IJATE

Volume 5 Issue 3 2018

International Journal of Assessment Tools in Education

International Journal of
Assessment Tools in Education

http://ijate.net/

http://dergipark.gov.tr/ijate

e-ISSN: 2148-7456

e-ISSN 2148-7456

http://www.ijate.net/index.php/ijate/index

Volume 5 Issue 3 2018

Dr. Izzet KARA
Editor in Chief
International Journal of Assessment Tools in Education (IJATE)

Pamukkale University,
Education Faculty,
Department of Mathematic and Science Education,
20070, Denizli, Turkey

Phone : +90 258 296 1036
Fax : +90 258 296 1200
E-mail : ijate.editor@gmail.com

Publisher : İzzet KARA
Frequency : 4 Issues per year (starting from June 2018)
Online ISSN: 2148-7456
Website : http://www.ijate.net/index.php/ijate
Design & Graphic: IJATE

Support Contact
Dr. İzzet KARA
Journal Manager & Founding Editor
Phone : +90 258 296 1036
Fax : +90 258 296 1200
E-mail : ikara@pau.edu.tr

hosted by Turkish JournalPark ACADEMIC

International Journal of Assessment Tools in Education

International Journal of Assessment Tools in Education (IJATE) is published 4 issues per year starting from June 2018. (one volume per year, four issues per year - March, June, September and December). IJATE accepts original theoretical and empirical English-language manuscripts in psycho-educational assessment. Theoretical articles addressing new developments in measurement and innovative applications are welcome.

There is no submission or publication process charges for articles in IJATE.

IJATE is indexed in:

- Emerging Sources Citation Index (ESCI) (Web of Science Core Collection)
- TR Index (ULAKBIM),
- DOAJ,
- Google Scholar,
- Index Copernicus International
- Türk Egitim İndeksi,
- Open Access Journals,
- Akademik Dizin,
- Academic Keys,
- CiteFactor (ASJ),
- SIS (Scientific Index Service) Database,
- SCIPIO (Scientific Publishing & Information Online),
- MIAR 2015 (Information Matrix for Analysis of the Journals),
- I2OR Indexing Services,
- JournalTOCs
- Sosyal Bilimler Atıf Dizini (SOBIAD)
- International Innovative Journal Impact Factor (IIJIF)

ISSN: 2148-7456

ISSN: 2148-7456

Table of Contents

Research Article

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 389–402

DOI: 10.21449/ijate.379139

Published at http://www.ijate.net http://dergipark.gov.tr/ijate Research Article

Adaptation of the Gamification User Types Hexad Scale into Turkish

Özcan Erkan Akgün[2], **Murat Topal**[2,*]

[1]İstanbul Medeniyet Üniversitesi, Kuzey Kampüsü, Dumlupınar Mahallesi, D-100 Karayolu No:98, 34000 Kadıköy/İstanbul, Türkiye

[2]Sakarya Üniversitesi, Eğitim Fakültesi, E Blok, Oda No: 5215, Hendek / Sakarya, Türkiye

Abstract: The aim of this study is to adapt the Gamification User Types Hexad Scale (GUTHS) created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016) into the Turkish context. The original scale consists of six dimensions, each of which has four items. The dimensions — and the user types that they refer to — are "Free spirit," "Socializer," "Achiever," "Philanthropist," "Player," and "Disruptor." It is the motive of this study that identifying and studying these user types may prove useful for understanding the effects of gamification dynamics and mechanics and assist in designing specific gamification techniques corresponding to each user type. The adaptation of the instrument began with translation, continued with an examination of the linguistic equivalence, and finalized with analyses of validity and reliability. The scale items were initially translated by the researchers. The translation was examined by seven experts with good English proficiency to finalize the Turkish version. To verify the linguistic equivalence, both the Turkish and English versions were then administered to 30 English Language Education (ELE) students. The correlation findings showed a high degree of correlation between the Turkish and English versions. Next, the Turkish version was administered to 452 university students studying at the Faculty of Education, Sakarya University, to check its validity and reliability. The results obtained from a confirmatory factor analysis and the reliability analysis indicate that the Turkish version of the scale is valid and reliable. It is recommended to use the translated scale in research especially on determining the effects of factors related to user types and on designing more affective gamification strategies.

ARTICLE HISTORY

Received: 01 January 2018
Revised: 23 April 2018
Accepted: 30 April 2018

KEYWORDS

User type hexad framework,
Gamification,
User Types,
Scale,
Game

1. INTRODUCTION

Employment of new digital strategies might be effective in using present technologies in education, provided that such strategies determine how to use software and technology to develop in-class and out-of-class teaching and learning (Johnson et al., 2014). Gamification, widely used in bussiness and marketing, is considered among new digital strategies and a new method leading to active and continuous student engagement (Johnson et al., 2014). Gamification is mainly considered as a strategy to create interest in instruction, rather than as a strategy for efficiency and effectiveness (Reigeluth, 2013). Even so, it still carries a potential

CONTACT: Murat Topal ✉ mtopal@sakarya.edu.tr Sakarya Üniversitesi, Eğitim Fakültesi, E Blok, Oda No: 5215, Hendek / Sakarya, Türkiye

ISSN-e: 2148-7456 /© IJATE 2018

to directly impact the "individual performance of the learner." Therefore, gamification may significantly contribute to the effectiveness of instruction, because it is considered an interesting and effective strategy to motivate students during instruction by some researchers (Lee & Hammer, 2011).

Although the use of games in education is an old practice, the concept of gamification (or gameful design) was first introduced by Nick Pelling in 2002 (Marczewski, 2013). Gamification term was first mentioned in the literature in 2008, and it became popular in the second half of 2010 through conferences and applications (Deterding et al., 2011). Gamification can simply be defined as the use of thought processes, mechanics and dynamics to ensure the engagement of users and to solve problems (Zichermann & Cunningham, 2011). Another definition of gamification is the use of computer game mechanics, dynamics and structures to achieve the targeted learning and it is believed to become a million-dollar industry in the fields of politics, health and marketing (Macmillan, 2011; Wu, 2012). According to Bunchball (2010), it is a strategy used to motivate individuals - such as customers, employees and patients, as well as students - and influence their behaviors. In short, the target audience of gamification is any group of people; and the aim of gamification is to sustain the engagement of the target group to achieve a desired behavior (Bunchball, 2010).

Gamification in education aims to help integrate and use gaming elements that make computer games fun in learning and teaching processes, thereby increasing the participation of students in learning and teaching activities, leading to a more entertaining instruction (Simoes, Redondo & Vilas, 2013). The "fun" factor mentioned here allows users or students to focus on solving the real problems in life using the motivational potential of computer games (Lee & Hammer, 2011) because computer games can help individuals use their problem-solving skills voluntarily for several hours in the game environment (Gee, 2008). Gamification is designed by integrating game mechanics and dynamics into non-gaming related situations. Game mechanics are different activities, desired behaviors and control mechanisms in processes that are used for gamifying tasks. Game dynamics are associated with the individual desire and motivation that arises as a result of a gamified process (Bunchball, 2010).

It can be argued that learning and teaching processes already contain certain gamification elements. For example, a student who completes his or her homework properly earns points as a reward, and each of these points is called a "grade" (good, very good, etc.), instead of a "badge," and if a student fulfills the desired achievement, she or he moves to a higher level at the end of the academic year by passing to a higher grade. As seen in the example above, the school system includes certain basic gamification experiences (Lee & Hammer, 2011). However, schools may not be capable of helping students achieve the desired engagement (Lee & Hammer, 2011). In such situations, gamification is considered as an interesting and effective strategy to motivate students in instructional settings. However, adaptation of gamification into educational settings as a strategy requires the scrutiny of educational scientists as well as the identification and use of appropriate instructional methods. Simply using points and rewards and attempting to gamify without comprehension of the real reasons behind the educational problems might fail to motivate learners and might impede efficient learning (Lee & Hammer, 2011). Moreover, gamification is more than just adding game dynamics and mechanics such as badges and leaderboards to a learning material and waiting for users to utilize these elements (Yılmaz, 2015). That is, arbitrarily adding gamification elements to an educational environment does not mean that the said environment has been gamified.

Gamification experts claim that gamification designs should include a self-sustaining game cycle, as in digital game designs (Werbach & Hunter, 2012; Chou, 2015). It is important to have knowledge of user types and characteristics to be able to design an efficient game cycle. It is necessary to identify and adequately use appropriate gaming elements. To do that, the

gaming elements should target the needs and characteristics of learners, as the target audience. Gamification may not motivate people in educational settings without knowing the actual causes behind the educational problems. Moreover, gamification focuses on engagement and motivation. If the educational problem cannot be eliminated by improving learner engagement and motivation, gamification will not be a useful method, either. Another important point here is that it is necessary to know what motivates people. Recent research on gamification shows that a poor implementation may lead to a failure of students to achieve goals (Hamari, Koivisto & Sarsa, 2014; Nacke & Detering, 2017). Personalized gamification applications targeting user motivation provide better results than the one-size-fits-all approaches (Tondello et al., 2016). Moreover, recent studies suggest that personalized gamification applications show better results on motivation (Foucault et al., 2018), emotional engagement (Mora et al., 2018), and perceived persuasiveness (Tondello, Orji & Nacke, 2017; Orji, Tondello & Nacke, 2018; Orji, Nacke & Marco, 2017). To this end, the Gamification User Types Hexad Scale (GUTHS) created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016) aims to determine user types. The scale not only provides identification of user profiles, but also recommendations on motivational factors and appropriateness of gamification elements for any given user profile. Based on the said properties of the scale, the present study was conducted to adapt the scale to the Turkish language and culture to provide contribution to the literature in Turkish.

2. METHOD

The adaptation of the scale began with an initial translation. The initial translation was carried out by the authors. Next, seven experts reviewed the initial translation. Two of these experts had Ph.D. degrees in English language, one in measurement and evaluation, one in Turkish language, one in educational psychology, and two in educational technology. Based on the experts' opinions, 10 items were accepted without modification, and 14 items were rewritten. A consensus was established among the experts, and a pilot questionnaire form was designed.

2.1. Sample

The scale created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016) based on the data obtained from the graduate and undergraduate students at the University of Waterloo, Canada. To mimic the conditions, we used the convenience sampling method in the present study. There were no specific criteria for selecting participants in the convenience sampling. We tried to reach all freshmen students and invited them to voluntarily participate in the study. To test the linguistic equivalence, the sample included 30 junior and senior students attending the English Language Education (ELE) program of Sakarya University in Turkey. To carry out validity and reliability analyses, we collected data from a total of 452 freshmen students who were attending courses at Sakarya University, Faculty of Education, in fall 2016. These students were from the following departments: 65 students from Computer Education and Instructional Technology, 52 from Science Education, 31 from Mathematics Education, 46 from Pre-School Education, 43 from Special Education, 57 from Psychological Services in Education Department, 55 from Elementary Education, 49 from Social Studies Education, and 54 from Turkish Language Education.

2.2. The Original Scale

GUTHS validated based on Marczewski's (2015) user type hexad framework of Marczewski (2015). Tondello et al. (2016) was developed GUTHS to create and validate a

survey to assess an individual's Hexad user type and verify the association between the Hexad user types and the game design elements that suggested by Marczewski's (2015) research. Because of GUTHS (Tondello et al., 2016) has a validaty and reliability our research focus on adaptation of this assesment tool. The aim of the scale is to match commonly used and popular gaming elements with certain user profiles suggested on the research of Marczewski's (2015). Thus, the scale aims to select appropriate gamification elements based on the characteristics of a target audience and individualize the design of gamification when designing gamification systems. GUTHS has focused on six user types — and six elements that motivate these user types — while creating the scale. The first type, "Philanthropists," is the type of users who are humanitarian and like to help others. "Socializers" tend to prefer being social. "Free Spirits" prioritize their freedom. "Achievers" like to overcome obstacles and difficulties. "Players" like to play games that motivate using rewards. The sixth and the final type of users, "Disruptors," like to obstruct, interfere and sabotage things. The scale includes 24 items, with four items in each factor. In the Turkish version, the factors of the scale were translated to have the same meanings as those of the original scale. The Cronbach's alpha internal consistency coefficients of the factors in the original scale were as follows: Philanthropist: .89, Socializer: .83, Free Spirit: .72, Achiever: .75, Player: .69, and Disruptor: .73. Tondello et al. also utilized the test-retest reliability method to measure the reliability of the instrument.

Tondello et al. (2016) preferred expert opinions to check the content validity of the scale. They designed a workshop at the Austrian Institute of Technology to determine the item pool of GUTHS by a group discussion based on the research of Marczewski (2015). The preliminary form created earlier by Tondello et al. (2016) was modified and finalized based on the views of six experts who were specialists in scale development and game mechanics.

In order to check the criterion validity, Tondello et al. (2016) selected three different instruments, each as a criterion. The first one of these instruments had game elements and was frequently used in the field of gaming and gamification (see Table 1). They chose this as a criterion because personalization could be used in game/gameful designs to tailor users' game or interaction mechanics (Tondello et al., 2016). The correlation between the scores obtained for each 7-point Likert-type item in the instrument that included the 32 game elements and the user types that suggested by the research of Marczewski (2015) was examined. The distribution of game elements based on the obtained correlation values is given in Table 1.

Previous studies have shown that there is a relationship between personality and player behavior in video games (Johnson & Gardner, 2010; Yee et al., 2011). Starting from this fact, the second instrument Tondello et al. (2016) selected as a criterion was the Five Factor Personality Scale developed by Rammstedt and John (2007). This scale had five factors: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The factors of the scale were matched with the factors of GUTHS (Tondello et al., 2016) that created based on Marczewski's (2015) user type hexad framework. The correlations between the factors of the two scales are presented in Figure 1.

The "Balanced Inventory of Desirable Responding (BIDR)", developed by Winker, Kroh, and Martin (2006) was selected as the third criterion to also check the criterion validity by Tondello et al. (2016). This was the short version of an earlier instrument developed by Paulhus (1994). It has been reported in the literature that the scale is used to verify and assess the objectivity of individuals' self-declarations (Winker, Kroh & Martin, 2006). The scale has two dimensions. One dimension measures the positive exaggeration of honest response tendencies (increase in self-deception), while the other measures the tendency of individuals to construct themselves in a premeditated manner for their audiences (Tondello et al., 2016).

Correlation analyses were conducted on the data obtained from all the abovementioned scales (on each dimension of all scales) and Tondello et al. (2016) stated that the analysis results supported the hypothesized theories.

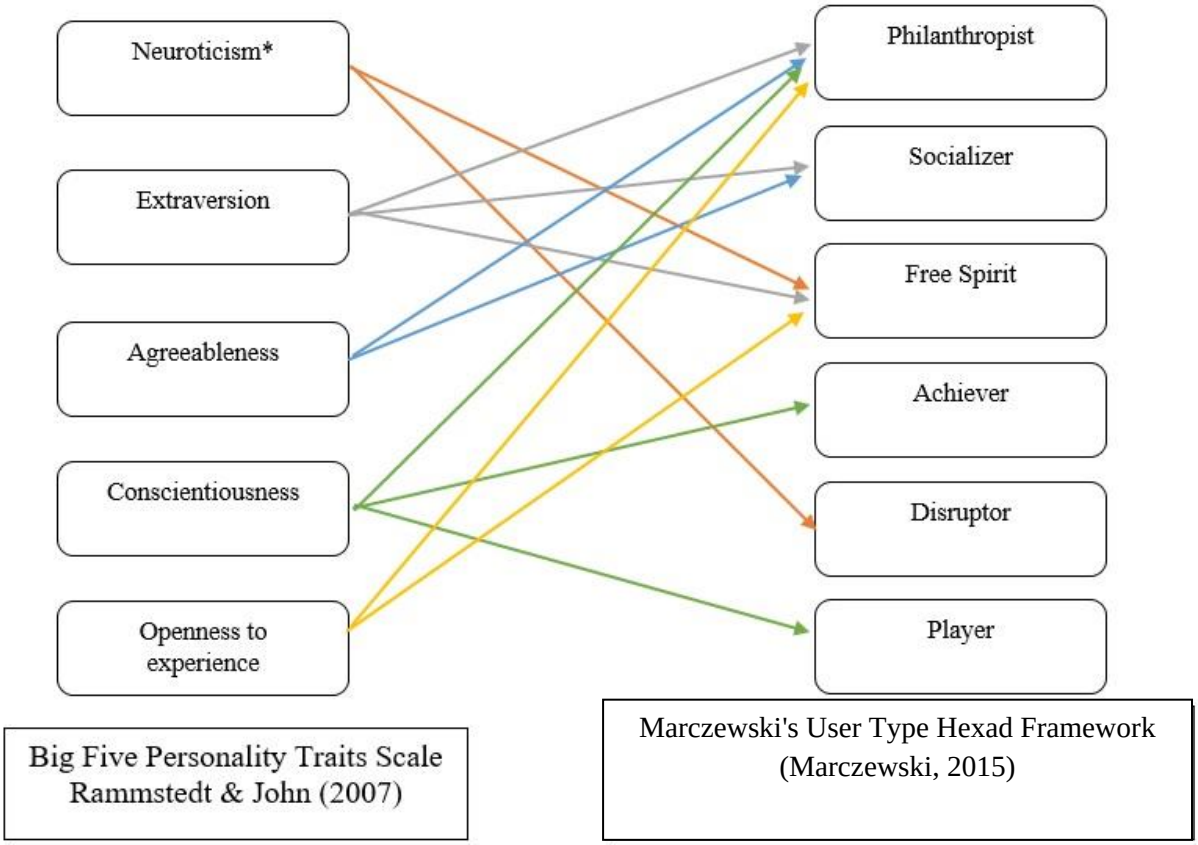


Figure 1. Relationships between factors of Big Five Personality Traits Scale (Rammstedt & John, 2007) and Marczewski's (2015) user types.

*Correlations with neuroticism are negative, other correlations are positive. All correlations indicated weak relationships (values ranging from .14 to .29) and were significant at .05 level.

2.3. Factors of the scale and sample items

Table 1 summarizes the factors and the corresponding scale items, game elements and factor definitions included in GUTHS, which was created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016).

Table 1. Factors, scale items, related game elements and factor definitions of GUTHS created based on Marczewski's (2015) user type hexad framework

Factor/User Type	Scale items	Suggested game elements	Definition
Philanthropist (4 items)	1. I like helping others to orient themselves in new situations. 3. The well-being of others is important to me.	- Collection and Trading - Gifting - Knowledge sharing	- Philanthropists are motivated by purpose. They are altruistic and willing to give without expecting a reward.

Factor/User Type	Scale items	Suggested game elements	Definition
	13. It makes me happy if I am able to help others. 22. I like sharing my knowledge.	- Administrative roles	
Socializer (4 items)	5. It is important to me to feel like I am part of a community. 10. I enjoy group activities. 18. I like being part of a team. 20. Interacting with others is important to me.	- Guilds or Teams - Social networks - Social comparison or pressure - Social competition - Social discovery	- Socializers are motivated by relatedness. They want to interact with others and create social connections.
Free Spirit (4 items)	2. I like to try new things. 6. Being independent is important to me. 9. It is important to me to follow my own path. 11. I often let my curiosity guide me.	- Exploratory tasks - Nonlinear gameplay - Easter eggs - Unlockable or rare content - Creativity tools - Customization	- Free Spirits are motivated by autonomy, meaning freedom to express themselves and act without external control. They like to create and explore within a system.
Achiever (4 items)	12. I like overcoming obstacles. 16. It is difficult for me to let go of a problem before I have found a solution. 17. It is important to me to always carry out my tasks completely. 23. I like mastering difficult tasks.	- Challenges - Certificates - Learning - Quests - Levels or Progression - Boss battles	- Achievers are motivated by competence. They seek to progress within a system by completing tasks, or prove themselves by tackling difficult challenges.
Disruptor (3 items)	14. I like to question the status quo. 15. I dislike following rules. 19. I see myself as a rebel.	- Innovation platforms - Voting mechanisms - Development tools - Anonymity - Anarchic gameplay	- Disruptors are motivated by the triggering of change. They tend to disrupt the system either directly or through others to force negative or positive changes. They like to test the system's boundaries and try to push further.
Player (3 items)	4. I like competitions where a prize can be won. 7. Return of investment is important to me.	- Points - Rewards or Prizes - Leaderboards - Badges or Achievements	- Players are motivated by extrinsic rewards. They will do whatever to earn a reward within a system, independently of the type of the activity.

Factor/User Type	Scale items	Suggested game elements	Definition
	24. Rewards are a great way to motivate me.	- Virtual economy - Lotteries or Games of chance	

* This table was prepared using the study of Tondello et al. (2016: pp. 231–243).

2.4. Interpretation of the data collected using the scale

Once factor scores are calculated using the data obtained through GUTHS, the factor with the highest score shows the dominant user type of an individual (Tondello et al., 2016). Two items were excluded from the adapted scale. The obtained scores were converted into percentages with the following formula to determine each individual's dominant user type:

$$Dominant\ User\ Type\ Score = \frac{Score\ of\ the\ Factor}{The\ Highest\ Possible\ Score\ of\ the\ Factor} \times 100$$

2.5. Data Collection and Analysis

Data were collected from the participants through paper-and-pencil form after they were given information about the purpose of the study. Each participant filled a scale application form during the data collection. The filled forms were reviewed to eliminate the forms with items that were not marked or had multiple marks per item. Such forms were excluded from the study. The subsequent analyses were conducted on the dataset that included the remaining 417 forms.

