- Ministry of Transport and Communication. (2015). *Driving in Finland: Licences, driving schools, rules & vehicles*. Retrieved January 21, 2017, from www.expat-finland.com/living_in_finland/driving.html
- Miser, R. (1999). *Halk eğitimi ve toplum kalkınması (Public education and community development)*. Ankara: MEB-Türk Tarih Kurumu Pub.
- Mulkaat-Ajibola, Y. (2015). Impact assessment of road traffic accidents on Nigerian economy. *Journal of Research in Humanities and Social Science*, 3(12), 8-16. Retrieved from <http://eprints.federalpolyilaro.edu.ng/615/1/YUSUFFM.A12.pdf>
- Nydick, S. & Weiss, D. (2009). A hybrid simulation procedure for the development of CATs. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved May 15, 2018, from www.psych.umn.edu/psylabs/CATCentral
- Orcutt, V. L. (2002, February). *Computerized adaptive testing: Some issues in development*. Paper presented at the annual meeting of the Educational Research Exchange, University of North Texas, Denton, Texas.
- Özbaşı, D. & Demirtaşlı, N. (2015). Development of computer literacy test as computerized adaptive testing. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2), 218–237. <https://doi.org/10.21031/epod.79491>
- Öztuna D. (2012). *A computerized adaptive testing software (CAT): SmartCAT*. European Rasch Training Group (ERTG) Meeting, 17-19 April 2012, Leeds, UK.
- Powers, D. E., & O'Neill, K. (1993). Inexperienced and anxious computer users: Coping with a computer-administered test of academic skills. *Educational Assessment*, 1(2), 153-173. https://doi.org/10.1207/s15326977ea0102_4
- Revuelta, J. & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4) 311–327. <https://doi.org/10.1111/j.1745-3984.1998.tb00541.x>
- Rudner, L. M. (1998). *An online, interactive, computer adaptive testing tutorial*. 11/98. Retrieved April 24, 2017, from <http://EdRes.org/scripts/cat>
- Schmidt, F. L., Urry, V. W., & Gugel, J. F. (1978). Computer assisted tailored testing: Examinee reactions and evaluations. *Educational and Psychological Measurement*, 38(2), 265–273. <https://doi.org/10.1177/001316447803800208>
- Sutton, R. E. (1991). Equity and computers in the schools: A decade of research. *Review of Educational Research*, 61(4), 475–503. <https://doi.org/10.3102/00346543061004475>
- Tanrikulu, S. (2002). Trafik kazalarının önlenmesi bağlamında trafik güvenliği eğitiminin rolü ve trafik kültürü. (The role of traffic safety education in the prevention of traffic accidents, and traffic culture). *Polis Bilimleri Dergisi*, 5(1), 45-60.
- Thissen, D. & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 103-134). Mahwah (NJ): Lawrence Erlbaum.
- Thompson, N., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1), 1–9. Retrieved from <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1242&context=pars>
- Turkish Statistical Institute [Türkiye İstatistik Kurumu (TÜİK)]. (2018). *Highway traffic accident statistics*. Retrieved November 27, 2018, from http://www.tuik.gov.tr/PreTablo.do?alt_id=1051
- Türkoğlu, M. (2002). *The Effect of the training in the private driving schools on traffic accidents*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.
- Tütüncü, M. (2001). *The Importance of traffic education in formal education for Turkey*. (Unpublished master thesis). Gazi University, Graduate School of Natural and Applied Sciences, Ankara.

- Vispoel, W. P. (1998). Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement*, 35(4), 328-345. <https://doi.org/10.1111/j.1745-3984.1998.tb00542.x>
- Vispoel, W. P. (2000). Reviewing and changing answers on computerized fixed-item vocabulary tests. *Educational and Psychological Measurement*, 60(3), 371-384. <https://doi.org/10.1177/00131640021970600>
- Vursavaş, F. (2004). *Evaluation of the driving school curriculum*. (Unpublished master thesis). Ankara University, Graduate School of Educational Science, Ankara.
- Wainer, H. & Mislevy, R. J. (1990). *Item response theory, item calibration and proficiency estimation*. In H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, R. J. Mislevy, L. Steinberg, & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (pp. 1-21). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6(4), 473-492. <https://doi.org/10.1177/014662168200600408>
- 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. <https://doi.org/10.1080/07481756.2004.11909751>
- Weiss, D. J. & Guyer, R. (2012). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing*. St. Paul MN: Assessment Systems Corporation.
- Weiss, D. J. & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361–375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>
- Wilder, G., Mackie, D., & Cooper, J. (1985). Gender and computers: Two surveys of computer-related attitudes. *Sex Roles*, 13(3), 215-228. Retrieved from <https://link.springer.com/content/pdf/10.1007/BF00287912.pdf>
- Yaşar, M. (1999). *A study about individualized tests [Bireyselleştirilmiş testler üzerine bir çalışma]*. (Unpublished doctoral dissertation). Hacettepe University, Ankara, Turkey.
- Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. In: I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J. D. Irwin (Eds.), *Distracted driving* (pp. 379-405). Sydney, NSW: Australasian College of Road Safety.