A confirmatory factor analysis (CFA) was conducted to determine the construct validity of the instrument. The convergent and discriminant validities were examined. In order to determine the reliability of the scale, the Cronbach's alpha coefficient was calculated.

3. FINDINGS

This section presents the results on the linguistic equivalence, CFA, convergent validity, discriminant validity and Cronbach's alpha reliability of the Turkish version of the scale.

3.1. Linguistic Equivalence

In order to examine the linguistic equivalence of the GUTHS, the English original and the translated version of the scale were administered to 30 junior and senior students in the ELE program. The resultant Pearson correlation coefficients between each of the items of the scales are presented in Table 2.

All the correlations between the Turkish and English items presented in Table 2 were significant ($p<.05$), positive and with coefficients greater than .30. It was found that the correlation coefficient between the total scores of the Turkish and English versions was .94. The above mentioned findings demonstrated that the responses given by the participants in the Turkish and English versions of the scale were equivalent.

Table 2. Correlation coefficients for comparing the linguistic equivalence of GUTHS versions

Item number	r	Item number	r
1	.85	13	.90
2	.60	14	.71
3	.89	15	.73
4	.60	16	.64
5	.65	17	.66
6	.66	18	.84
7	.69	19	.80
8	.82	20	.61
9	.83	21	.71
10	.82	22	.67
11	.81	23	.93
12	.83	24	.79
Dimensions		Dimensions	
Philanthropist	.81	Achiever	.89
Socializer	.85	Player	.86
Free Spirit	.88	Disruptor	.79

3.2. Confirmatory Factor Analysis

As mentioned earlier, the six-factor structure of the Turkish version of the scale, as in the original scale, was confirmed by expert opinions. A CFA was also conducted using the data collected from the 417 undergraduate students to further verify this six-factor structure. During the analysis, two items (items 7 and 21) were found to load weakly on their relevant latent variables (the factors) and have high error rates. These items were removed, and the analysis was replicated after these items were removed from the model. The resultant goodness of fit values are shown in Table 3 along with perfect and acceptable fit values from the literature for comparison.

Table 3. CFA results of GUTHS

Fit Index	Perfect Fit Values	Acceptable Fit Values	Values from CFA
χ^2/SD	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 3$	2.742
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI \leq .95$	.90
AGFI	$.90 \leq AGFI \leq 1.00$	$.85 \leq AGFI \leq .90$	.86
CFI	$.95 \leq CFI \leq 1.00$	$.90 \leq CFI \leq .95$	.97
NFI	$.95 \leq NFI \leq 1.00$	$.90 \leq NFI \leq .95$	.95
RMSEA	$.00 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$	.06
SRMR	$.00 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$	.05

The minimum Chi-square value based on the CFA was χ^2=531.97, *SD*=194, p<0.01, and the ratio of the Chi-square to the degree of freedom was 2.742, indicating a significant deviation from an acceptable fit. We believe that the deviation was due to the sample size. A review of the other goodness of fit values demonstrated that the AGFI, GFI, CFI, NFI, and RMSEA values were in the acceptable fit range, and the SRMR value was in the perfect fit range (Bentler & Bonett, 1980; Marsh et al., 2006; Schermelleh-Engel & Moosbrugger, 2003; Byrne & Campbell, 1999). These values based on the CFA indicated that the data fit the six-factor structure as specified in the model These values indicated that the observed construct of the scale was compatible with the expected construct, thus the scale had an acceptable construct validity. However, the 2 items mentioned above that were included in the original

scale were excluded from the Turkish version. Factor loadings of the six-factor model are presented in Figure 2.

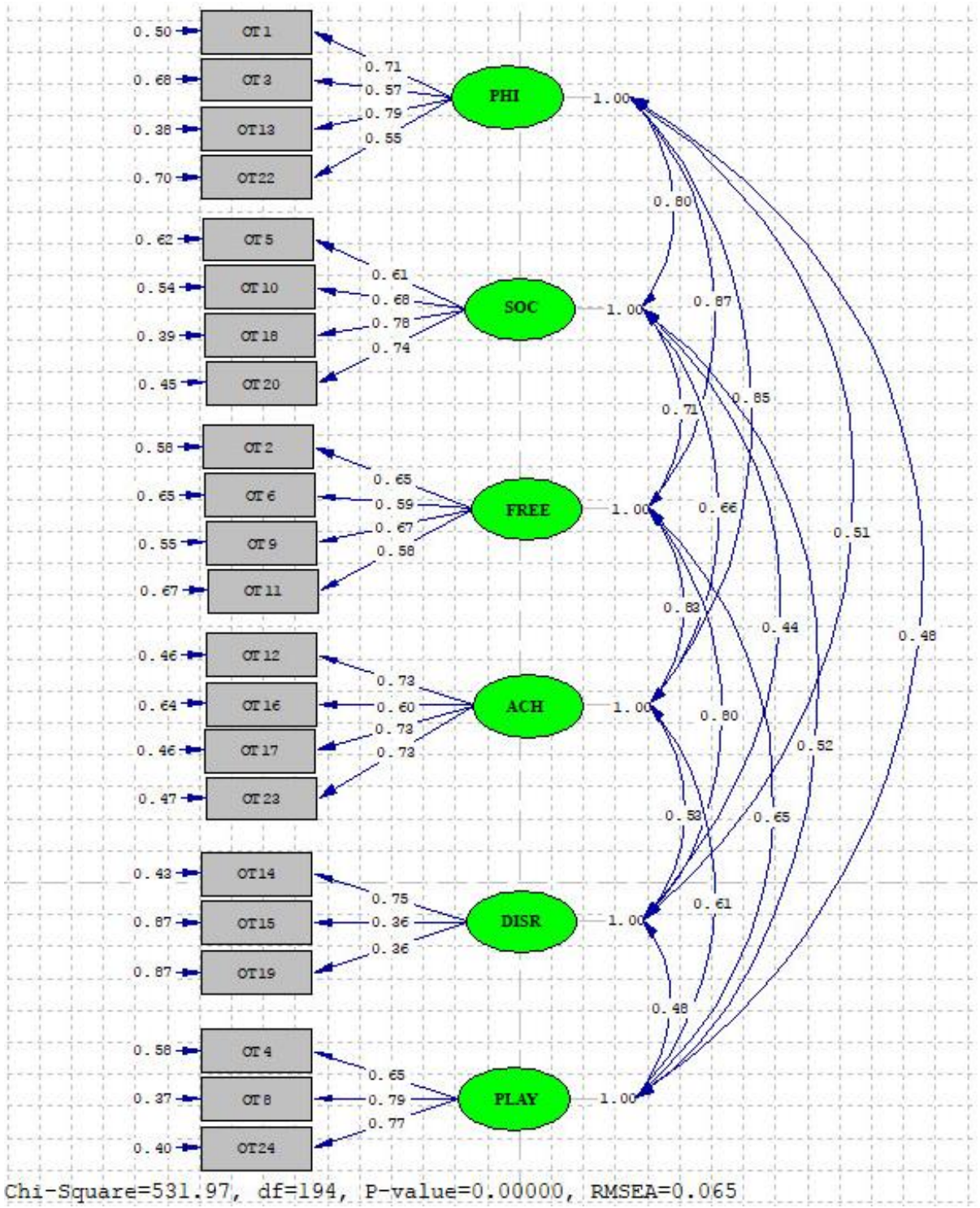


Figure 2. CFA results and Path Diagram of the Turkish version of GUTHS (Phi=Philantropsist, Soc=Socialiser, Free=Free Spirit, Ach=Achiever, Disr=Disruptor, Play=Player)

3.3. Convergent and Discriminant Validities

The scale was tested for convergent validity and discriminant validity. For the convergent validity, the average variance extracted (AVE) values were as follows: Philantropist: .57, Socializer: .61, Free Spirit: .53, Achiever: .61, Disruptor: .51, and Player: .69. Since the values were greater than .50, the findings indicated an acceptable level of convergent validity (Fornel & Larcker, 1981).

For the discriminant validity, square roots of the AVE values were calculated (see the diagonal values in bold in Table 4). Fornel & Larcker (1981) stated that square root values higher than .50 and higher than the correlations between the other factors in the same column of each factor may be an evidence for discriminant validity. Table 4 shows the correlations between the factors, in addition to the square roots of the AVE values. As can be seen in the table, the square root values are the greatest values in their respective columns. These findings indicated that the discriminant validity of the instrument was acceptable.

Table 4. Correlation between the square roots of AVE values of the GUTHS factors

	Philantropist	Socializer	Free Spirit	Achiever	Disruptor	Player
Philantropist	**.75**					
Socializer	.63	**.78**				
Free Spirit	.63	.54	**.72**			
Achiever	.64	.51	.61	**.78**		
Disruptor	.21	.23	.28	.28	**.71**	
Player	.36	.42	.47	.47	.28	**.83**

3.4. Scale Reliability

The reliability of the gamification user types scale was examined by calculating the Cronbach's alpha internal consistency coefficient. The Cronbach's alpha coefficient calculated for the whole scale was .89. The coefficients for the factors were as follows: Philanthropist: .76, Socializer: .79, Free Spirit: .72, Achiever: .80, Disrupter: .71, and Player: .78. All these values were greater than .71. There is not a certain rule of thumb for judging the Cronbach'a alpha coefficients (Cho & Kim, 2015). It can be said that the higher the α coefficient, the more the items have shared covariance and probably measure the same construct. These coefficients varied between .70 and .89 in the original study (Tondello et al., 2016). The coefficients of the Turkish version vary between .71 and .79, and are very similar to the values found in Tondello et al. (2016). A value of .70 and higher is considered a cut-off point commonly accepted in social sciences (Büyüköztürk, 2012). Thus, it can be claimed that the instrument was reliable at an acceptable level.

4. DISCUSSION AND CONCLUSION

The use of digital game features that promote game playing in different fields has introduced the concept of gamification (Bunchball, 2010). Learning and teaching processes have also been influenced by gamification as a new strategy, especially in terms of engagement and motivation. GUTHS created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016) has been used to determine what motivates six different user types and the appropriate gamification elements that can be used for these user types. The scale has the potential to become a key instrument for individuals who aim to develop gamification-assisted processes and/or products. It can also offer principles for designing facilities in the future. Thus, the present study aimed to adapt GUTHS to the Turkish language and culture.

GUTHS that created based on Marczewski's (2015) user type hexad framework and validated by Tondello et al. (2016) is a fairly new instrument contributed to the literature, so it is hard to find extensive studies discussing the relationship between gamification and user types based on GUTHS. This adaptation study was completed not long after GUTHS was published in the proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. This study may contribute e-learning studies to enrich it with gamification strategies in Turkish.

This adaptation study has some limitations with its sample size, predictive validity, and test-retest reliability. It does not yet have prooven and effective suggestions for user types and gamification. Nevertheless, the international literature also suffers from similar issues. This attempt to offer a scale on user types — even just for the sake of initiating discussions — is important for future studies to improve the scale and user-type models as well as investigating gamification dynamics and mechanics.

The participants of the study were assumed to be bilingual individuals. Their answers to the Turkish and English items with a one-week interval were positively and significantly correlated, with correlation coefficients ranging from medium to high. The overall correlation between the two versions of the scale was .94. These significant, positive and high levels of correlations between the answers of participants constitute an evidence for the equivalence of the English and Turkish versions. In the original study, the authors examined the validity of the scale with workshops and through a predictive validity study. In this study, we conducted a CFA to confirm that the expected model of the scale has a good fit with the observed model based on the data from 417 participants. With this result, this study also shows that the Gamification User Types Hexad Scale Turkish version has construct, convergent and discriminant validities. These results may constitute comparable examples for studies in other languages. In this study, the reliability of the scale was examined only through the Cronbach's alpha internal consistency coefficients, relying on one-time data collection. The alpha coefficients that were obtained in this study were at acceptable levels as stated in the literature and were similar to the coefficients of the original study.

The validity and reliability analyses demonstrated that the six-factor structure of the Turkish version was compatible with the factor structure in the original form with 22 items.

In conclusion, the Gamification User Types Scale Turkish version has 22 items and 6 factors, and it is a reliable and valid measurement instrument that can be used to determine gamification user types of undergraduate students. It is recommended that the scale be used in correlational or comparative research on user profiles as well as gamification mechanics and dynamics. We believe that the scale will contribute to the efficiency of gamification and provide ideas for appropriate gamification designs.

ORCID

Özcan Erkan AKGÜN https://orcid.org/0000-0002-6486-0486
Murat TOPAL https://orcid.org/0000-0001-5270-426X

5. REFERENCES

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588-606.

Bunchball (2010). Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior [White paper]. Retrieved from http://www.bunchball.com/sites/default/files/downloads/gamification101.pdf.

Büyüköztürk, Ş. (2012). *Veri analizi el kitabı* (17. Baskı). Ankara: Pegem Akademi.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30*(5), 555-574.

Cho, E & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, *18*(2), 207-230.

Chou, Y. (2015). Octalysis: Complete gamification framework-Yu-kai Chou. *Octalysis media, Fremont*.

Deterding, S. (2012). Gamification: designing for motivation. *Interactions*, *19*(4), 14-17.

Deterding, S., Khaled, R., Nacke, L. E., & Dixon, D. (2011). *Gamification: Toward a definition*. CHI 2011 Gamification Workshop Proceedings, 12-15.

Foucault, M., Blanc, X., Storey, M. A., Falleri, J. R., & Teyton, C. (2018). *Gamification: a Game Changer for Managing Technical Debt? A Design Study*. arXiv preprint arXiv:1802.02693. Retrieved from https://arxiv.org/pdf/1802.02693.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.

Gee, J. P. (2008). Learning and games. K. Salen (Ed.) *The ecology of games: Connecting youth, games, and learning (John D. and Catherine T. MacArthur Foundation series on digital media and learning)*. Cambridge, MA: The MIT Press.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. *In System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3025-3034). IEEE.

Johnson, L., Becker, S., Estrada, V., & Freeman, A. (2014). *Horizon report: 2014 higher education.* retrieved from http://cdn.nmc.org/media/2014-nmc-horizon-report-he-EN-SC.pdf .

Johnson, D., & Gardner, J. (2010, November). Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (pp. 276-279). ACM.

Lee, J. J., & Hammer, J. (2011). Gamification in education: What, how, why bother?. *Academic exchange quarterly*, *15*(2), 146.

MacMillan, D. (2011). *Gamification': A growing business to invigorate stale websites.* 15.02.2016 retrieved from http://www.bloomberg.com/news/articles/2011-01-19/gamification-a-growing-business-to-invigorate-stale-websites

Marczewski, A. (2013). *Gamification: a simple introduction.*

Marczewski, A. C. (2015). User Types. In *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design* (1st ed., pp. 65-80). CreateSpace Independent Publishing Platform.

Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*(4), 311-360.

Mora, A., Tondello, G. F., Nacke, L. E., & Arnedo-Moreno, J. (2018). Effect of personalized gameful design on student engagement. *In Proceedings of the IEEE Global Engineering Education Conference - EDUCON 2018*. Tenerife, Spain. IEEE.

Nacke, L. E., & Deterding, S. (2017). The maturing of gamification research. *Computers in Human Behavior*, *71*, 450-454.

Orji, R., Nacke, L. E., & Di Marco, C. (2017, May). Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1015-1027). ACM.

Orji, R., Tondello, G. F., & Nacke, L. E. (2018). Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. *In Proceeding of the 2018 ACM Conference on Human Factors in Computing Systems - CHI 2018*. Montreal, QC, Canada. ACM.

Paulhus, D. L. (1994). Balanced inventory of desirable responding: Reference manual for BIDR version 6. *Unpublished manuscript, University of British Columbia, Vancouver, Canada*.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, *41*(1), 203-212.

Reigeluth, C. M. (2013). Öğretim-tasarımı teorisi nedir ve nasıl değişir?. K. Çağıltay ve İ. Göktaş (Ed.), *Öğretim Teknolojilerinin Temelleri: Teoriler, Araştırmalar, Eğilimler* (1-22). Ankara: Pegem Akademi.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, *8*(2), 23-74.

Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, *29*(2), 345-353.

Tondello, G. F., Orji, R., & Nacke, L. E. (2017, July). Recommender systems for personalized gamification. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 425-430). ACM.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (229-243). ACM, October.

Tuckman, B. W. (2007). The effect of motivational scaffolding on procrastinators' distance learning outcomes. *Computers & Education*, *49*(2), 414-422.

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

Winkler, N., Kroh, M., & Spiess, M. (2006). *Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit* (No. 579). DIW Discussion Papers.

Wu, M. (2012). *The gamification backlash + Two long term business stategies*. 03.03.2016 retrieved from http://community.lithium.com/t5/Science-of-Social-blog/The-Gamification-Backlash-Two-Long-Term-Business-Strategies/ba-p/30891

Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011, May). Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 753-762). ACM.

Yılmaz, E. A. (2015). *Oyunlaştırma.* Abaküs yayınları, İstanbul.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Inc.

Appendix. Turkish Version of The Gamification User Type Hexad Scale

Oyunlaştırma için Oyuncu Tipleri Ölçeği

Maddeler		Kesinlikle katılmıyorum	Katılmıyorum	Kısmen Katılmıyorum	Karasızım	Kısmen Katılıyorum	Katılıyorum	Tamamen Katılıyorum
1.	Başkalarına yeni durumlara uyum sağlamaları için yardım etmeyi severim.							
2.	Yeni şeyler denemekten hoşlanırım.							
3.	Başkalarının maddi-manevi iyi olması benim için önemlidir.							
4.	Karşılığında kazanılacak bir ödül olduğunda rekabetten hoşlanırım.							
5.	Bir topluluğun parçası olduğumu hissetmek benim için önemlidir.							
6.	Bağımsız olmak benim için önemlidir.							
8.	Ödül beni tatmin ediyorsa çaba gösteririm.							
9.	Kendi yolumu izlemek benim için önemlidir.							
10.	Grup aktivitelerinden hoşlanırım.							
11.	Çoğunlukla merakımın beni yönlendirmesine izin veririm.							
12.	Zorlukların üstesinden gelmekten hoşlanırım.							
13.	Başkalarına yardım edebilirsem bu beni mutlu eder.							
14.	Hayatımdaki mevcut durumumu sorgulamaktan hoşlanırım.							
15.	Kurallara uymaktan hoşlanmam.							
16.	Bir problemi çözmeden bırakmak beni rahatsız eder.							
17.	Görevlerimi eksiksiz bir şekilde yerine getirmek benim için önemlidir.							
18.	Bir takımın parçası olmaktan hoşlanırım.							
19.	Kendimi asi biri olarak görürüm.							
20.	Diğer insanlarla etkileşim içinde olmak benim için önemlidir.							
22.	Bilgimi başkalarıyla paylaşmaktan hoşlanırım.							
23.	Zor görevleri başarmayı severim.							
24.	Ödüller benim için önemli bir motivasyon kaynağıdır.							

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 403–416

DOI: 10.21449/ijate.430720

Published at http://www.ijate.net http://dergipark.gov.tr/ijate Research Article

Evaluating Performance of Missing Data Imputation Methods in IRT Analyses

Ömür Kaya Kalkan [1,*], **Yusuf Kara** [2], **Hülya Kelecioğlu** [3]

[1] Pamukkale University, Education Faculty, Measurement and Evaluation Department, 20160, Denizli, Turkey
[2] Anadolu University, Education Faculty, Measurement and Evaluation Department, 26470, Eskişehir, Turkey
[3] Hacettepe University, Education Faculty, Measurement and Evaluation Department, 06800, Ankara, Turkey

Abstract: Missing data is a common problem in datasets that are obtained by administration of educational and psychological tests. It is widely known that existence of missing observations in data can lead to serious problems such as biased parameter estimates and inflation of standard errors. Most of the missing data imputation methods are focused on datasets containing continuous variables. However, it is very common to work with datasets that are made of dichotomous responses of individuals to a set of test items to which IRT models are fitted. This study compared the performances of missing data imputation methods that are IRT model-based imputation (MBI), Expectation-Maximization (EM), Multiple Imputation (MI), and Regression Imputation (RI). Parameter recoveries were evaluated by repetitive analyses that were conducted on samples that were drawn from an empirical large-scale dataset. Results showed that MBI outperformed other imputation methods in recovering item difficulty and mean of the ability parameters, especially with higher sample sizes. However, MI produced the best results in recovery of item discrimination parameters.

ARTICLE HISTORY

Received: 10 May 2018
Accepted: 04 June 2018

KEYWORDS

Missing Data,
IRT Model-Based Imputation
Multiple Imputation,
Expectation-Maximization,
Regression Imputation

1. INTRODUCTION

Majority of research data in psychology, sociology, and education are collected from human subjects. Missing data is one of the most important problems for researchers working in these fields. Missing data generally occurs in situations where participants do not respond to some items in data collection process due to lack of knowledge, hesitation, and lack of motivation or planned missing data designs, etc. (Enders, 2010; Finch, 2010; Graham, Taylor, Olchowski, & Cumsille, 2006; Sijtsma & van der Ark, 2003; Little & Rubin, 2002). It is difficult to conduct statistical analyses and calculate total scale scores in the presence of missing data. Moreover, existence of missing observations in data results in problems such as biased parameter estimates, inflation of standard errors, loss of information, and weak generalizability of results (De Leeuw, Hox, & Huisman, 2003; Dong & Peng, 2013; Finch, 2010; Rubin 1987; Schafer, 1997). While determining a proper method for handling missing observations, researchers need to address the rate, mechanism, and pattern of missing data

CONTACT: Ömür Kaya Kalkan ✉ kayakalkan@gmail.com Pamukkale University, Education Faculty, Measurement and Evaluation Department, 20160, Denizli, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

(Dong & Peng, 2013; Enders, 2010; Little & Rubin, 2002; Rubin, 1987; Schafer, 1997; Schafer & Graham, 2002).

1.1. Rate of Missing Data

Rate of missing data is the first issue that needs to be taken into account. Although there is not a certain criterion in the literature with regard to an acceptable rate of missing data in order to obtain valid inferences from a dataset, Tabachnick and Fidell (2007) and Schafer (1999) suggest that 5% or lower rate of missing data in a large sample would be inconsequential. On the other hand, Bennet (2001) suggests that parameter estimates through a statistical analysis are more likely to be biased, when more than 10% of data is missing. After addressing the rate of missing data, another issue that needs to be considered is the mechanism, namely how the missing data are distributed.

1.2. Missing Data Mechanism

Mechanism of missing data in a dataset can be Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) (Little & Rubin, 2002; Rubin, 1976). In MCAR case, missing data do not follow a specific pattern and variables containing missing observations do not depend on any other observed variables. Although Little (1988) proposed a test for detection of MCAR data mechanism, in practice, it is quite difficult to determine whether missing data are MCAR or not (Little & Schenker, 1995; Schlomer, Bauman, & Card, 2010). MCAR can also be regarded as a special case of MAR or a random sample of a complete dataset (Schafer & Graham, 2002). While omitting MCAR data does not lead to bias, standard errors of the point estimates increase since the sample size decreases (Dong & Peng, 2013). In case of MAR mechanism, missing observations in a variable are related to other variables in a dataset, rather than the latent level of the measured variable itself (Allison, 2001; De Leeuw, Hox, & Huisman, 2003; Enders, 2004, 2010). As a result, MCAR and MAR do not follow a specific pattern and missing observations are rather a result of randomness. In MNAR, however, missing data have a specific pattern and cannot be easily omitted because they are associated with the latent levels of the measured variable (Enders, 2004; Tabachnick & Fidell, 2007). Besides the aforementioned mechanisms, there are also planned missing data designs that are widely preferred by practitioners, especially in large-scale assessment studies. In such data collection designs, not all items are responded by all individuals, due to time and cost limitations. Interested readers are encouraged to see Enders (2010) and Graham et al. (2006) for more detailed information about planned missing data.

1.3. Patterns of Missing Data

Enders (2010) stated that missing data pattern simplistically explains position of the "holes" in a data matrix. There are three types of missing data patterns that are referred to as univariate, monotone, and arbitrary, respectively. If responses of individuals to a set of items have missing observations related to one or more variable(s), this indicates a univariate pattern of missing data. Monotone missing data pattern is often observed in longitudinal studies, where data are collected repeatedly from same individuals. If missing data are randomly observed for any individual in any of the variables, this indicates an arbitrary missing data pattern. Computation of a univariate or monotone missing data pattern is easier than an arbitrary pattern (Dong & Peng, 2013; Little & Rubin, 2002). Below, widely-used methods for handling various types of missing data are briefly explained.

1.4. Methods for Handling Missing Data

According to Little and Rubin (2002), most of the missing data handling methods can be classified into four general categories as follows: procedures based on completely recorded units, weighting procedures, imputation-based procedures, and model-based procedures.

Primitive methods like List-wise Deletion (LD) and Pair-wise Deletion (PD) simply suggest deletion of cases that have missing observations. LD, which proposes deletion of all cases that contain missing values for any variable, is not a recommended method in general, because it leads to loss of statistical power. It is also known that PD, which excludes cases that have missing values for variables covered, causes problems in correlation based calculations and multivariate analyses (Schlomer, Bauman, & Card, 2010). In sum, these two methods, which are based on exclusion of cases that have missing data, often lead to biased and inefficient parameter estimates (Rubin, 1987; Schafer, 1997).

Apart from deletion-based solutions, several methods have been developed for imputing the missing values by various techniques. Mean Substitution (MS) is one of the simplest imputation methods, where missing values are replaced with mean of the existing observations of the relevant variable. Although it sounds convenient to implement, MS results with reduced variance, which in turn leads to biased parameter estimates (Allison, 2001; Bennett, 2001; Graham, Cumsille, & Elek-Fisk, 2003; Pallant, 2007). Another relatively primitive method, which is called as Regression Imputation (RI), uses the variable(s) that does not contain any missing values as predictors in a regression model, in order to impute missing observations. Although it yields unbiased mean values in MCAR and MAR, RI leads to bias in variance and covariance (Graham, Cumsille, & Elek-Fisk, 2003).

There are more sophisticated imputation methods that have a wide range of usage including Multiple Imputation (MI), Expectation-Maximization (EM), Pattern-Matching Imputation (PMI), Hot-deck Imputation (HDI) and many more. In addition to RI, we only included MI and EM among more sophisticated imputation methods in this study. Thus, a brief information is provided below for MI and EM, as well as another method that is called as Model Based Imputation (MBI). For a detailed overview of the all mentioned imputation methods, interested readers can see Finch (2010), Little and Rubin (2002), Schafer and Graham (2002), Schlomer, Bauman, and Card (2010).

1.5. MI and EM

Rather than replacing missing observations directly, MI and EM combine the existing information obtained from data that meet some statistical assumptions to estimate the missing data mechanism. MI first creates a pre-specified number of copies of a dataset, in which missing observations are appropriately imputed. In other words, MI produces multiple datasets that have different imputed observations for missing cells. Then, parameters are estimated from each of the imputed datasets based on the model to be fitted. In the final step, multiply estimated parameters are averaged to provide a single estimate (Royston, 2004).

EM is an iterative procedure, where initial estimates of the missing values are imputed by a regression model that includes a random error term. In the following step, the covariance matrix and set of means is estimated. Then, previously obtained covariance matrix and means are used in another regression model for the estimation of missing observations. These steps (i.e., Expectation [E] and Maximization [M]) continue until the change in the covariance matrix reaches a minimum value. When this condition is satisfied, the iteration stops and final imputations are used for the desired analysis (Allison, 2001; Finch, 2008). It is important to note that with categorical data analysis, it is necessary to round these imputed values, as they will typically not be integers.

1.6. MBI

Briefly stated, MBI imputes the missing values based on statistical model that is applied to the data. There are various MBI methods that can be used for imputation of missing data in IRT applications. In one of them, Mellenbergh (2002) proposed a procedure based on computation of a π parameter by the response pattern, score, and frequency of the observed

data. Missing responses then imputed by randomly sampling an observation from a Bernoulli distribution with that parameter. Another method based on non-parametric IRT models proposed by Sijtsma and van der Ark (2003), and called as response function imputation (RFI). RFI requires computation of a fraction that is based on total score, summary score, and rest score (which is obtained by omitting the selected item from the total test score) for the imputation of missing observations. For a more detailed information, interested readers can see the referred studies.

MBI that is adopted in this study is rather a simpler procedure compared to mentioned approaches. Strictly speaking, we employed 3 parameter IRT model's equation for imputing missing observations by using the information from the complete data (i.e., dataset with no missing observations). Details about this procedure are elaborated in method section.

1.7. Purpose of the Current Study

A review of the literature reveals that the missing data itself and missing data handling methods in different sample sizes have a direct effect on the IRT model applications (Enders, 2004; Finch, 2008; Glas & Pimentel, 2008; Huisman, & Molenaar, 2001; Sijtsma & van der Ark, 2003). In addition, Dong and Peng (2013) underlie the fact that majority of the studies that have been conducted in the last decade do not contain sufficient information about missing data. Namely, they have not defined a certain approach in handling missing data, and have not tested assumptions about missing data methods.

Besides the limited reports about procedures of handling missing data issues in social science researches, majority of the proposed methods focus on continuous data. However, binary response patterns obtained from achievement tests that are administered to a group of students are one of the most common types of categorical data in educational sciences. Moreover, most of the widely used psychological tests are made of items that have multiple category response options (e.g., likert-type structured response scales). Common procedure for handling missing observations in such datasets is using the mentioned methods that are developed for continuous missing observations. Appropriateness of these imputation techniques for categorical missing data has not been clearly expressed in the literature. This study primarily aims to fill this gap by inspecting the performance of these methods in datasets that consist of binary response patterns to which IRT models are fitted. In addition to widely-used EM, MI, and RI methods, MBI was also evaluated in the study. Thus, the performance of the mentioned methods under different amounts of MCAR data were compared based on the recovery of item and ability parameter estimates. On the other hand, it is worth noting that there is a limited number of studies that use MBI for imputing dichotomous missing observations in datasets to be analyzed by IRT models. Therefore, we believe that this study provides significant contributions both to the existing literature and practitioners about handling missing data in IRT modeling applications.

2. METHOD

2.1. Data Set

Research data consist of binary (i.e., 0-1) coded responses of 480,691 candidates to 19 multiple choice items in the Turkish language test, which is a subtest of the 6th Grade Placement Exam (SBS) conducted by the Ministry of National Education of the Republic of Turkey in 2008. SBS is an annual nation-wide exam that measures achievement levels of 6th, 7th and 8th graders in Turkish, Maths, Science and Technology, and Social Science courses. Although 958,879 candidates took the exam in 2008, only a subset of 480,691 candidate's data were obtained with official permission. This raw dataset contained many missing observations, as was expected. Thus, a clean dataset of 306,757 individuals who answered all items (namely,

the dataset that didn't have any missing observations) was drawn from the mentioned raw dataset. We named this cleaned dataset as "full dataset" in order to prevent any conflicts with the raw dataset that includes missing cells. In that case, the full dataset can be realized as the population data for the sake of our recovery evaluation study. In other words, the "true" values of the item and ability parameters are estimates that would be obtained by fitting the relevant IRT model to the "full dataset", namely the population dataset.

Obtaining random samples of complete data sets. As a second step of data preparation, samples of 250, 500, 1,000, and 5,000 (25 replications for each sample size) were drawn repeatedly from the full dataset by simple random sampling employing SPSS (IBM, 2011). In order to determine the best fitting IRT model, -2LL values that were obtained by applying unidimensional 1PL, 2PL, and 3PL models to the full dataset were evaluated. The 3PL model was found to be the one that fits best among the three models. Therefore, unidimensional 3PL model was used in all subsequent analyses with the random samples that were drawn from the full dataset.

Obtaining datasets that contain missing observations. Only the missing data that would show MCAR were considered in this study. "missForest" (Stekhoven, 2016) package of R (R Core Team, 2016) was used to obtain datasets with MCAR data. Thereby, datasets with missing observations were obtained by deleting 5%, 10%, and 15% of the observations from complete datasets sampled in different sizes from the full dataset. Thus, 25 datasets with MCAR patterns were created for each of the 12 conditions (4 sample sizes x 3 missing data rates), leading to a total of 300 datasets.

2.2. Missing Data Imputations

The EM, MI, RI, and MBI based on unidimensional 3PL IRT model probability function were used to impute the missing observations. SPSS (IBM, 2011) was used to impute missing data through the EM, MI, and RI. Imputations with EM were performed based on normal distribution and the maximum number of iterations was 25 as default. This quantity specifies the number of iterations that is used to estimate the true covariance. When the specified number of iterations is reached, the procedure stops even if the estimates have not converged (IBM, 2014). Since the EM algorithm is an iterative process, decimal values are likely to be obtained in categorical data. Hence, all decimal values are rounded to nearest integers in order to end up with a categorical outcome (Finch, 2008).

Imputations based on MI were performed through the Markov chain Monte Carlo (MCMC) algorithm and maximum number of iterations was set to the default value of 10 as defined in SPSS (IBM, 2011). Maximum number of iterations taken by MCMC is used by fully conditional specification (FCS) method. FCS fits a univariate model using all other available variables in the model as predictors. The method then imputes missing values for the variable being fit for each iteration and each variable. This procedure continues until the maximum number of iterations is reached (IBM, 2014). Finally, FCS iteration history data are examined to evaluate the convergence of the model. For any missing observation in the missing datasets, as suggested by Royston and White (2011), the number of imputations was in accordance with the rate of missing data (5, 10, and 15). Thus, for a dataset containing 5% missing data, 5 different datasets were obtained by MI. Similarly, 10 and 15 different datasets were obtained for datasets containing 10% and 15% missing data, respectively. Each MI dataset for each missing rate (e.g., 5 MI datasets for 5% missing rate) was analyzed separately and results (e.g., item parameter estimates) were combined as

$$\bar{Q} = \sum_{m} \hat{Q}_m \, / \, N$$

where N is the number of unique analyses (Finch, 2008).

Imputation that was performed by RI was also accomplished using SPSS (IBM, 2011). RI in SPSS (IBM, 2011) uses multiple linear regression for estimating the missing observations (IBM, 2011, 2014). Residuals of that regression model are also included for adjustment of the estimates. Thus, error terms are drawn randomly from the observed residuals of complete cases and added to the regression estimates (IBM, 2014). Then, decimal values obtained for each missing observation are rounded to nearest integers, just as the final step of EM.

In imputations with MBI, 3PL IRT model was used because it provided the best model fit with the full dataset, as mentioned in the preceding section. The statistical equation of 3PL IRT model is shown below.

$$P(X_{is}=1|\theta_s,\beta_i,\alpha_i,\gamma_i)=\gamma_i+(1-\gamma_i)\frac{\exp[\alpha_i(\theta_s-\beta_i)]}{1+\exp[\alpha_i(\theta_s-\beta_i)]} \quad (1)$$

where;

X_{is}= response of person s to item i (0 or 1),

θ_s= trait level for person s,

β_i= difficulty of item i,

α_i= discrimination for item i,

γ_i= lower asymptote (guessing) for item i (Embretson & Reise 2000; Hambleton & Swaminathan 1985).

When imputing values by MBI, a two-step procedure is followed as follows: In the first step, datasets with missing observations that were crafted for each condition were analyzed based on 3PL model via BILOG controlled by the "irtoys" (Partchev, 2016) package of R (R Core Team, 2016). “NA” value was assigned to the missing cells of the datasets and these cells were handled as missing in the analyses (Partchev, 2016). In that case, item responses that were missing for an individual were not included in the estimation of item and ability parameters. Item and person parameter estimations were performed using ML and Marginal Maximum Likelihood (MML), respectively. In the second step, in order to impute the missing observations, the item and ability parameters obtained from the first step were used. Namely, for a missing response of an individual to an item, the probability was calculated by substituting the relevant item and ability parameters into equation 1. Then, if the calculated probability is equal to or greater than 0.5, “1” was placed into the missing item response for that individual. Otherwise, “0” was imputed for the relevant cell.

2.3. Evaluating the Accuracy of the Imputation Methods

A total of 48x25=1,200 analyses were performed with the 25 datasets crafted for each of the 48 conditions (4 sample sizes x 3 missing data rates x 4 imputation methods). Item and mean of the ability estimates were compared with the “true values” that were obtained from the population, namely from the “full dataset”. Population estimates of the item parameters that are considered as true values for recovery evaluation are provided in Table 1.

Table 1. Population estimates of item parameter and standard errors

	Parameter Estimate (SE)		
Item	a	b	c
1	3.312 (0.025)	-0.118 (0.004)	0.374 (0.002)
2	2.497 (0.018)	-0.320 (0.006)	0.304 (0.003)
3	1.209 (0.011)	-0.659 (0.019)	0.212 (0.008)
4	1.864 (0.012)	-0.823 (0.009)	0.150 (0.005)
5	2.038 (0.013)	-0.688 (0.007)	0.184 (0.004)
6	2.518 (0.015)	-0.666 (0.005)	0.170 (0.003)
7	2.528 (0.014)	-0.587 (0.005)	0.123 (0.003)
8	1.139 (0.006)	-1.170 (0.006)	0.003 (0.001)
9	1.585 (0.013)	0.480 (0.007)	0.205 (0.003)
10	1.904 (0.013)	-0.387 (0.007)	0.169 (0.003)
11	2.321 (0.016)	-1.467 (0.009)	0.117 (0.007)
12	1.889 (0.014)	-0.382 (0.008)	0.244 (0.004)
13	1.925 (0.015)	0.724 (0.005)	0.175 (0.002)
14	2.249 (0.016)	-0.638 (0.007)	0.297 (0.004)
15	1.627 (0.010)	-0.662 (0.009)	0.092 (0.005)
16	2.165 (0.014)	-0.311 (0.005)	0.184 (0.003)
17	2.690 (0.018)	-0.744 (0.006)	0.243 (0.004)
18	2.471 (0.022)	0.543 (0.005)	0.383 (0.002)
19	2.240 (0.015)	0.039 (0.005)	0.207 (0.002)

As can be inspected from Table 1, SE's of the population estimates are considerably lower than .05, which can be regarded as a sign of accurate estimation. Majority of the difficulty parameters are lower than 0, indicating a relatively easy test. Discrimination parameters are all higher than 1, which is a sign of good discriminating test for low and high ability groups of individuals. Guessing parameters on the other hand, varied between 0.003 and 0.383.

Due to varying number of the sample sizes for each condition, recovery of the ability parameters was evaluated based on their calculated means, which should be zero according to 3PL IRT models' identification constraint. Thus, we compared the mean of the estimated ability parameters from each condition by the mean of population estimates, which was calculated as 0.18 in the full dataset. Average of the absolute differences (AAD) among replications were calculated as a recovery indicator for ability parameters by the equation provided below.

$$AAD = \frac{\sum |\bar{\hat{\theta}}_m - 0.18|}{25} \quad (2)$$

$\bar{\hat{\theta}}_m$ is the mean of the estimated ability parameters at the *m*th step of the replications. Performance of the imputation methods based on recovery of the true item parameters were compared through RMSE (Root Mean Square Error) by the formula given below.

$$\text{RMSE} = \sqrt{\frac{\sum(\hat{\alpha}_m - \alpha)^2}{25}} \quad (3)$$

$\hat{\alpha}_m$ is the item parameter value estimated from imputed dataset at the *m*th step and α is the parameter value estimated from population, namely the so-called true value. RMSE for item parameters is regarded as indicator of the variability in the estimates. We reported the mean of the RMSE's for each type of item parameter. Thus, mean RMSE values for difficulty, discrimination, and guessing parameter, as well as, an AAD value for mean of the ability parameters were reported.

3. RESULTS

3.1. Results for Item Parameters

Recovery performance for discrimination, difficulty and guessing parameters are presented separately by tendency graphs of mean RMSE's below. Looking at the all figures for all item parameters, one can observe that mean RMSE's decreases for all methods by the increase of the sample size, with any rate of missing data. This is an expected improvement thanks to increase of available data that provides extra information for more accurate parameter estimations. On the contrary, as the rate of missing data increases, mean RMSE's for all methods also increases with any of the sample sizes. This phenomenon is also in line with the previous fact that we just mentioned. Namely, the amount of information lost increased due to increased number of missing observations, which in turn leads accuracy of the parameter estimates to decrease.

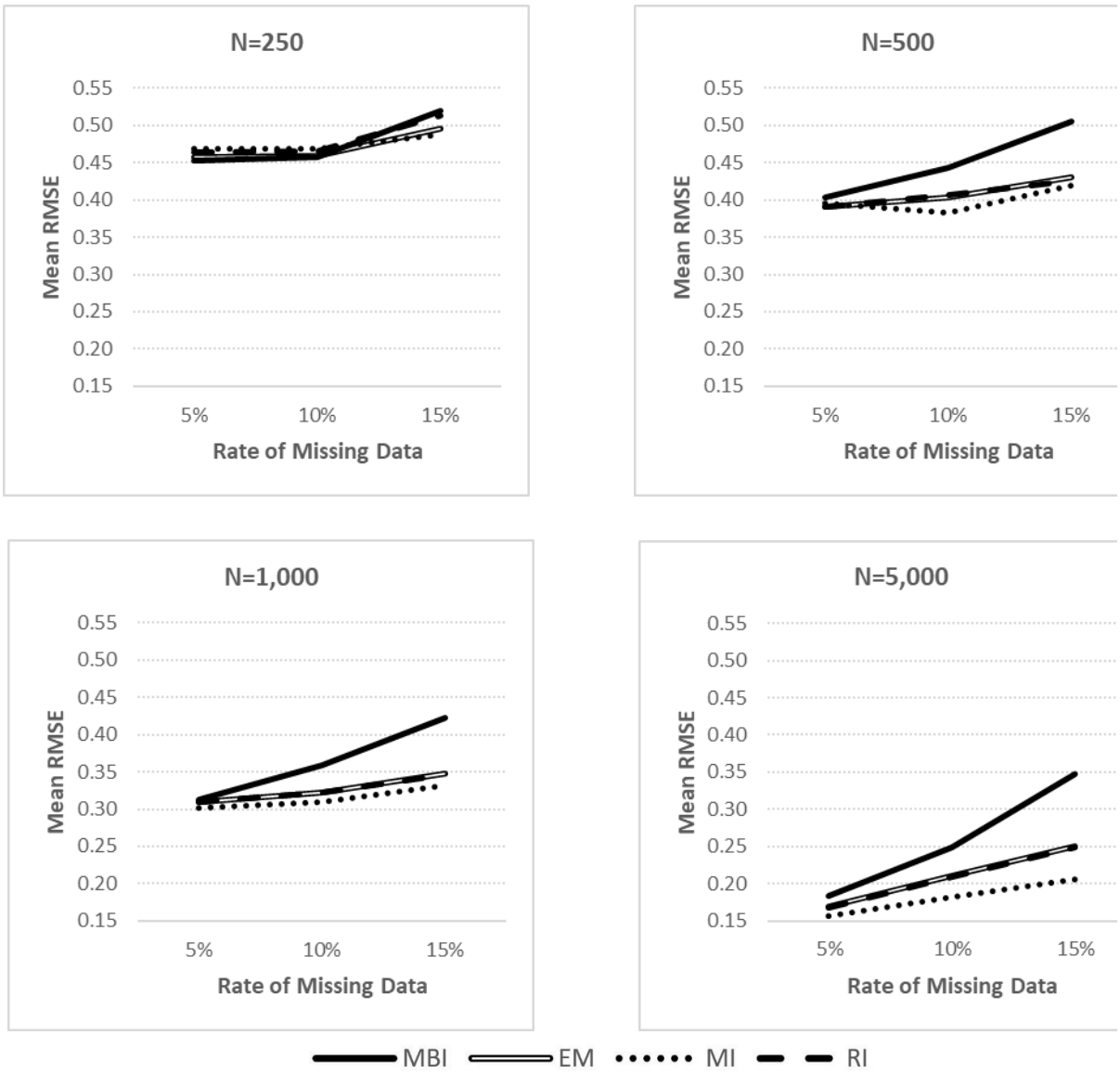


Figure 1. Mean RMSE values for item discrimination parameters

Looking at Figure 1 for discrimination recoveries, it can be observed that all imputation methods had nearly identical mean RMSE's when missing data rate is 5%, with all sample sizes. With N=250, all methods seemed to behave nearly in a similar fashion, in terms of mean

RMSE change. With all other sample sizes, EM and RI showed nearly undistinguishable trend of mean RMSE change. On the other hand, MBI apparently had higher mean RMSE's with all sample sizes (except for N=250) with all rates of missing data, while MI had slightly lower ones. Moreover, increase in mean RMSE was steeper for MBI, compared to all three methods, especially with three larger sample sizes.

Trend of the mean RMSE's for difficulty parameters are illustrated in Figure 2. Different from the previous findings for discriminations, MBI was the method that had the lowest mean RMSE's nearly in all conditions for item difficulty recoveries. Moreover, the increase in mean RMSE by the inflation of missing rates, was not as dramatic for MBI as other method's, especially with the two larger sample sizes.

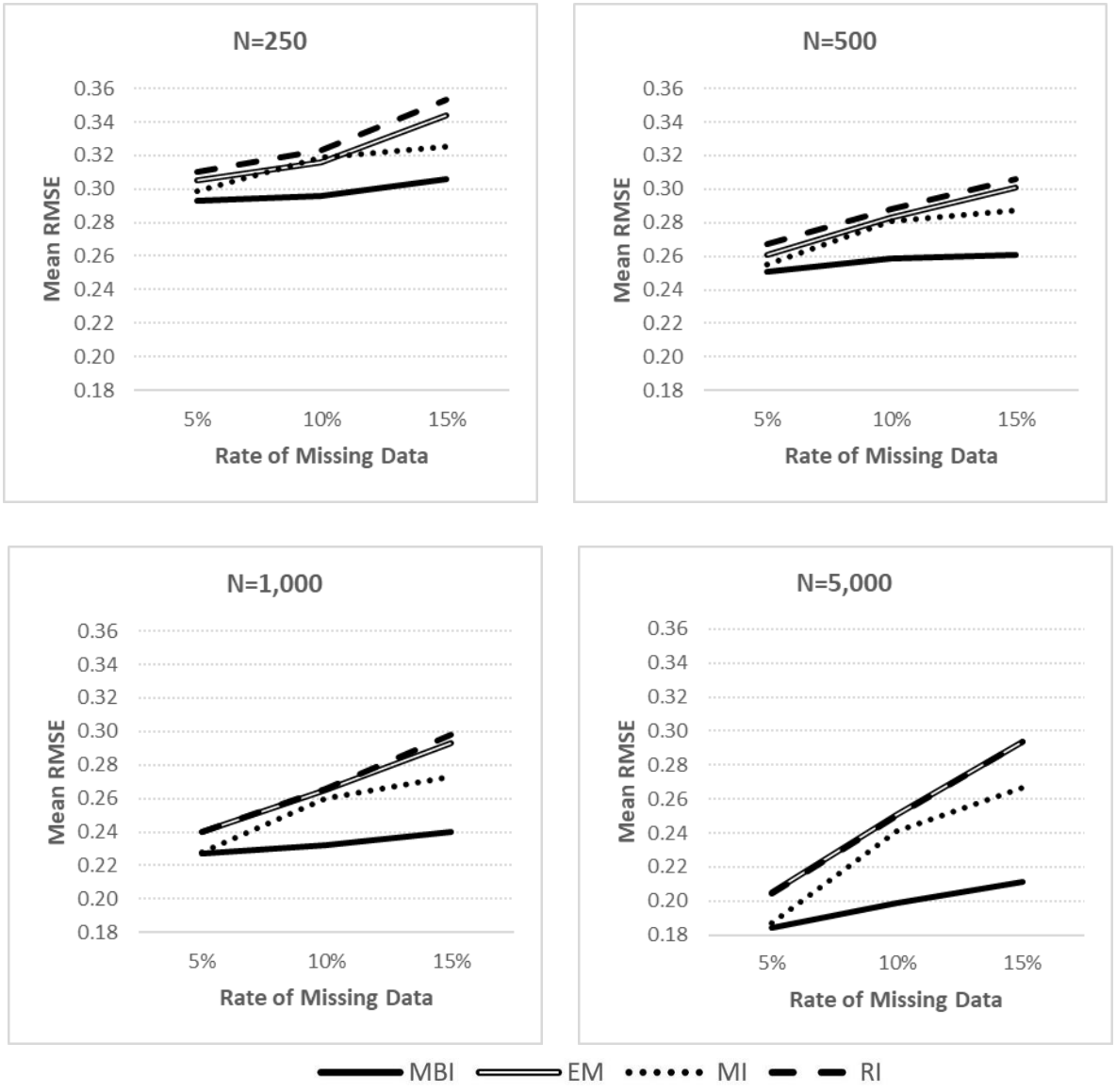


Figure 2. Mean RMSE values for item difficulty parameters

Similar to previous finding in discrimination parameters, EM and RI had nearly identical mean RMSE trend for the recovery of item difficulties.

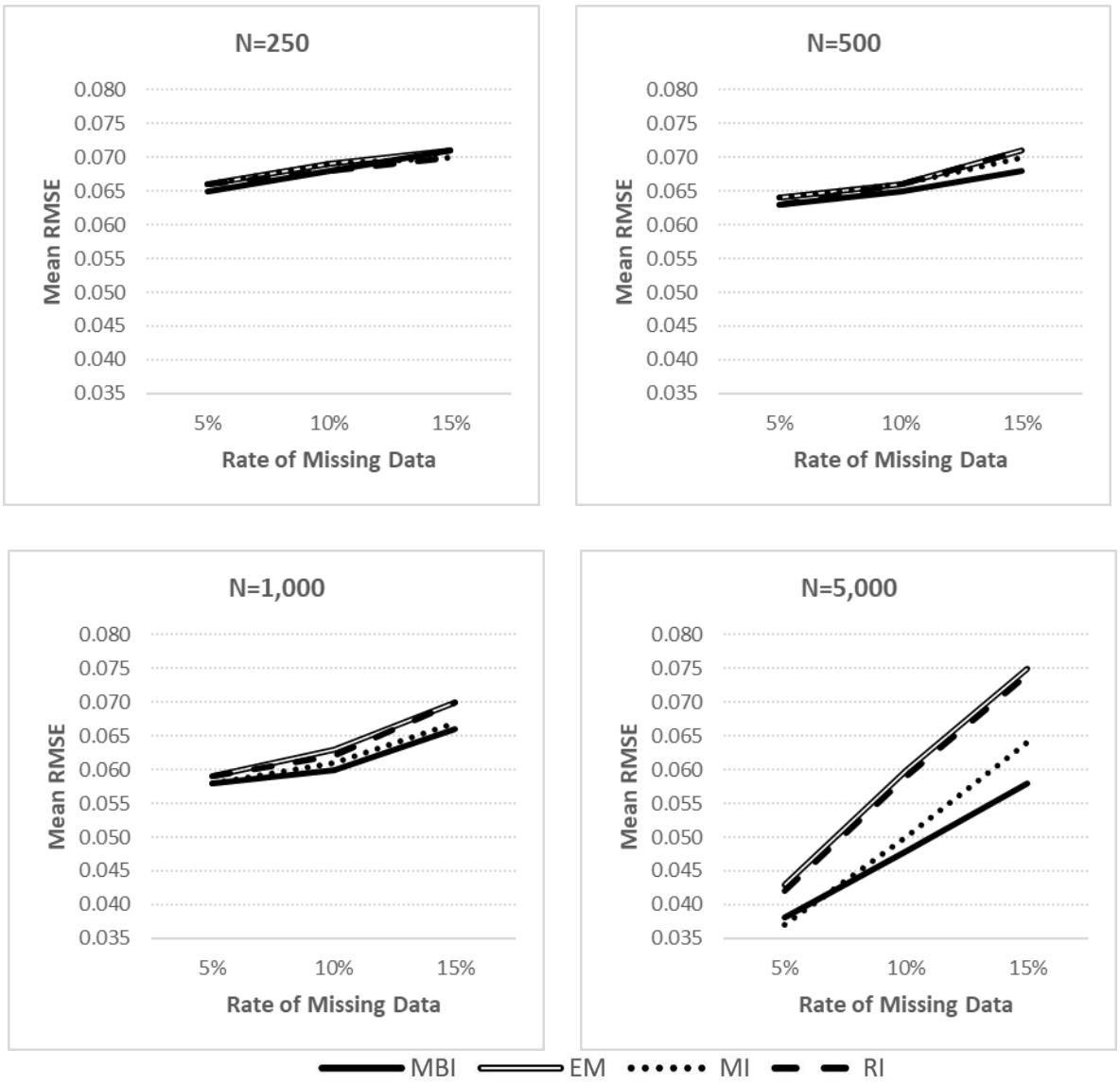


Figure 3. Mean RMSE values for guessing parameters

Trend of the mean RMSE's for guessing parameters are illustrated in Figure 3. Difference among methods is nearly indistinguishable for N=250, and it is quite similar for N=500 and N=1,000. For the largest sample size, differences among methods became more observable, except MI and RI, which behaved nearly identical. Moreover, MBI had relatively smaller mean RMSE's compared to other methods, especially with larger rates of missing data, when N=5,000.

3.2. Results for Ability Parameters

Graphs for the change of the averaged absolute differences (AAD) between means of the true and estimated ability parameters are presented in Figure 4. All methods had close AAD values with the lowest rate of missing data in all sample sizes. MBI had relatively smaller AADs with 10 and 15% rates of missing data. EM and RI again had nearly identical AAD trends, with N=1,000 and N=5,000. As a general inspection, it can be said that RI can be considered the method that had relatively higher AAD's compared to other ones. Last, it is also important to imply that the increase in AAD's was nearly linear when N=5,000 for all methods.

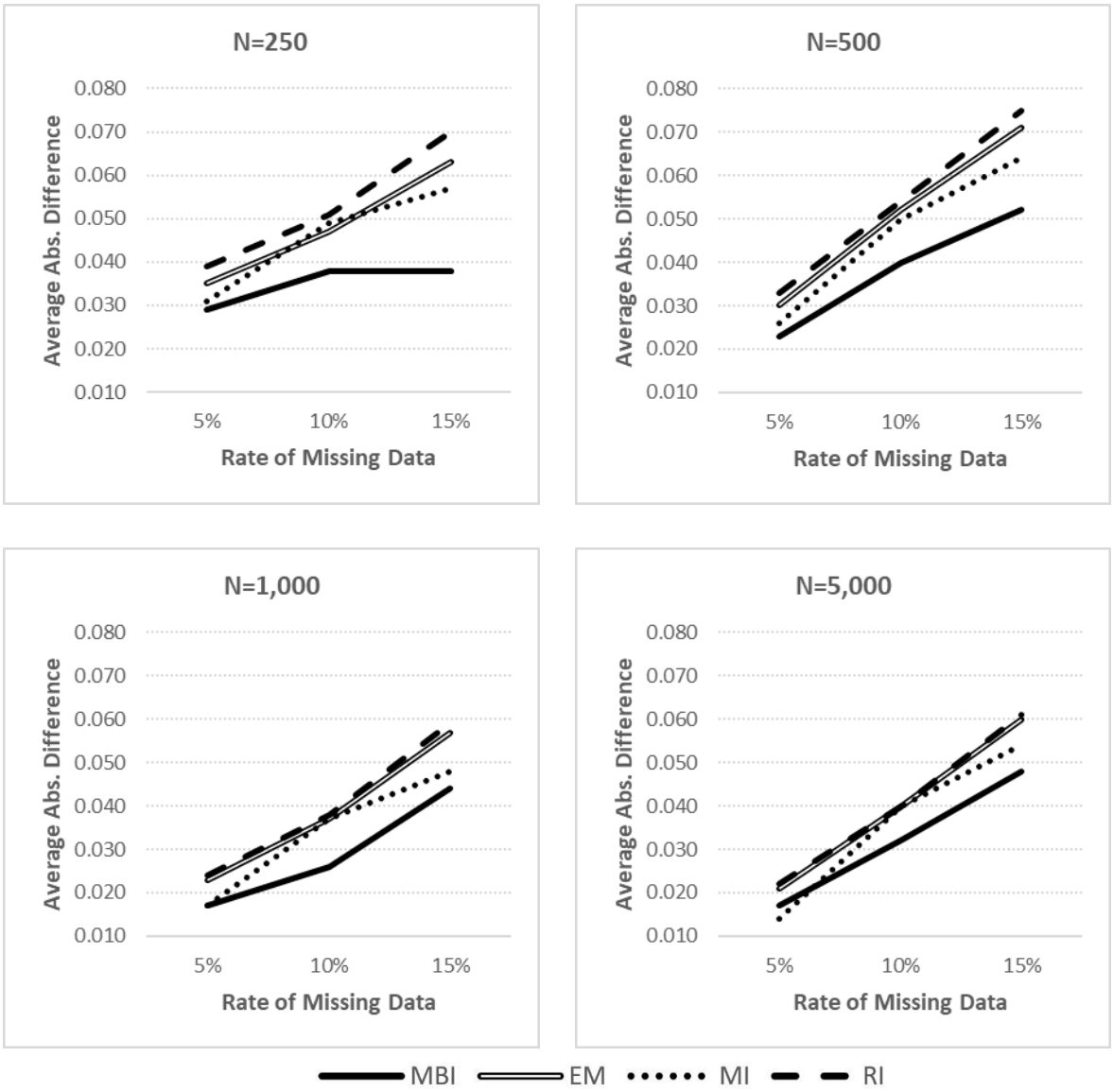


Figure 4. Average absolute difference values for the mean of the ability parameters

4. DISCUSSION AND CONCLUSION

In this study, the performance of various missing data imputation methods in terms of IRT models' parameter recovery were compared under different sample sizes and missing data rates. For the lowest rate of missing data (5%), it can be concluded that all methods showed close recovery performance for item and mean of the ability parameters in all sample sizes. Moreover, this similar performance occurred at lower levels of mean RMSE's for the two largest sample sizes. These conclusions in fact are parallel with Tabachnick and Fidell (2007). Namely, they stated that 5% or less MCAR data would not cause serious problems in large datasets. Moreover, they also stated that various missing data imputation methods are more likely to perform similarly in such conditions.

Methods showed different performances across different types of item parameters. Namely, MI generally was more robust to the increase of the missing data rate in recovery of the discrimination parameters. In fact, there are numerous studies which report that MI provides generally less-biased estimates compared to single-imputation methods. For example, Schlomer, Bauman, and Card (2010), and Schafer and Graham (2002) reported that the

accuracy of the parameter and standard error estimates makes MI the best choice for handling missing data. Similarly, it was also reported that MI produces unbiased parameter estimates in datasets that contain MAR data pattern (Peugh & Enders, 2004) and performs better than EM (Acock, 2005). Finch (2008) also verified MI's less-biased recovery performance in IRT model estimates, compared to other methods. MI's for discrimination parameters can also be attributed to performing imputation to each dataset according to the rate of missing data (Royston & White, 2011). EM and RI performed nearly identical and took the second place after MI in terms of recovery of the discrimination parameters, while MBI took the last place. It is important to note that these prominent implications were true especially for sample sizes 500 or larger, because all methods behaved quite similar with the smallest sample size N=250. As implied by Finch (2008), EM's relatively low recovery performance can be attributed to its assumption of normality, which is not the case for the dichotomous item responses.

When it comes to item difficulties, MBI showed the best recovery performance with all sample sizes that have missing data especially 10% and more. MI can be placed in the second order, while EM and RI again showed nearly identical behavior as the last two methods. For guessing parameters, all methods behaved nearly identical at all rates of missing data for the sample sizes of 250, 500 and 1,000. In fact, Finch (2008) reported that the estimation of the guessing parameters were not dramatically differed among the methods that he covered. The distinction among methods became more apparent with the maximum sample size of 5,000, especially for 10% or larger missing data. MBI again was slightly more robust to the increase of missing data, compared to other methods for the sample size of 5,000. MBI recovered the mean of the ability parameters better than other methods, with the 10% and 15% of missing data. For the 5% missingness, the distinction among methods was not that considerable.

Unfortunately, as also implied by other researchers (e.g., Finch, 2008; Smits, Mellenbergh, & Vorst, 2002) it is hard to recommend a single imputation method that will work in every scenario. Depending on our results, we can say that apart from the well-known methods that are MI, EM and RI, MBI can also be used as an alternative for dichotomous data containing missing observations in IRT model applications. MBI will particularly be more effective in scenarios where the ability and item difficulty estimations are more of interest. Thus, the use of MBI is expected to be more effective in Rasch and 1 PL models, where item discrimination parameters are not modeled. Nevertheless, if item discrimination parameters are of more interest, it would be more reasonable to prefer MI. Moreover, it is also important to imply that differences among the performances of MBI and MI was not so dramatic in many conditions. Thus, if two or three parameter models are desired and all model parameters would have the same level of interest for the practitioner, MI would be more rationale to adopt. Moreover, MBI requires long calculations for each of the missing observations. Considering this limitation of MBI, MI also offers a more reasonable solution in terms of efficiency.

There are some further research suggestions that we draw depending on the results of that study. First, evaluating the performance of MBI with one and two parameter IRT models and comparing the results with the current study is thought to be worth further investigation. In addition, the performance of MBI can also be examined in multidimensional datasets that contain missing data. Last, the consequences of assigning 0 to the missing observations (treating missing observations as false) rather than handling them as missing can also be explored for MBI.

Acknowledgements

We declare that a part of this study was presented as an oral presentation at the IIIrd International Eurasian Educational Research Congress (EJER, 2016) held on 31 May 03 June 2016 in Muğla, Turkey.

ORCID

Ömür Kaya Kalkan https://orcid.org/0000-0001-7088-4268
Hülya Kelecioğlu https://orcid.org/0000-0002-0741-9934

5. REFERENCES

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family, 67(4),* 1012-1028.

Allison, P. D. (2001). *Missing data*. Thousand Oaks, *CA, US: Sage publications*.

Bennett, D. A. (2001). How can i deal with missing data in my study? *Australian and New Zealand journal of public health*, *25*(5), 464-469.

De Leeuw, E. D., Hox, J., & Husman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics, 19*(2), 153-176.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222.

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

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and psychological measurement*, *64*(3), 419-436.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, *45*(3), 225-245.

Finch, W. H. (2010). Imputation methods for missing categorical questionnaire data: A comparison of approaches. *Journal of Data Science*, *8*(3), 361-378.

Glas, C. A., & Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educational and Psychological Measurement*, *68*(6), 907-922.

Graham J.W., Cumsille, P.E., & Elek-Fisk, E. (2003). *Methods for handling missing data.* In Research Methods in Psychology, ed. JA Schinka, WF Velicer, pp. 87–114. Volume 2 of Handbook of Psychology, ed. IB Weiner. New York: Wiley

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological methods*, *11*(4), 323.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications* (Vol. 7). Boston, MA: Kluwer-Nijhoff Publishing.

Huisman, M., & Molenaar, I. W. (2001). Imputation of missing scale data with item response models. In *Essays on item response theory* (pp. 221-244). Springer, New York, NY.

IBM. (2011). IBM SPSS statistics base 20. *Chicago, IL: SPSS Inc*.

IBM (2011). IBM SPSS missing values 20. Retrieved January 24, 2018, from https://www.csun.edu/sites/default/files/missing-values20-64bit.pdf

IBM (2014). IBM SPSS missing values 22. Retrieved January 24, 2018, from http://www.sussex.ac.uk/its/pdfs/SPSS_Missing_Values_22.pdf

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley. New York.

Little, R. J., & Schenker, N. (1995). *Missing data*. In Handbook of statistical modeling for the social and behavioral sciences (pp. 39-75). Springer US.

Mellenbergh, G. J. (2002). *Measurement model-based imputation of missing item responses*. Unpublished manuscript.

Pallant, J. (2007). *SPSS survival manual* (3rd ed.). New York, NY: Open University Press.

Partchev, I. (2016). Irtoys: simple interface to the estimation and plotting of IRT models. *R package version 0.2.0*

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research*, *74*(4), 525-556.

R Core Team. (2016). R: A language and environment for statistical computing [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Royston, P. (2004). Multiple imputation of missing values. *Stata journal*, *4*(3), 227-41.

Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): implementation in Stata. *Journal of Statistical Software*, *45*(4), 1-20.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592.

Rubin, R. B. (1987). *Multiple imputation for nonresponse in surveys* (J Wiley & Sons, New York, NY).

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Schafer, J. L. (1999). Multiple imputation: a primer. *Stat Methods in Med 8*(1), 3–15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, *7*(2), 147.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling psychology*, *57*(1), 1.

Sijtsma, K., & Van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, *38*(4), 505-528.

Smits, N., Mellenbergh, G. J., & Vorst, H. (2002). Alternative missing data techniques to grade point average: Imputing unavailable grades. *Journal of Educational Measurement*, *39*(3), 187-206.

Stekhoven, D. J. (2016). MissForest: nonparametric missing value imputation using random forest. *R package version 1.4.*

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statisti*cs. Pearson Education. Boston, MA.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 417–427

DOI: 10.21449/ijate.409826

Published at http://www.ijate.net http://dergipark.gov.tr/ijate Research Article

A Comparison of Traditional and Kernel Equating Methods

Çiğdem Akın Arıkan [1], **Selahattin Gelbal** [2,*]

[1]Ordu University, Education Faculty, Measurement and Evaluation Department, 52200, Ordu, Turkey
[2]Hacettepe University, Education Faculty, Measurement and Evaluation Department, 06800, Ankara, Turkey

Abstract: In this study, the equated score results of the kernel equating (KE) method compared with the results of traditional equating methods—equipercentile and linear equating and 9th grade 2009 ÖBBS Form B of Social Sciences and 2009 ÖBBS Form D of Social Sciences was used under an equivalent groups (EG) design. Study sample consists of 16.249 students taking booklets B and another 16.327 students taking D in that test. The analysis of the test forms was carried out in four steps. First, descriptive statistics were calculated for the data and then it was checked whether the data obtained from the two booklets satisfy the equating conditions. In the second step, the booklets were equated according to methods. Lastly, the errors for each equating methods were calculated. Kernel equating results were nearly same to the results from the corresponding traditional equating methods. In Kernel equating, when parameter h was selected as optimal, equated scores provided almost identical results as traditional equipercentile equating. When it was selected large, this time the equated scores provided results almost identical to traditional linear equating. It is concluded that Kernel equating methods are relatively more the most appropriate equating method method than traditional equating methods.

ARTICLE HISTORY

Received: 26 March 2018
Revised: 13 May 2018
Accepted: 29 May 2018

KEYWORDS

Kernel equating,
Traditional equating Methods,
The equivalent groups design

1. INTRODUCTION

Comparison of test scores obtained from different forms has been a centre of attention for psychometrics for nearly a century. Further, discovery of methods for comparing scores is almost as old as the field of psychometrics itself (Holland, 2007). Discussion and development on the topic of comparison of scores were started in 1910s by Otis (1916, 1918), Kelley (1914), Starch (1913), Weiss (1914) and Pinter (1914) and have not come to an end up to day (Holland, 2007). In order to compare scores from different forms, various equating methods have been introduced which are based on different theories and statistics. The purpose of all equating methods is to compare scores collected from different test forms. Although different test forms are developed with similar content and statistics, they may vary in difficulty, so test equating is needed. The statistical process which allows comparing and interchangeable use of scores

CONTACT: Çiğdem AKIN ARIKAN ✉ akincgdm@gmail.com Ordu University, Education Faculty, Measurement and Evaluation Department, 52200, Ordu, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

obtained from different test forms is called equating (Holland, 2007; Kolen & Brennan, 2004). There are certain conditions that must be met for equating test forms. These conditions include equality, symmetry, group invariance, equal construct and equal reliability (Dorans and Holland, 2000).

- Equal construct: Both tests must be measure of the same characteristics, trait or skills.
- Equal reliability: The tests must have the same level of reliability.
- Symmetry: In equating function, form X can be transformed into form Y and vice versa at the same time.
- Equality: Lack of difference resulting of taking form X or Y of individuals.
- Group invariance: Equating relationship is independent of groups.

After meeting the equating conditions, test equating design is selected, one of the most important steps of test form equating. Equating designs are divided into three as single-group design, equivalent groups design, and common-item test design. In single-group design, the oldest and the simplest equating design, the same individuals are given both test forms. Since forms are answered by the same individuals, no error emerges due to ability levels of individuals (Kolen & Brennan, 2004). In equivalent groups design, the test forms are divided into two randomly and administered separately to the groups regarded equivalent. In this design, individuals in each group responds to only one test (Livingston, 2014). However, since each of the groups consists of different individuals, the difference of distribution of individuals' ability is reflected as bias in equating (Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004). In common-item test design, two tests consist of different questions and groups are given either test which contain common items. On the common items in the forms, equating is performed. For this reason, common items should be selected in a way to represent the test properly (Kolen, 1988; Kolen & Brennan, 2004). Once the equating design is determined, decision is made regarding which equating method to use. Equating methods are divided as true-score equating and observed-score equating. While true-score equating relates to equating methods based on the Item Response Theory; whereas observed-score equating includes linear, average, equipercentile equating, IRT observed-score equating, and Kernel equating, a new approach. Present study was conducted with linear equating, equipercentile equating, Kernel linear equating and Kernel equipercentile equating methods among observed-score equating methods.

In linear equating, it is assumed that all properties except the mean and standard deviation of test forms are equal (Kolen & Brennan, 2004). In addition, the difference in difficulty between test forms varies in a constant amount through the entire score scale (Albano, 2016). Concerning X new form and Y reference (old) form, , μ_X ve μ_Y gives means of the forms, and S_X ve S_Y gives standard deviations of the forms. On this basis, linear equating is obtained from the following equation.

$$Y = \frac{S_Y}{S_X}x + \left[\mu_Y - \frac{S_Y}{S_X}\mu_X\right] \quad (1)$$

In the case of equipercentile equating, cumulative frequency of each test form is first calculated and then the scores corresponding to the same percentile based on the frequencies are equated. When the forms are equated with equipercentile equating, the scores obtained from the forms are similar in mean, standard deviation, and distribution (kurtosis, skewness, etc.) (Kolen, 1988). However, Livingston (2014) stated that, in equipercentile equating, equated points obtained from the new form have almost identical distribution with the points on the reference form, and that the distributions are not identical because the scores are discrete.

In equivalent groups design, both linear and equipercentile equating methods can be used. While equipercentile equating methods use score distribution curves to explain difficulty

difference between forms, linear equating methods use linear estimates. Hence equipercentile equating is more general than linear equating methods (Kolen & Brennan, 2004).

As for the Kernel equating, an observed-score equating method, was defined by Holland and Thayer (1989) and improved by von Davier, Holland and Thayer (2004). In traditional equipercentile equating methods, cut-off score distribution is made continuous by using linear estimates. On the other hand, Kernel equating employs the Gauss Kernel approach after which it is also named. In the latter, discrete distributions are made continuous so that scores are equated on the basis of the continuous distributions (Lee & von Davier, 2011; Ricker & von Davier, 2007). Kernel equating also includes linear equating functions even though it is an equipercentile equating function (von Davier et al. 2004). One of the important parameters in the Kernel is the parameter h, which is the continuation parameter. If the parameter h is chosen as ideal, equipercentile equating is obtained; but if it is selected as large, linear equating function is obtained. In addition, the results obtained through the ideal h parameter approach traditional linear equating, while those obtained through the large h parameter approach traditional equipercentile equating.

In the Kernel model, test forms are equated in five steps:

Presmoothing: It refers to using the log-linear statistical model for smoothing of score distributions. In this step, estimation of score probabilities varies depending on the score equating design. Equivalent groups design is a univariate distribution; however, common-item test design is a bivariate distribution in nonequivalent groups. von Davier et al. (2004) indicated four statistical properties in selection of estimating point probabilities as;

- **Consistency;** as the sample size increases, estimated values approach the population parameter.
- **Efficiency;** deviation of the score probabilities estimated from the population values is at the minimum level possible.
- **Positivity;** score probabilities estimated for each score are positive.
- **Integrity;** smoothed score distributions match with observed score distribution. To get good fit in univariate distributions, five or six moments of test forms must be used (von Davier et al., 2004).

Estimation of score probabilities according to the equating design Gauss Kernel approach is used to make the cut-off score distributions continuous at the relevant stage. Still, Lee and von Davier (2011) suggested logistics and uniform kernel approaches as alternatives. Test forms are equated by using the continuous distributions obtained in the previous step. It is calculation of the standard error of equating and difference of the standard error of equating. Kernel equating can be used in common-test designs with single-group, equivalent groups, balanced group, and nonequivalent groups (von Davier et al., 2004).

It is used for not only international tests such as PISA and TIMMS but also nation-wide examinations held by the Ministry of National Education and Student Selection and Placement Centre in Turkey. Student Achievement Determination Exam at National Scale (SADE-ÖBBS) is among them. ÖBBS has taken every three years since 1992 at elementary education. It applies to different areas and grade levels. Then, the applications were reported in 2002, 2005 and 2008, respectively. ÖBSS helps determine the adequacy levels of the education and instruction environment offered to students, objectives and skills and make an assessment so as report to relevant authorities accordingly (EARGED, 2010). It was last held in year 2008 to determine achievement levels of elementary students in subjects such as Turkish language, science and technology, mathematics, social studies and English language. Then, in 2009, the Secondary Education Development Project was launched to identify achievements of students in lessons, monitor their progress, and propose recommendations according to results. The

project was implemented on the 9th and 10th grade students (EARGED, 2010). In this scope, in order to ensure test security in practice and to avoid violations of the rules; booklets A, B, C and D were prepared which are very similar in both scope and difficulty. However, these booklets were not equated although they consist of different questions. Even though test developers introduce test forms similar in content and statistical aspects, these forms may differ in difficulty.

This study was conducted on the scores obtained from the ÖBBS given to 9th graders in 2009. In particular, the scores obtained from social studies test in booklet B were equated to booklet D by using linear equating, equipercentile equating, Kernel linear equating and Kernel equipercentile equating methods. It was aimed to find out the best method as a result. The other aim of the study is to compare results obtained from the Kernel equating and traditional equating methods.

2. METHOD

2.1. Population and Sample

The population of the study is composed of the elementary 10 th grade students who took the 2009 ÖBBS covering Turkish Literature - Language and Expression, Mathematics, Sciences, Social Studies and English Language tests. There are four booklets as A, B, C and D, for each of the tests in ÖBSS. The booklets were put into two pairs by arranging A and C, and B and D in parallel, respectively (EARGED, 2010). Study sample consists of 16.249 students taking booklets B and another 16.327 students taking D in that test.

Data is constituted by scores obtained from social studies test as a part of ÖBBS held in 2009. The test contains 15 (fifteen) questions which target history and geography lessons. Each booklet contains different questions in ÖBBS. So, there are two pairs of booklets considering the traits they measure resulting in booklets A and C as a pair and B and D as another pair. This study was carried out on the latter pair of booklets, B and D.

2.2. Equating Design and Data Analysis

Despite containing different questions, the booklets were prepared in parallel and applied spirally to the students. For this reason, equivalent groups were formed for B and D booklets at random. In this study, the social studies test was equated through the use of the equating pattern for equivalent groups.

The analysis of the test forms was carried out in four steps. First, descriptive statistics were calculated for the data and then it was checked whether the data obtained from the two booklets satisfy the equating conditions. In the second phase, the two booklets were equated according to the equivalent group design and equated according to the methods. In the second step, the booklets were equated with equivalent groups design and equated socres were gained accordingly. Lastly, the amount of error resulting from each equating method was calculated. Analysis of the data was done with SPSS and FACTOR (Lorenzo-Seva & Ferrando, 2006). The equate R package (Albano, 2016) was used for traditional equating methods analyses and the kequate R package (Andersson, Branberg & Wiberg, 2013) was used for kernel equating methods analyses (R Core Team, 2017).

Step I: Descriptive Statistics and Testing of Equating Conditions

At first, descriptive statistics of the social studies tests were calculated. The findings are given in Table 1. According to Table 1, for the booklets B and D, score distributions exhibit positive coefficients of skewness below 1. On the other hand, kurtosis and skewness coefficients are negative and smaller than 1 in score distribution of both forms. Büyüköztürk (2007) stated that kurtosis and skewness coefficients between -1 and +1 refer to normal distribution.

Table 1. Descriptive statistics

Descriptive Statistics	Booklet B	Booklet D
N	16249	16327
Mean	7.60	7.97
Standard Deviation	3.50	3.52
Variance	12.28	12.45
Skewness	.172	.050
Skewness Standard Error	.019	.019
Kurtosis	-.81	-.89
Kurtosis Standard Error	.038	.038

Test forms can be equated provided that certain requirements are satisfied. Dorans and Holland (2000) list these requirements as unidimensionality, equal reliability, and similar difficulty. To confirm unidimensionality of the data, principal components factor analysis based on the tetrachloric correlation used for two-category data was performed. The factor analysis was conducted with FACTOR 10.7 (2017) program.

Table 2. Factor Analysis Results for Booklets B and D

	Booklet B		Booklet D	
Component	Eigenvalue	V.A.O (%)	Eigenvalue	V.A.O (%)
1	4.91	32.7	5.02	33.5
2	0.96	6.4	0.99	6.6

In Table 2, when factors with eigenvalue greater than 1 are taken to calculate the number of factors, there is only one factor with eigenvalue greater than 1 for both booklets. Explanatory variance of the first factor is 32.7% for booklet B, while it is 33.5% for the other booklet. In single-factor scales, explained variance ratio at and above 30% is regarded adequate (Büyüköztürk, 2007). So, both booklets can be said to have one single general factor.

In order to test whether reliability of booklets B and D is equal, the reliability coefficient of KR-20 was calculated. Again, reliability coefficients were accepted as the correlation coefficients and Fischer's Z transformation was performed to check if there is a difference between the two reliability coefficients (Akhun, 1984). The results are presented in Table 3.

Table 3. Comparison Results of Reliability of Booklets

Booklet	KR-20	Zr	Z	p
B	0.75	0.99	1.29	0.51
D	0.76	1.00		

Table 3 suggests that the booklets B and D meet the prerequisite of equal reliability ($p>.05$). The difference between average difficulties of the booklets B and D was examined with two ratio difference test (Baykul, 1996), indicating no significant difference between difficulty levels of the two booklets ($p>.05$). It can be suggested that both booklets are equal in average difficulty.

Step II: Booklet equating: The booklets B and D were equated by using linear and equipercentile equating, Kernel equating method (linear and equipercentile) among traditional equating methods in equivalent grups design.

Step III: At this stage, the weighted error squares mean (WMSE) and RMSD (Root Mean Squared Difference) were calculated for evaluating the errors randomly involved in the

equating process and MSD (Mean Signed Difference) indices were calculated for evaluating the systematic errors.

Below are shown the equations for calculating WMSE, RMSD and MSD coefficients:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{k-1} fi(X_E - X_{Crit})^2}{\sum_{i=1}^{k} f_i}} \quad (2)$$

$$\text{WMSE} = \frac{\sum_{i=1}^{k-1} fi(X_E - X_{Crit})^2}{\sum_{i=1}^{k} f_i \, S^2 y} \quad (3)$$

$$\text{MSD} = \frac{1}{N} \sum_{j=1}^{N} (X_E - X_{crit}) \quad (4)$$

3. FINDINGS

In 2009 ÖBBS social studies test, booklet B was equated with booklet D by using linear equating, equipercentile equating, Kernel linear equating and Kernel equipercentile equating methods. The parameter h, which is the continuation parameter in the kernel equating methods, was selected by the kequate package. It was found to be hx=0.539 and hy=0.538 for Kernel equipercentile equating method; and hx=3503.590 and hy=3530.941 for Kernel linear equating method. Equated scores obtained from the equating methods and raw scores are given in Table 4.

Table 4. Equivalent scores of Booklet D corresponding to raw scores in Booklet B

Booklet B	LE	Dif.	EE	Diff.	KE-LE	Diff.	KE-EE	Dif.
0	0.31	0.31	0.01	0.01	0.31	0.31	0.02	0.02
1	1.32	0.32	1.12	0.12	1.32	0.32	1.11	0.11
2	2.33	0.33	2.19	0.19	2.33	0.33	2.18	0.18
3	3.34	0.34	3.24	0.24	3.34	0.34	3.23	0.23
4	4.35	0.35	4.29	0.29	4.34	0.34	4.28	0.28
5	5.35	0.35	5.31	0.31	5.35	0.35	5.31	0.31
6	6.36	0.36	6.41	0.41	6.36	0.36	6.40	0.40
7	7.37	0.37	7.50	0.50	7.37	0.37	7.49	0.49
8	8.38	0.38	8.53	0.53	8.37	0.37	8.52	0.52
9	9.38	0.38	9.54	0.54	9.38	0.38	9.54	0.54
10	10.39	0.39	10.53	0.53	10.39	0.39	10.53	0.53
11	11.40	0.40	11.45	0.45	11.4	0.40	11.47	0.47
12	12.41	0.41	12.35	0.35	12.41	0.41	12.35	0.35
13	13.41	0.41	13.21	0.21	13.41	0.41	13.21	0.21
14	14.42	0.42	14.10	0.10	14.42	0.42	14.11	0.11
15	15.43	0.43	15.02	0.02	15.43	0.43	15.04	0.04

**Dıff: difference; LE: Linear Equating; EQ: Equipercentile Equating; KE-LE: Kernel Linear Equating; KE-EQ: Kernel Equipercentile Equating*

According to Table 4, the raw scores belonging to Booklet B take values in the range of 0-15 points, but the equated scores obtained through traditional linear equating and Kernel linear equating methods vary from 0.31 to 15.43. The equated scores obtained from both methods demonstrate that the equated scores are the same except for a few conditions. As a result of the linear equating methods, the scores equated throughout the entire score scale were

found larger than the raw scores and difficulty level did not vary throughout the scale. So, it can be said that booklet B is more difficult than D. The equated scores obtained from traditional equipercentile equating method were found to be between 0.01 and 15.02, while Kernel equipercentile equating method yielded results in the range of 0.02 to 15.04. The two methods were seen to generate the same equated scores except in a few conditions. It was found out that the scores equated with equipercentile equating methods were greater than the raw scores throughout the entire score scale and difficulty level did not vary throughout the scale. This finding suggests that booklet B is more difficult than booklet D according to equipercentile methods. Moreover, a linear relationship was detected between raw scores from booklet B and equated scores obtained through both linear equating methods and equipercentile equating methods. Such relationships are shown in Figure 1 and Figure 2.

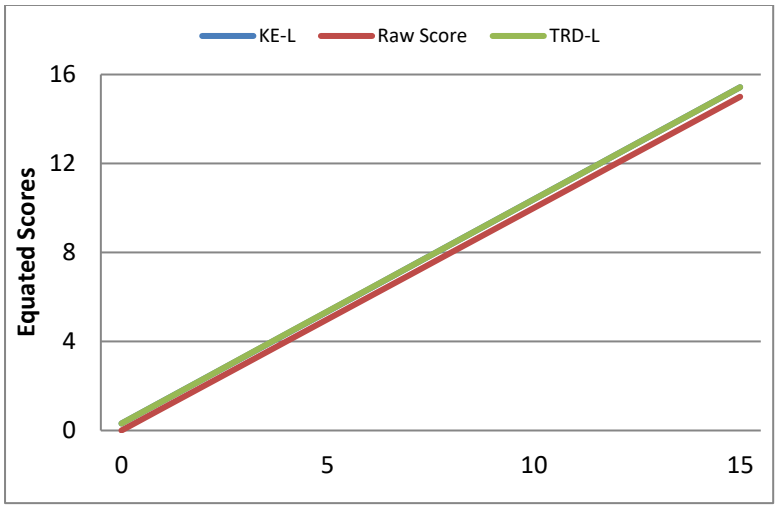


Figure 1. Raw scores in booklet B and equated scores based on linear equating methods

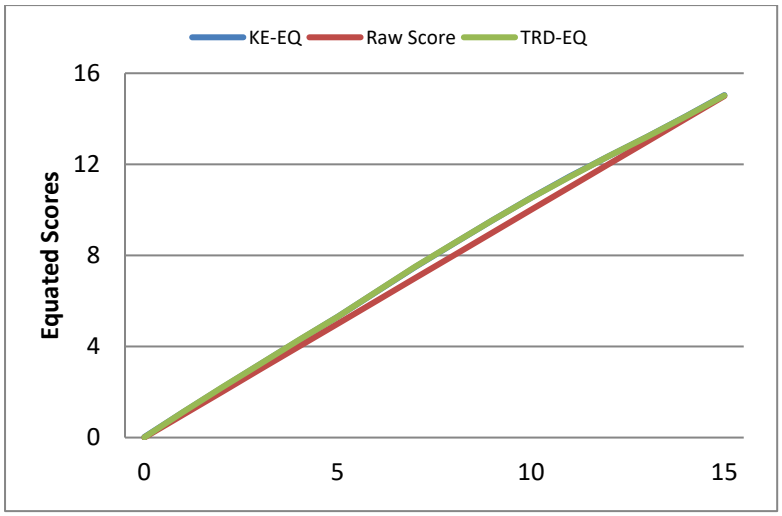


Figure 2. Raw scores in booklet B and equated scores based on equipercentile equating methods

The graphs above indicate almost the same difference between the equated scores based on linear equating methods and raw scores throughout the whole scale. In the case of equipercentile equating methods; while the difference between equated scores and raw scores is smaller in extreme scores, the difference increases in in the medium score scale (in the range of 5 to 12). One superiority of Kernel equating to traditional equating methods is that it

calculates the standard error of equating for each score. Figure 3 displays the graph for the standard error of the equating obtained over the entire scale score according to Kernel equipercentile and Kernel linear equating methods.

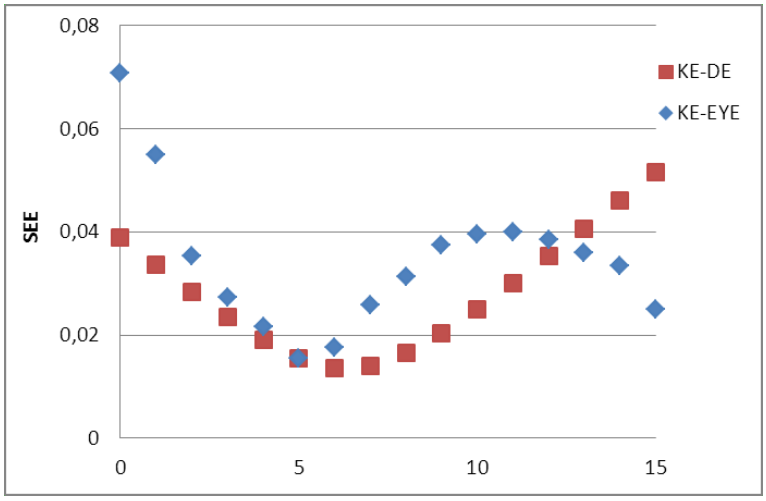


Figure 3. Standard errors from Kernel linear and equipercentile equating methods

Figure 3 shows that linear equating methods yield higher standard errors in marginal ends but lower standard errors in intermediate band. In the case of equipercentile equating method, standard error proved high in lower end but low in higher end. Also in linear equating, it has a decreasing tendency from 0 to 7 points but increasing from 7 to 15 points. In equipercentile equating, it again shows a decreasing tendency from 0 towards 5 points while increasing from 5 towards 11 and then falling from 11 to 15 back. The average standard error was calculated as 0.035 and 0.028 for Kernel equipercentile equating and linear equating, respectively, which shows a smaller error in linear equating method.

The booklets B and D for social studies test from 2009 ÖBBS were were equated with linear and equipercentile equating methods. WMSE, MSD and RMSD were calculated to find out which of the equating methods includes a lower level of error. The obtained values are given in Table 5.

Table 5. WMSE, RMSD and MSD values from the equating methods

Equating Method	WMSE	RMSD	MSD
GE-EYE	0.013	0.400	0.373
GE-DOĞ	0.011	0.377	0.376
KE-EYE	0.012	0.399	0.372
KE-DOĞ	0.011	0.375	0.374

Table 5 shows the smallest WMSE and RMSD values as a result of Kernel linear equating method, while the largest WMSE and RMSD values are given by traditional equipercentile equating method. In addition, in terms of WMSE and RMSD values while traditional linear equating and Kernel linear equating provided similar results; traditional equipercentile equating and Kernel equipercentile equating methods provided nearly same results. Apart from that, MSD index referring to systematic error indicates that Kernel equipercentile equating has the smallest error while the opposite is reported by traditional linear equating method. In consideration of all findings here, the smallest random error was reached with Kernel linear equating method and the smallest systematic error was achieved by Kernel equipercentile equating method.

4. DISCUSSIONS AND CONCLUSION

In this study, we compared the results obtained through traditional equating methods with Kernel equating with the intention of finding out the superior one.

It was found out that Kernel equating methods as a recently introduced approach produced results comparable to traditional equating methods. In Kernel equating, when parameter h was selected as optimal, equated scores provided almost identical results as traditional equipercentile equating. When it was selected large, this time the equated scores provided results almost identical to traditional linear equating. These results seem to be in accord with Mao, von Davier and Rupp (2006), von Davier et al (2006) and Grant, Zhang and Damiano (2009). In their study, Mao et al. (2006) compared traditional equating methods and Kernel equating methods in equivalent groups and nonequivalent groups in common-item test pattern, and noted quite similar results particularly as a result of linear equating in equivalent groups design. In another sample, von Davier et al. (2006) compared Tucker, chained equating, frequency estimation, Levine observed-score equating and post-stratification equating from traditional equating methods against Kernel chained equating (h optimal), Kernel chained equating (h large), and Kernel post-stratification (h large) from Kernel equating methods. They reported comparable results in both Kernel and traditional equating methods**.** Furthermore, Grant et al. (2009) comparing performances of Kernelchained equating, Kernel post-stratification, Tucker, chained equipercentile and Levine equating methods found out that Kernel chained equating yield similar results to chained linear equating method with large bandwidth. In addition, they came up with minor differences only between Kernel post-stratification equating and Tucker and Levine equating results.

In relation with error; comparison of WMSE, RMSD, and MSD values from Kernel equipercentile, Kernel linear equating, traditional linear and equipercentile equating methods recorded the smallest random error with linear equating methods applying to both traditional and Kernel equating methods. The smallest systematic error was found in Kernel equating methods. In particular, Kernel linear equating method generated the lowest random error, while the highest level of the same type of error was generated by traditional equipercentile method. It can be argued that our results show similarity with the literature. Kelecioğlu and Öztürk Gübeş (2013) carried out equating on 2009 ÖBBS social studies test A and C books according to linear and equipercentile equating with random group design and found out that linear equating involves the smallest random error. In a comparative study by Zhu (1998) on RMSD and MSD coefficients from linear equating and unsmoothed and postsmoothed equipercentile equating methods, it was found out that linear equating method involves the least random and systematic error but unsmoothed equipercentile equating show the most random and systematic errors. Yet, our findings seem to be in dispute with Zhu (1998) in relation with systematic error because Kernel equipercentile equating and traditional linear equating methods yielded the smallest and the largest random error, respectively. The result does not coincide with the findings by Zhu (1998).

When the equated scores obtained from the linear equating methods are examined, it is seen that the scores take values out of the raw score scale range. Kolen and Brennan (2014) suggested that it is an expected effect and Livingston (2014) argued that it is peculiar to linear equating methods. Kolen and Brennan (2014) proposed two alternative ways in this case. The first is to allow the points that are not in the raw score range. The second is to accept the scores outside the raw score range as the lowest and highest raw scores. In other words, it refers to taking all equated points below as 0 and taking all equated points above 15 as 15. Concerning equipercentile equating method, Kolen and Brennan (2014) claim that equated scores might deviate from the raw score range by -.5 up to +5, which is a desirable feature of equipercentile equating. That is to say, the points obtained with equipercentile equating methods need to be

valued between -0.5 and 15.5. Considering this range for both Kernel and traditional equipercentile equating, both methods seem to be in the range determined.

We also found out that the standard error obtained from Kernel equipercentile equating method is greater than Kernel linear equating and it tends to increase in marginal points. This seems to be consistent with findings from other studies in the literature. In a study by Choi (2009) comparing Kernel equating and traditional equating methods, Kernel linear equating methods were found to give lower standard error compared to Kernel equipercentile equating methods. According to Mao (2006), higher errors in end points with Kernel equating methods are due to the fact that the point scale in Gauss Kernel continuation method falls in the range of $+\infty$ and $-\infty$.

In the light of the all findings, it is understood that Kernel linear equating method has the least random error and Kernel equipercentile equating has the least systematic error in equating of booklets B and D of social studies test from 2009 ÖBBS. Yet, the error values obtained seem to be very close, which may be due to similar distribution of the test forms. Livingston (2014) contended that if the new and old forms exhibit the same distribution of points, linear and equipercentile equating methods would yield almost identical results and even equated scores could overlap.

Departing from the discussion above, it was concluded that Kernel equating methods are more suitable than others for equating booklets B and D of social studies test in the context of 2009 ÖBBS. Present study was planned to compare equatings of the foregoing documents through the use of traditional and Kernel equating methods. In the future, a similar study could be conducted in other equating designs and methods (equating methods based on Item Response Theory, local equating, etc.). Besides RMSD, MSD and WMSE, other evaluation criteria such as invariance indices and DTM could be employed in future studies as well.

ORCID

Çiğdem AKIN ARIKAN https://orcid.org/0000-0001-5255-8792
Selahattin GELBAL https://orcid.org/0000-0001-5181-7262

5. REFERENCES

Akhun, İ. (1984). İki korelasyon katsayısı arasındaki manidarlığın test edilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *17*, 1-7.

Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, *74*(8), 1-36.

Andersson, B., Branberg, K., & Wiberg, M. (2013). Performing the Kernel Method of Test Equating with the Package kequate. *Journal of Statistical Software*, *55*(6), 1–25.

Baykul, Y. (1996). *İstatistik: Metodlar ve uygulamalar* (3. Baskı). Ankara: Anı Yayıncılık

Büyüköztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı* (8.Baskı). Ankara: Pegem A Yayıncılık.

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

Dorans, J. N., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Measurement*, *37*, 281-306.

Eğitim, Araştırma ve Geliştirmesi Daire Başkanlığı (EARGED). (2010). Ortaöğretim ÖBBS raporu 2009. Ankara, Milli Eğitim Bakanlığı.

Grant, M. C., Zhang, L., & Damiano, I. (2009). An Evaluation of Kernel Equating: Parallel Equating With Classical Methods in the SAT Subject Tests™ Program. *ETS Research Report Series, 2009* (1), i-25.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications.* Boston: Academic Puslishers Group.

Holland, P. W. (2007). A framework and history for score linking. In Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.), *Linking and aligning scores and scales* (pp. 5-30). Springer, New York, NY.

Kelecioğlu, H., & Öztürk Gübeş, N. (2013). Comparing linear equating and equipercentile equating methods using random groups design. *International Online Journal of Educational Sciences*, *5*(1), 227-241.

Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice, 7*, 29-36.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd. ed.). New York: Springer

Lee, Y. H., & von Davier, A. A. (2011). Equating through alternative kernels. In von Davier, A. (Ed.) *Statistical models for test equating, scaling, and linking* (pp. 159-173).Springer New York.

Livingston, S. A. (2014). Equating test scores (without IRT), (2nd. ed.). *Educational testing service.*

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, *38*(1), 88-91.

Mao, X. (2006). *An investigation of the accuracy of the estimates of Standard errors for the kernel equating functions*. Unpublished doctoral dissertation, The University of Iowa.

Mao, X., von Davier, A. A., & Rupp, S. (2006). Comparisons of the Kernel Equating Method with the Traditional Equating Methods on Praxis™ Data. *ETS Research Report Series, 2006(2).*

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Ricker, K. L.,& Davier, A. A. (2007). The impact of anchor test length on equating results in a nonequivalent groups design. *ETS Research Report Series, 2007(2).*

von Davier, A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). An Evaluation of the Kernel Equating Method: A Special Study with Pseudo tests Constructed from Real Test Data. *ETS Research Report Series, 2006(1).*

von Davier, A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel method of equating*. New York, NY: Springer.

Zhu, W. (1998). Test equating: What, why, how?. *Research Quarterly for Exercise and Sport, 69*(1), 11-23.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 428-442

DOI: 10.21449/ijate.404205

Published at http://www.ijate.net http://dergipark.gov.tr/ijate Research Article

Examination of Factors Affecting Students' Reading-Comprehension Achievement with Structural Equation Modeling

Hakan Ülper [1,*], **Gökhan Çetinkaya** [2], **Ayhan Dikici** [3]

[1]Burdur Mehmet Akif Ersoy University, Education Faculty, Turkish Language Education, 15100, Burdur, Turkey
[2]Pamukkale University, Education Faculty, Turkish Language Education, 20700, Denizli, Turkey
[3]Nigde Omer Halisdemir University, Education Faculty, Department of Educational Sciences, 51100, Nigde, Turkey

Abstract: It is noted in the light of several studies associated with the reading-comprehension skills that fluent word reading, connectives knowledge and fluent text reading are closely related to the comprehension skill. Hence it seems possible to create a model over these concepts and their interrelationships. Within the scope of this research two different models are proposed considering theoretical information. The study group of the research was composed of 263 secondary school students at a state school in Burdur province. Three different confirmatory factor analysis (CFA) were performed to see whether factor structures were confirmed on the sample by using the maximum estimate method of the AMOS 18 software. Analysis of Covariance and Multivariate Analysis of Covariance were performed to determine whether the demographic variables grade level and gender had a significant effect on the scale scores. These analyzes were also performed in the SPSS 18 program. It has been found that as the grade level of the students increases, connective usage konowledge increases, the fluent silent word reading, the fluent silent text reading, the reading comprehension and the academic achievements increase as the grade level increases. This research concluded that reading comprehension strongly and significantly predicted academic achievement. But however, fluent silent text reading did not predict reading comprehension significantly. According to the results of the research, the model formed by the sub-dimensions of binding usage information shows better fit.

ARTICLE HISTORY

Received: 11 March 2018
Revised: 13 May 2018
Accepted: 09 June 2018

KEYWORDS

Fluent reading,
Comprehension,
Academic achievement,
Structural equation modelling

1. INTRODUCTION

According to "the simple view of reading theory", reading is composed of two main categories: word reading and language comprehension. Word reading category includes the subcategories of letter knowledge-phonics, reading accuracy and automatic reading whereas language comprehension is constituted by the subcategories of word comprehension, sentence

CONTACT: Hakan ÜLPER ✉ hakanulper@gmail.com Burdur Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Türkçe Eğitimi Anabilim Dalı, 15100, Burdur, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

comprehension and text comprehension. Hence, activation of word meaning, understanding the sentences, inference, control of comprehension and understanding the text structure are covered by this subcategory. When either of these categories does not work, reading performance is not achieved in the desired quality (*see* Oakhill, Cain, & Elbro, 2015).

In this sense, fluent processing particularly on the word level during the process of reading-comprehending a text is regarded as being of top priority for cognitive processing required for the text comprehension. Because it would be possible to focus mental sources on text comprehension with the automatization of word recognition (NRP, 2000; Perfetti, 1998), relating the word reading skill to comprehension seems probable in the theoretical framework.

It is remarkable in the literature that there is a high-level relationship between fluent word reading and comprehension in the first periods of reading particularly (Kim, Wagner, & Lopez, 2012). In the research conducted with 2143 students, Verhoeven and Leeuwe (2008) achieved comparable results. According to their research, word reading has an important effect of comprehension in early stages. In the study, it is concluded that there is a high-level relationship between fluent word reading and comprehension; the relationship is shown .50 at the lowest and .87 at the highest (*see* Castillo, Torgesen, Powell-Smith, & Al-Otaiba, 2009).

On the other hand, according to the Lexical Quality theory, lexical quality refers to how important formats and meaning elements of the word seen by a reader and the lexical knowledge are. Accordingly, a reader should be able to know both textual, phonological and morphological-syntactic attributes and semantic attributes of a word and include the words in the textual context in the text processing (Perfetti, 2007). Then, the reader should be able to read simple words fluently and process these words in accordance with the meaning required by the context and fluently. The consequent type of reading is not the act of reading from a list of simple words but reading of a text is the exact opposite.

In such reading, two main concepts are focused on in silent reading for fluency: reading accuracy and reading rate. Hence, it is possible to define this reading as fluent silent text reading.

A reader who gets into the fluent text reading process should perform inter-sentence cognitive transactions such as understanding small structures (word and phrase recognition, etc.) as well as integrating the meanings (inter-sentence grammatical and semantic relations), inferring the semantic information from text by focusing on inter-proposition relation networks, making the connections of references and filling the gaps in regard to cohesion through bridging so that he/she can comprehend a text, that is, create the mental representation of the text content (McNamara, Kintsch, Songer and Kintsch, 1996). In other words, he/she should make consistent semantic relations on the sentence level other than the word level. To that end, readers need to understand the directions for the pronouns and the referred expressions in the text and comprehend how sentences are interrelated and how clauses, paragraphs, parts and groups of part are related to the general subject (Caccamise, Snyder, & Kintsch, 2008). Again, as stated by van Silfhout, G., Evers-Vermeul, J. & Sanders, T. (2015), it is very important for reader to establish the proper relations both locally and wholly between the units of the text for structuring the mental representation of text. In this process, connectives come across as some of the most important guides in inter-sentence connection.

Linguists divide words into two groups as function words and content words. In addition to this classification, it is observed that functional and content words are also defined as open/closed class or grammatical and substantial words. Open class words are basic lexical groups such as nouns, verbs and adjectives. Closed class words are syntactic processors including grammatical groups such as determiners, prepositions and connectives (d'Arcadais, 1984). Semantic contribution of functional words is more abstract and less referential than

content words. They are the latest acquired constructs in language acquisition (Littlefield, 2005; Smith & Witten, 1993).

Hence, readers should have internalized content words like connectives for efficient reading. Because readers receive information required about the function of connectives from their long-term memory when reading. If a reader is not informed of what functions connectives have in his/her long-term memory, it will not matter when he/she reads connectives properly and fluently, that is, simply performs a decoding process. As in fluent word reading, the reader will read the word properly first, then access the information in its function/meaning in his/her mental glossary. In the next stage, this will enable the reader to use this function/meaning in accordance with the context properly and achieve fluent text reading. In this case, it is obvious that the reader needs to read connectives in the text automatically and fluently.

As stated by Kurtul (2011), connectives function to demonstrate the logical connection between propositions or between a sentence and a noun phrase in some cases. In this sense, the meaning conveyed by connectives is very important for coherent connections among discourse sections. It will be difficult to achieve desired coherence when there is an incoherent relation between connective and semantic content of connected elements (Zufferey, Mak, Degand, & Sanders, 2015: 390). Process of comprehension will be affected negatively, too.

Several research results coincide with this theoretical information. Sanders and Noordman (2000) state that presence of connectives in texts makes text processing easier, referring to numerous studies on online texts. On the other hand, Geva and Ryan (1985) state based on the related research that competent readers have good connectives knowledge and use this knowledge to establish logical relations among the parts of text. This is an outcome which is expected to increase the speed of processing during reading. Noordman, Vonk and Kempf (1992) accordingly concluded that connectives accelerate readers' reading time. Likewise, van Silfhout, Evers-Vermeul and Sanders (2015) state that many young readers cannot construct the text in their minds properly because of their incompetence in making connections when reading. According to their research, connectives not only accelerate the processing of later information but also reduce the number of rereading the current information in the text. These findings indicate that connectives have a function of an initial "processing directive". Thus, the findings also refer to that having sufficient amount of connectives knowledge during reading will contribute to a reader's fluent reading and text comprehension.

As addressed above, it is noted in the light of several studies associated with the reading-comprehension skills that fluent word reading, connectives knowledge and fluent text reading are closely related to the comprehension skill. Two models can be accordingly suggested for these relations: The first model recognizes connectives knowledge as one dimension; the second model recognizes connectives knowledge along with its subdimensions. The following is the illustration of this model (see Figure 1). In this study, it was aimed to investigate the effect of connective usage knowledge, fluent silent word reading and fluent silent text reading on reading comprehension and direct effect of reading comprehension on academic achievement. For this purpose, the following hypotheses were developed:

H1: Connective usage knowledge has a positive significant effect on fluent silent word reading, fluent silent text reading, and reading comprehension.

H2: Fluent silent word reading and fluent silent text reading have a positive significant effect on reading comprehension.

H3: Reading comprehension has a positive significant effect on academic achievement.

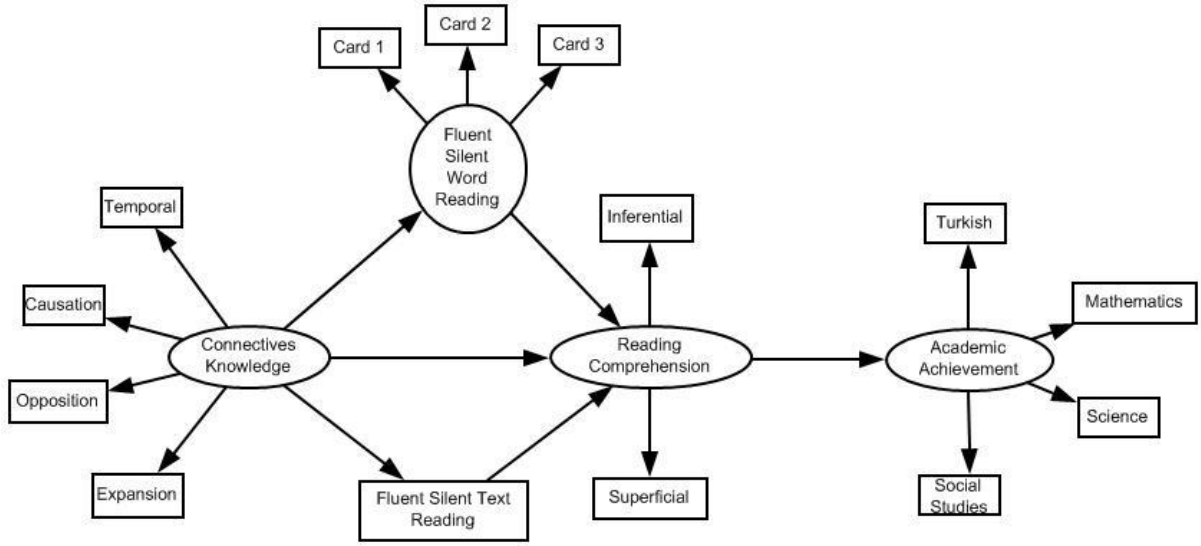


Figure 1. Proposed Model

2. METHOD

2.1. Participants

The study group of the research was composed of 263 secondary school students at a state school in Burdur province. This school was chosen because it is an easily accessible sample. 25.1% (f=66) of the students were fifth-graders, 24.7% (f=65) were sixth-graders, 24.3% (f=64) were seventh-graders, and 25.9% (f=68) were eighth-graders. 51.7% (f=136) of the students are girls and 48.3% (f=127) of them are boys. Ages of the students varied between 11 and 14. Accordingly, arithmetic means of the ages were 12.8 while the standard deviation was .26.

2.2. Instruments

Four assessment instruments and official achievement scores were used in this study:

2.2.1. Connective Knowledge Test

The items of the Connective Test developed by Çetinkaya, Ülper and Bayat (2014) were produced from sentences using side-by-side, top-and-bottom and discourse-marker connectives. Items under these three groups are based on another semantic classification. This classification is temporal, causation, opposition and expansion. There are 28 four-choice items in the instrument. Kuder-Richardson 20 (KR-20) reliability coefficient of the whole scale was found to be .74. Due to low reliability coefficients of the subscales, it was ensured that each factor was composed of three items at least (Comrey, 1988), and KR-20 reliability coefficients were .66 for the temporal factor, .71 for the causation factor, .62 for the opposition factor, and .89 for the expansion factor. Item difficulty values ranged from .47 to .73 and item discriminations ranged from .28 to .43.

2.2.2. Fluent Silent Word Reading Test

Developed by Ülper and Yağmur (2016), this test is composed of three reading cards. The first card includes 110 (+110), the second card 100 (+100), and the third card 90 (+90) actual words and fake words derived from those actual words in the same amount. In this test, students are expected to mark the actual words as many as possible in one minute. Cronbach's Alpha coefficient for the scores obtained from three cards is .96.

2.2.3. Fluent Silent Text Reading Test

Fluent Silent Text Reading Test measures the rate of silent and fluent reading of a given text. In this test, students are expected to read as many words as possible in one minute. The text in the test is narrative. The words used in the text can be known to that group of age. The text is on the independent reading level according to the formulation by Çetinkaya-Uzun (2014) in terms of readability. The experts were also consulted for opinion when choosing the text. It was seen in the preliminary practices that there was no significant difference between the scores obtained by two equivalent groups. This provides an important data of reliability.

2.2.4. Reading Comprehension Test

Developed by Ülper, Çetinkaya and Bayat (2017), the test includes 28 questions in total (11 multiple-choice, 3 fill-in-the-blank, 6 true-false, and 8 open-ended questions). Students are expected to answer all questions in the test. KR-20 reliability coefficients were found .76 for the Inference factor and .61 for the superficial text factor.

2.2.5. Academic Achievement

Official achievement scores obtained by the students in the courses of Turkish, Mathematics, Science and Social Studies were accepted as the academic achievement.

2.3. Procedure

All instruments were applied to the students under the control of the researcher on the basis of voluntariness within one week. It was explained to the students how to complete the tests before the application of all tests The Connectives Knowledge Test was applied first. All students completed the test in one class hour. On the next day, the Silent Word Reading Test was applied to the students. Card 1, card 2, and card 3 were handed to the students in this test, respectively. The students were admitted to the test in two separate groups and sat down to their places individually so that they could not be affected by each other. Following the necessary explanations, the practitioner started the one-minute duration of the test, and the students were asked to put down their pencils when the time was up. The Fluent Silent Text Reading Test was applied to the students on the next day. The students were admitted to the test one by one. It was explained to the students that they could go back and read any word again if they were to read it incorrectly. The text was handed to the students during the test, and they were asked to read it silently for one minute. The students were also asked to follow up what they read with their pencils, and if they were to go back, they would do it with their pencils so that their reading could be observed. In this process, the students were carefully watched and asked to draw a circle around the last read word once the time was up. On the last day, the Reading Comprehension Test was handed to the students, and they were asked to answer all questions. All students completed the test within one class hour. All the tests were applied in March 2017.

2.4. Data Analysis

Three different confirmatory factor analyses (CFA) were performed to see whether factor structures were confirmed on the sample by using the maximum estimate method of the AMOS 18 software. The first CFA examined whether the first-order factor structure (temporal, causation, opposition and expansion) was predicted by the second-order factor structure (Connectives Knowledge). The second CFA was carried out to find out whether the first-order factor structure (Card 1, Card, 2, and Card 3) was predicted by the second-order factor structure (Fluent Silent Word Reading). In the third CFA, it was examined whether the first-order factor structure (Superficial and Inference) was predicted by the second-order factor structure (Reading Comprehension).

Analysis of Covariance (ANCOVA) and Multivariate Analysis of Covariance (MANCOVA) and analyses were performed to determine whether the demographic variables grade level and gender had a significant effect on the scale scores. These analyses were conducted in SPSS 18 software. Levene's Test of Equality of Error Variances of ANCOVA analysis was not found siginicantly ($F_{(1-261)}$=.38, p>.05). Levene's Test values for MANCOVA analyses were ranged from .87 to 3.52 (p>.05). Box's Test of Equality of Covariance Matrices of MANCOVA was not significant ($F_{(10)}$=1.71, p>.05). Thus, ANCOVA and MANCOVA assumptions were meet. Two models were tested in the Structural Equation Modeling (SEM), which is a strong method. In the first model, connectives knowledge was recognized as the independent variable while fluent silent word reading and fluent silent text reading were accepted to be the mediatory variables, and the reading comprehension was recognized as the dependent variable while academic achievement was accepted to be the output variable. Temporal, causation, opposition and expansion are the indicators of connectives knowledge were recognized as independent variables in the second model, and other measures were organized in the same way as in the first model. While connectives knowledge was included in the first model as the latent variable, subdimensions of connectives knowledge were examined as observed variables in the second model. Fluent Silent Text Reading was examined as the observed variable in both proposed models. Acceptability limits of CMIN/x^2, RMR, RMSEA, NFI, CFI, GFI, and AGFI were considered in the evaluation of model fit (Byrne, 2010; Hair, Black, Babin, & Anderson, 2010; Hu & Bentler, 1995; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Schumacker & Lomax, 2004).

All scores were translated into standard scores for the analyses in the first place. No standard score lower than -3.0 and higher than +3.0 was observed (Bakeman & Robinson, 2014). It was additionally observed that linearity, singularity and multicollinearity assumptions of SEM analysis were met. Data of five students were excluded from the dataset due to high Mahalanobis d^2. Then, Mahalanobis d^2 varied between 14.61 and 39.09 (p>.001). These findings and data were included in the proposed model. Sobel's z test (Sobel, 1982) was conducted to test the possible effect in case fluent silent word reading and fluent silent text reading had the mediatory effect between connectives knowledge and reading comprehension (Baron & Kenny, 1986; Kim & Bentler, 2006; Preacher & Hayes, 2004).

3. FINDINGS

3.1. Preliminary Analyses

3.1.1. Factor Structure of Connectives Knowledge Test

The CFA results showed good fit ($x^2_{(2)}$ =2.15, p>.05; x^2/df=1.07; RMR=.03; RMSEA=.01; NFI=.99; CFI=.99; GFI=.99; AGFI=.98). The standardized parameter estimates differ between .38 and .71 and are significant at .001. In other words, connectives knowledge predicted temporal factor at .71, causation factor at .68, opposition factor at .38 and expansion factor at .66.

3.1.2. Factor Structure of Fluent Silent Word Reading Test

The CFA results indicated good fit of the Fluent Silent Word Reading Test ($x^2_{(2)}$ =01, p>.05; x^2/df=0.01; RMR=.00; RMSEA=.01; NFI=.1.00; CFI=.1.00; GFI=.1.00; AGFI=.99). The standard parameter estimates were found significant at .001. Fluent silent word reading predicted Card 1 at .93, Card 2 at .98 and Card 3 at .92.

3.1.3. Factor Structure of Reading Comprehension

The CFA results showed good fit of the factor structure of the Reading Comprehension Test ($x^2_{(2)}$ =01, p>.05; x^2/df=0.01; RMR=.00; RMSEA=.01; NFI=.1.00; CFI=.1.00; GFI=.1.00;

AGFI=.99). Reading comprehension predicted inferential at .93 and Superficial at .98. The standard parameter estimates were found significant at .001.

3.1.4. Effects of Demographics on Measures

Gender and grade level were found to be significant on dependent variables in analyses conducted without controlling gender or grade level. Wilks Lambda= 2.05, $p< .05$, η_p^2=.06 for the gender effect and Wilks Lambda= 26.14, $p< .001$, η_p^2=.45 for the grade level effect. Grade level was found to have a statistically significant effect on subscales of connectives knowledge in the results of MANCOVA performed with students' genders being kept constant (Wilks Lambda= 16.06, $p< .001$, η_p^2=.20). Grade level's partial eta-squared levels were found (η_p^2=.25) on the temporal factor, (η_p^2=.32) on the causation factor, (η_p^2=.22) on the opposition factor, and (η_p^2=.27) on the expansion factor. Bonferroni test showed that fifth-grade level was different from all other levels in all factors. The results indicated that connectives knowledge increased as the grade level increased. Gender was found to have no statistically significant effect on subscales of connectives knowledge in the results of MANCOVA performed with students' grade levels being kept constant (Wilks Lambda= .88, $p> .05$, η_p^2=.01).

Grade level was found to have a statistically significant effect on subscales of silent word reading test in the results of MANCOVA performed with students' genders being kept constant (Wilks Lambda= 59.21, $p< .001$, η_p^2= .40). Grade level's partial eta-squared levels were found (η_p^2=.71) on Card 1, (η_p^2=.73) on Card 2 and (η_p^2=.72) on Card 3. Bonferroni test showed that fifth-grade level was different from all other levels in all factors. The results indicated that achievement of fluent silent word reading increased as the grade level increased. Gender was found to have a statistically significant effect on subscales of connectives knowledge in the results of MANCOVA performed with students' grade levels being kept constant (Wilks Lambda= 2.22, $p< .05$, η_p^2=.02). The results were in favor of the girls.

Grade level was found to have a statistically significant effect on fluent silent text reading test in the results of ANCOVA performed with students' genders being kept constant (F= 175.25, $p< .001$, η_p^2=.67). Bonferroni test showed that fifth-grade level was different from all other levels in all factors. According to the results, achievement of fluent silent word reading increased as the grade level increased. Gender was found to have a statistically significant effect on fluent silent text reading in the results of ANCOVA performed with students' grade levels being kept constant (F=.7.96, $p< .01$, η_p^2=.03). The results were in favor of the girls.

Grade level was found to have a statistically significant effect on subscales of reading comprehension test in the results of MANCOVA performed with students' genders being kept constant (Wilks Lambda= 22.69, $p< .001$, η_p^2=.21). Grade level's partial eta-squared levels were found (η_p^2=.28) on the inferential factor, (η_p^2=.25) on the superficial factor, and (η_p^2=.72) on Card 3. Bonferroni test showed that fifth-grade level was different from all other levels in all factors. The results indicated that reading comprehension increased as the grade level increased. Gender was found to have no statistically significant effect on subscales of reading comprehension test in the results of MANCOVA performed with students' grade levels being kept constant (Wilks Lambda= .11, $p> .05$, η_p^2=.001).

Grade level was found to have a statistically significant effect on subscales of academic achievement in the results of MANCOVA performed with students' genders being kept constant (Wilks Lambda= 9.42, $p< .001$, η_p^2=.13). Grade level's partial eta-squared levels were found (η_p^2=.28) on the Turkish course, (η_p^2=.30) on the Mathematics course, (η_p^2=.31) on the Science course, and (η_p^2=.26) on the Social Studies course. Bonferroni test showed that fifth-grade level was different from all other levels in all factors. According to the results, students' academic achievement increased as the grade level increased. Gender was found to have no statistically significant effect on subscales of reading comprehension test in the results of

MANCOVA performed with students' grade levels being kept constant (Wilks Lambda= .48, p> .05, η_p^2=.008). Arithmetic means and standard deviations obtained in the measures are shown in Table 1.

Table 1. Arithmetic Means and Standard Deviations of Measures by Gender and Grade Levels

			Connectives Knowledge		Silent Word Reading		Silent Text Reading		Reading Comprehension		Academic Achievement	
		n	M	SD	M	SD	M	SD	M	SD	M	SD
Grade Level	5	64	5.42	3.33	54.35	5.06	125.78	6.87	11.20	3.91	293.04	55.95
	6	64	9.32	1.79	65.95	4.44	137.85	6.48	14.35	2.34	342.03	31.90
	7	63	9.76	.92	69.79	1.52	141.85	4.07	15.65	1.88	345.79	25.70
	8	67	10.20	1.14	74.32	4.37	153.31	9.53	16.61	1.93	356.26	25.54
Gender	Female	134	8.79	2.80	66.62	8.05	140.73	12.00	14.47	3.06	336.26	40.14
	Male	124	8.58	2.75	65.71	8.92	138.90	12.29	14.47	3.61	332.58	48.01

3.2. Measurement Model

The measurement model was established on four latent variables and one observed variable. Temporal, causation, opposition and expansion which are four indicators of connectives knowledge; Card 1, 2, and 3 which are three indicators of silent word reading; and silent text reading were included in the measurement model as the observed variables whereas inferential and superficial which are two indicators of reading comprehension; and Turkish, Mathematics, Science and Social Studies courses' achievement grades which are four indicators of academic achievement were included in the measurement model. The goodness-of-fit indexes were found to be within the good fit limits ($x^2_{(61)}$ =87.24, p>.05; x^2/df=1.43; RMR=.06; RMSEA=.04; NFI=.99; CFI=.99; GFI=.99; AGFI=.92). Factor loads varied between .53 and .97. It was also observed that all standardized parameter estimates were significant at .001.

3.3. Structural Model

Regarding the goodness-of-fit indexes of the two proposed model, the fit indexes of Model 2 ($x^2_{(54)}$=57.64, p<.05; x^2/df=1.06; RMR=1.04; RMSEA=.01; NFI=.99; CFI=.99; GFI=.97; AGFI=.94) were found better than the fit indexes of Model 1 ($x^2_{(65)}$=161.15, p<.05; x^2/df=2.47; RMR=2.19; RMSEA=.07; NFI=.96; CFI=.97; GFI=.92; AGFI=.87). Causations between connectives knowledge and reading comprehension and between silent text reading and reading comprehension were excluded from Model 1. After the proposed modifications had been done, the fit indexes achieved the good fit. According to Figure 2 and Table 2, connectives knowledge predicted fluent silent word reading (β=.57, t=6.12, p<.001) and fluent silent text reading (β=.77, t=9.29, p<.001); fluent silent word reading predicted reading comprehension (β=.83, t=13.21, p<.001), fluent silent text reading predicted fluent silent word reading (β=.48, t=6.80, p<.001), and reading comprehension predicted academic achievement (β=.84, t=12.68, p<.001) significantly.

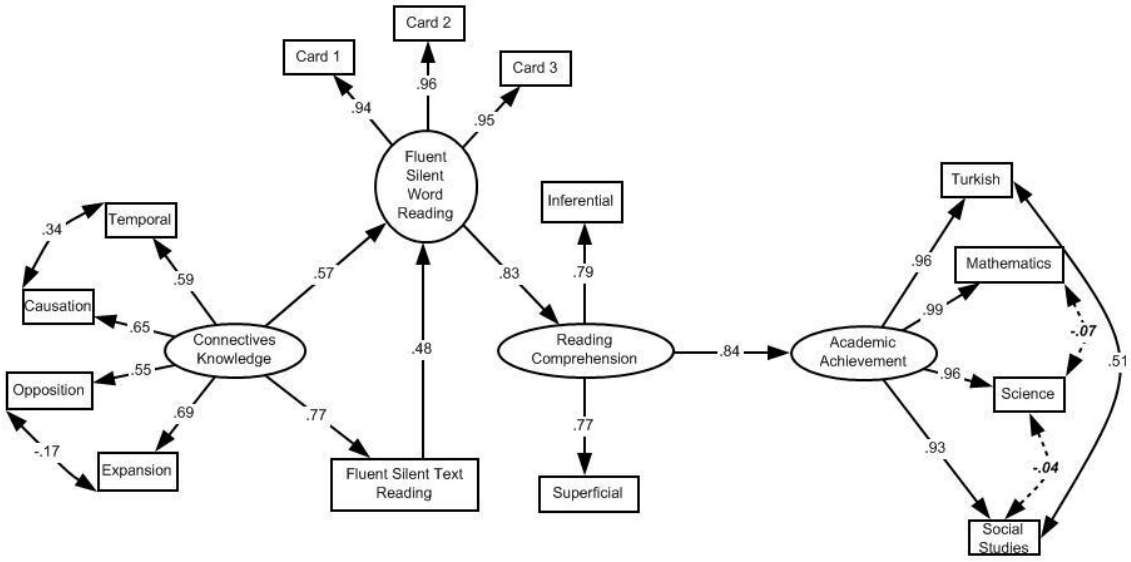


Figure 2. Results of Proposed Model 1

Note: Dashed lines and bold italic parameters are not statistically significant ($p>.05$).

Table 2. Results of Model 1's SEM Analysis

Dependent Variables		Independent Variables	Estimate	Error	*t*	*β*	*p*
SITEREAD	<---	Conn._Know.	19.99	2.15	9.29	.77	***
SIWOREAD	<---	Conn._Know.	3.59	.58	6.12	.57	***
SIWOREAD	<---	SITEREAD	.11	.01	6.80	.48	***
Read._Comp.	<---	SIWOREAD	.45	.03	13.21	.83	***
Acad._Achieve.	<---	Read._Comp.	5.50	.43	12.68	.84	***

***$p< .001$

The results obtained in the analyses of Model 2 are shown on Figure 3 and presented in Table 3. Following the proposed modifications, the fit indexes of Model 2 was found to be better than the fit indexes of Model 1. The proposed modifications are shown on Figure 3. Two of the standardized parameter estimates were not found statistically significant. One of these estimates, the temporal factor of connectives knowledge did not significantly predict fluent silent word reading (β=.05, t=1.61, $p>.05$), and the other one, which is fluent silent text reading did not significantly predict reading comprehension (β=-.10, t=-.89, $p>.05$). All other standardized parameter estimates were found statistically significant. Differently from Model 1, it was seen in Model 2 that all factors of connectives knowledge significantly predicted reading comprehension. Moreover, fluent silent text reading predicted fluent silent word reading (β=.72, t=22.38, $p<.001$) more strongly. Sobel's z test was performed to test whether fluent silent word reading had a mediatory effect between connectives knowledge factors and reading comprehension. Fluent silent word reading was found to have a mediatory effect between causation and reading comprehension (z=2.41, $p<.01$), opposition and reading comprehension (z=2.31, $p<.05$), and expansion and reading comprehension (z=2.73, $p<.01$).

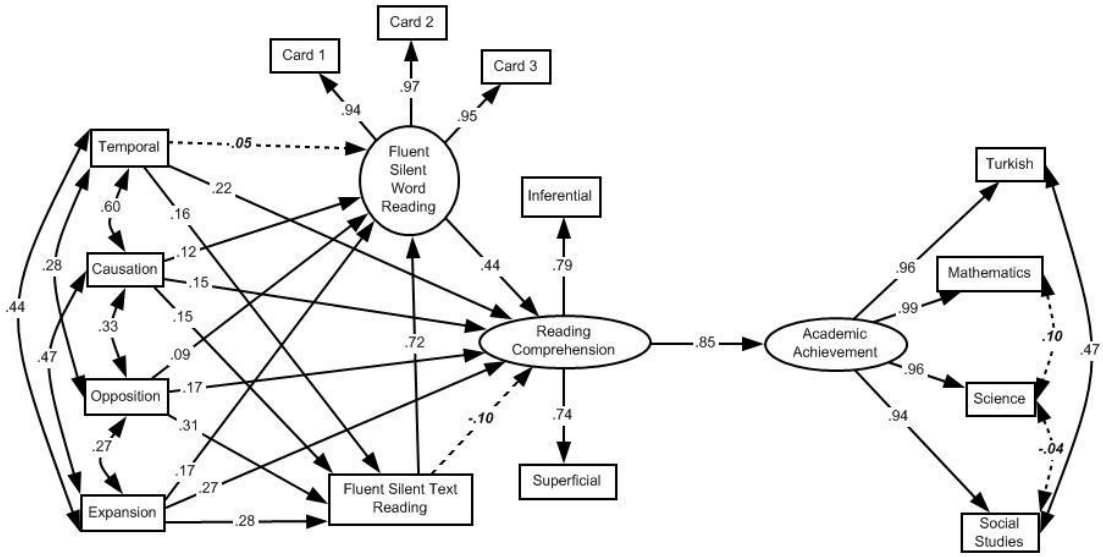


Figure 3. Results of Proposed Model 2

Note: Dashed lines and bold italic parameters are not statistically significant (p>.05).

Table 3. Results of Model 2's SEM Analysis

Dependent Variables		Independent Variables	Estimate	Error	*t*	*β*	*p*	*Sobel's z*
SITEREAD	<---	Expansion	5.07	.99	5.08	.28	***	
SITEREAD	<---	Opposition	3.51	.58	6.04	.30	***	
SITEREAD	<---	Causation	1.72	.73	2.34	.14	.019	
SITEREAD	<---	Temporal	2.03	.74	2.73	.16	.006	
SIWOREAD	<---	Temporal	.13	.08	1.61	.05	.107	
SIWOREAD	<---	Causation	.33	.08	3.94	.11	***	2.41**
SIWOREAD	<---	Opposition	.24	.07	3.54	.09	***	2.31*
SIWOREAD	<---	Expansion	.73	.11	6.14	.16	***	2.73**
SIWOREAD	<---	SITEREAD	.17	.00	22.38	.72	***	
Read._Comp.	<---	SIWOREAD	.23	.07	3.06	.44	.002	
Read._Comp.	<---	Temporal	.35	.07	4.59	.22	***	
Read._Comp.	<---	Causation	.23	.07	2.98	.15	.003	
Read._Comp.	<---	Opposition	.24	.06	3.78	.17	***	
Read._Comp.	<---	Expansion	.64	.12	5.29	.27	***	
Read._Comp.	<---	SITEREAD	-.01	.01	-.89	-.10	.370	
Acad._Achieve.	<---	Read._Comp.	5.62	.40	13.82	.85	***	

***p<.001, **p<.01, *p<.05

4. DISCUSSION AND CONCLUSION

It was concluded that students' connectives knowledge increased as their grade levels increased. Higher grade levels also meant higher levels of fluent silent word reading, fluent silent text reading, reading comprehension and academic achievement. It was found that fluent silent text reading was in favor of the girls and their scores were higher.

The model established with the factors of connectives knowledge showed better fit. Only the temporal factor of connectives knowledge did not predict fluent silent word reading significantly. Factors of causation, opposition and expansion significantly predicted fluent silent word reading, fluent silent text reading and reading comprehension. These results highly coincide with the findings of few researches in the Turkish literature. In the study conducted

by Gençer (2013) to examine the relationship between connectives in the text and reading comprehension, reading comprehension scores of the poor readers were found negative significantly correlated to the connectives that report "opposition" and "expansion". In other words, higher number of connectives in the text meant lower reading comprehension of the poor readers. The main reason for this is that poor readers cannot process connectives in the text.

It was found that fluent silent word reading was a partial mediator in the relationship between connectives knowledge factors of causation, opposition and expansion and reading comprehension. In the Turkish literature (*see* Yılmaz, 2008), it was observed that the activity of having students read the words in a simple way which they misread in the text by writing down them on cards with the word repetition technique had a positive impact both on reading fluency and elimination of reading errors and comprehension. The study performed by Çetinkaya, Ülper and Yağmur (2015) concluded that fluent silent word reading was correlated to comprehension. It is also known that studies in foreign languages have similar results (*see* Yılmaz, 2008). How the students who began to read the words automatically through the word repetition technique improved their reading skills refers to a similar situation that fluent word reading skill is closely related to comprehension. The presence of such relationship is remarkable in the foreign literature (Kim, Wagner, & Lopez, 2012; Verhoeven & Leeuwe, 2008). In a study, which concluded that there is a high-level relationship between fluent word reading and comprehension, the relationship is shown .50 at the lowest and .87 at the highest (*see* Castillo et al., 2009). Coincidence of these findings with the findings of our research is important in that it refers to similar situations in foreign languages. Another study that can be associated with the prediction of comprehension by word reading skill was the study of Yıldırım, Yıldız and Ateş (2011). In their study, participants' lexical knowledge was found moderately correlated to narrative text comprehension and highly correlated to informative text comprehension. Since high levels of lexical knowledge are important for student to recognize words and read more fluently, it is obvious that this relationship indirectly refers to the fluent word reading-comprehension relationship. Hence, both oral and silent reading fluency is regarded as being associated with comprehension in the literature.

Considering the studies on fluent oral word reading both in Turkish and foreign literature, fluent word reading positively affects comprehension, individuals who can read a text fluently or become a fluent reader through the education have higher comprehension levels or there is an interaction in between (*see* Baştuğ & Keskin; 2013; Başaran, 2013; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Pikulski & Chard, 2005;). According to the finding of this research, fluent silent text reading did not predict reading comprehension significantly. There can be two explanations to the case: The first is about the subject being discussed in the literature. Accordingly, do reading fluency predicts comprehension, or can individuals read fluently because they can comprehend well? Some researchers argue that this is a two-way relationship (*see* Dowhower, 1987; Yıldız, 2013). Yıldız (2013) explored a two-way relationship for the Turkish language. The second is that the studies in the literature generally conducted through oral reading. The general opinion is that oral reading predicts comprehension. However, fluent reading is not performed only orally but also in a silent manner. It is, on the other hand, apparent that prosodic qualities cannot be observed in silent reading fluency. In the study conducted by Jenkins and Jewell (1993) on the relationship between fluent silent reading and comprehension, fourth-grade students' fluent silent reading levels were found weakly correlated to their comprehension levels (r=.38), and second-grade students' fluent silent reading levels were strongly correlated to their comprehension levels (r=.76). This finding does not coincide with the finding of our research. Yet, it is remarkable that there is no study on this subject in the Turkish literature. As many studies on the fluent silent text reading-comprehension relationship as in oral reading are required to achieve a literal knowledge on the subject. Only then it will

be possible to provide more concrete explanations of the relationship between fluent silent text reading and comprehension.

This research concluded that reading comprehension strongly and significantly predicted academic achievement. Similarly, Yıldız (2013) confirmed the hypothesis that reading comprehension has a significant effect on academic achievement. Likewise, Ural and Ülper (2013) drew attention to the relationship between reading comprehension skill and mathematical modeling. Akay (2004) and Özdemir and Sertsöz (2006) remarkably explored a similar relationship in terms of mathematical problem solving. Bayat, Şekercioğlu and Bakır (2014) found that reading comprehension achievement significantly predicted the science course achievement.

Another related study compared students' score averages of reading comprehension with their achievements in the courses of Turkish, Mathematics, Social Studies, and Science and Technology. The study in question observed that students' reading comprehension achievements were statistically correlated to their achievements in the courses of Turkish, Mathematics, Social Studies, and Science and Technology (Yılmaz, 2011). Again, it was explored in another research that grade averages of the students who like reading books were higher than the students who do not (Yılmaz, 2012). The common reference point of these studies is that academic achievement requires sufficient level of comprehension. Coincidence between all these findings and the findings achieved in our research indicates the dominant role of comprehension in academic achievement.

It is possible that this research could be taken up in a wider frame and also be a source for other studies. In this context, similar studies can be done with different sample groups, and another model proposal can be made based on some other variables that explain the reading. The limitation of the present study is that the data were obtained from secondary school's students at a state school in Burdur province. Future studies should be carried out based on a larger number of students obtained from different schools and different provinces and in order to provide more comprehensive results.

ORCID

Hakan Ülper https://orcid.org/0000-0002-7067-599X

5. REFERENCES

Akay, A. A. (2004). *İlköğretim 2.sınıf öğrencilerinin okuduğunu anlama becerilerinin matematik problemlerini çözme başarısına etkisi [Effect of reading comprehension skills on mathematical problem solving among elementary second-grade students]*. İstanbul: Marmara University, Institute of Educational Sciences, Postgraduate Thesis.

Bakeman, R. & Robinson, B. F. (2014). *Understanding statistics in the behavioral science.* Mahwah, New Jersey: Psychology Press.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Başaran, M. (2013). Okuduğunu anlamanın bir göstergesi olarak akıcı okuma [Fluency reading as an indicator of reading comprehension]. *KUYEB, 13(4)*, 2277-2290.

Baştuğ, M. & Keskin, H. K. (2012). Akıcı okuma becerileri ile anlama düzeyleri (basit ve çıkarımsal) arasındaki ilişki [Relationship between fluency reading skills and comprehension levels (literal and inferential)]. *Ahi Evran University, Journal of Kırşehir Faculty of Education (KEFAD), 13*(3), 227-244.

Bayat, N., Şekercioğlu, G., & Bakır, S. (2014). The Relationship Between Reading Comprehension and Success in Science. *Education and Science, 39 (176)*, 457-466.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge, Taylor & Francis Group.

Caccamise, D., Snyder, L., & Kintsch, E. (2008). Constructivist theory and the situation model: Relevance to future assessment of reading comprehension. In C. C. Block, & S. R. Parris (Eds.), *Comprehension instruction: Research-based best practices* (2nd ed., pp. 80-97). NY: The Guilford Press. Chapple

Castillo, J.M., Torgesen, J.K., Powell-Smith, K.A., & Al-Otaiba, S. (2009). Examining the decision reliability and validity of three reading fluency measures for predicting outcomes on statewide reading accountability tests. In R.K. Wagner, S. Schatsneider, & C. PhythianSence (Eds.), Beyond decoding: The behavioral and biological foundations of reading comprehension (pp. 195-223). New York, NY: Guilford Press.

Comrey, A. L. (1988). Methodological contributions to clinical research. Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*(5), 754-761.

Çetinkaya, G. & Uzun, L. (2011). Türkçe Ders Kitaplarındaki Metinlerin Okunabilirlik Özellikleri [Readability Qualities of Turkish Course Book Texts], Ed. Hakan Ülper, *Türkçe Ders Kitabı Çözümlemeleri [Analyses of Turkish Course Book]*, Ankara: PegemA yay.

Çetinkaya, G., Ülper, H., & Bayat, N. (2014). Bağlayıcı testinin geliştirilmesi ve öğrencilerin bağlayıcı bilgisinin çeşitli değişkenlere göre incelenmesi [*Developing Connectives Test and examining students' connectives knowledge based on various variables*] *International Journal of Language Academy, 2(3)*, 88-98.

Çetinkaya, G., Ülper, H., & Yağmur, K. (2015). Öğrencilerin Doğru ve Akıcı Sessiz Sözcük Okuma Becerileriyle Kavrama Başarıları Arasındaki İlişki [Relationship between Students' Reading Accuracy and Fluency and Comprehension Achievements], *İlköğretim Online*, *14*(3), 993-1004.

Dowhower, S. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22*, 389-406.

Flores d'Arcadais, G. D. (1984). Lexical knowledge and word recognition: Children's reading of function words. *Visible Language, 18*(4), 359-377.

Fuchs, L.S., Fuchs, D., Hosp, M.D., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading, 5, 239– 259. Geva, E.; Ryan, E., B. (1985). Use of conjunctions in expository text by skilled and less skilled readers. *Journal of Reading Behavior, 17*(4), 331-345.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hu, L. T. & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp.76-99). Thousand Oaks, CA: Sage Publications.

Jenkins, J. R. & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*, 421-432.

Kim, K. H. & Bentler, P. M. (2006). Data modeling: Structural equation modeling. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 161-175). Mahwah, NJ: Lawrence Erlbaum.

Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology, 113*, 93-111.

Kurtul, K. (2011). *Türkçe ve İngilizce'deki Bağlaçların Yazılı Metinlerde Kullanımı [Use of Turkish and English Connectives in Written Texts]*. (Unpublished doctoral thesis) Ankara University, Institute of Social Sciences, Ankara.

Littlefield, H. (2005). Lexical and functional prepositions in acquisition: Evidence for a Hybrid Category. Boston University Conference on Language Development 29, Online Proceedings Supplement.

McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assesment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: The National Instute of Child Health and Human Development.

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language, 31*(5), 573-590.

Oakhill, J., Cain, K., & Elbro, C. (2015). *Understanding and teaching reading comprehension: A handbook.*, New York: Routledge.

Özdemir, A.Ş. & Sertsöz, T. (2006). Okuduğunu anlama davranışının kazandırılmasının matematik başarısına etkisi [How bringing reading comprehension behavior to students affect mathematics achievement]. M.U. *Atatürk Faculty of Education Journal of Educational Sciences, 23*, 237-257.

Perfetti, C. A. (1998). Two basic questions about reading and learning to read. In P. Reitsma & L. Verhoven (Eds.) *Problems and interventions in literacy development* (pp. 15-48). Dordrecht, The Netherlands: Kluwer-Academic Publisher.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.

Pikulski, J. J. & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *Reading Teacher, 58*, 510-519.

Preacher, K. J. & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36(4),* 717-731. http://dx.doi.org/10.3758/BF03206553

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Schumacker, R. E. & Lomax, R. G. (2004). *Beginner's guide to structural equation modeling* (2nd ed.). New York: Psychology Press, Taylor & Francis Group.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. *Sociological Methodology, 13,* 290-312.

Smith, T. C. & Witten, I. H. (1993). Language inference from function words (Computer Science Working Papers 93/3). Hamilton, New Zealand: Department of Computer Science, University of Waikato.

Sanders, T. J. M. & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29*(1), 37-60.

Ural, A. & Ülper, H. (2013). İlköğretim Matematik Öğretmeni Adaylarının Matematiksel Modelleme ile Okuduğunu Anlama Becerileri Arasındaki İlişkinin Değerlendirilmesi [Evaluation of the Relationship between Preservice Elementary Mathematics Teachers'

Mathematical Modeling and Reading Comprehension Skills]. *Kuramsal Eğitimbilim Dergisi - Journal of Theoretical Educational Science*, *6*(2), 214-241.

Ülper, H. & Yağmur, K. (2016) Akıcı sessiz sözcük okuma testinin geliştirilmesi [Developing the fluent silent word reading test]. *İlköğretim Online, 15(2)*, 581-593.

Ülper, H., Çetinkaya, G., & Bayat, Nihat (2017). Okuduğunu anlama testinin geliştirilmesi [Developing the reading comprehension test]. *Ahi Evran University, Journal of Kırşehir Faculty of Education (KEFAD), 18*(1), *175-187.*

van Silfhout, G., Evers-Vermeul, J., & Sanders, T. (2015) Connectives as processing signals: how students benefit in processing narrative and expository texts. *Discourse Processes, 52*(1), 47-76

Verhoeven, L. & Leeuwe, J. V. (2008). Prediction of the development of reading comprehension: a longitudinal study. *Applied Cognitive Psychology, 22*, 407–423.

Yıldız, M. (2013). Okuma motivasyonu, akıcı okuma ve okuduğunu anlamanın beşinci sınıf öğrencilerinin akademik başarılarındaki rolü [Role of reading motivation, reading fluency and reading comprehension on fifth-grade students academic achievements], *Turkish Studies - International Periodical for The Languages, Literature and History of Turkish or Turkic, 8/4, 1461-1478.*

Yıldırım, K., Yıldız, M., & Ateş, S. (2011). Kelime Bilgisi Okuduğunu Anlamanın Anlamlı Bir Yordayıcısı mıdır ve Yordama Gücü Metin Türlerine Göre Farklılaşmakta mıdır? [Is Lexical Knowledge a Predictor of Reading Comprehension and Do its Predictiveness Differ by Types of Text?] *Kuram ve Uygulamada Eğitim Bilimleri, 11*(3), 1531-1547.

Yılmaz, M. (2008). Kelime tekrar tekniğinin akıcı okuma becerilerini geliştirmeye etkisi [Effect of word repetition technique on improving the reading fluency skills]. *Turkish Journal of Educational Sciences*, *6*(2), 323-350.

Yılmaz, M. (2011). İlköğretim 4.sınıf öğrencilerinin okuduğunu anlama seviyeleri ile Türkçe, matematik, sosyal bilgiler ve fen ve teknoloji derslerindeki başarıları arasındaki ilişkinin belirlenmesi [Relationship between elementary fourth-grade students' reading comprehension levels and achievements in the courses of Turkish, mathematics, social studies, and science and technology]. *Dumlupınar University Journal of Institute of Social Sciences, 29*, 9-13.

Yılmaz, B. (2012). Okuma Alışkanlığının Okul Başarısına Etkisi: Ankara Keçiören Atapark İlköğretim Okulu Öğrencileri Üzerine Bir Araştırma [How Reading Habit Affects School Achievement: Research on Students of Ankara Keçiören Atapark Elementary School]. Külcü, Ö., Çakmak, T. & Özel, N. (Ed.), *Prof. K. Gülbün Baydur'a armağan [A Gift to Prof. K. Gülbün Baydur]* (210-218). Ankara: Özyurt Matbaacılık.

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Visible Language, 18*(31), 389-411.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 443–460

DOI: 10.21449/ijate.442399

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Examining the 4th Grade Students' Ability to Say "No" in the Framework of Life Studies Curriculum with Many-Facet Rasch Model

Ferat Yılmaz[1,*], **M. Akif Sözer**[2]

[1]Primary Education Department, Dicle University, Diyarbakır, Turkey
[2]Primary Education Department, Gazi University, Ankara, Turkey

Abstract: This study aims to reveal whether or not the ability to say "no", which is tried to be taught to students in the 2005 life studies curriculum, has been gained by children with regards to various situations. The survey model was utilized in the study. The study was conducted with 4th grade students who took the class of life studies in primary school 1st, 2nd and 3rd grades and therefore are expected to have gained the ability to say "no". In order to collect data, The Ability to Say "No"-Specific Cases Inventory (NSCI) developed by the researchers was employed. In order to be able to evaluate the responses given to the cases in the NSCI, the "NSCI Rubric" was developed. The data obtained from NSCI were analyzed according to the many-facet Rasch analysis using the FACETS packet program. The findings suggest that students may be able to demonstrate the basic ability to say "no" expected of them in response to specific cases; however they will not focus too much on the behavior of making an explanation about the reason why they have said "no", or, on performing alternative behaviors after saying "no".

ARTICLE HISTORY

Received: 14 January 2017
Revised: 11 April 2018
Accepted: 25 May 2018

KEYWORDS

Life studies curriculum,
The ability to say "no",
Many-facet Rasch analysis

1. INTRODUCTION

Regardless of circumstances, people behave reactively in the face of threats against themselves and their values (Aytaç, 2005). One of these reactions is the demonstration of the ability to say "no", the most effective way of expressing the boundaries of personal space clearly and seriously (Bolton, 1979). The significance of the ability to say "no" underlies the circumstances when having trouble saying "no" can be improper and unsafe while examining children as well as adults. Some cases in which it is crucial to say "no" are when a child receives substance use offers, is faced with the risk of being sexually abused, needs to manage time, feels moral problems, and/or is confronted with inappropriate requests from online environments. These cases overall include the circumstances when the child should exhibit the ability to say "no."

Substance refers to any chemical that can cause addiction when used abusively, resulting in deterioration of the mood, mental processes and various functions of the brain (Ceyhun, Oğuztürk and Ceyhun, 2001). When the studies carried out on substance use (Boztaş and Arısoy, 2010; Özyurt and Dinç, 2006; Ünsal and Sezgin, 2009; Korkmaz Ekren, Başarık, and Özhan,

CONTACT: Ferat Yılmaz ferat.yilmaz@dicle.edu.tr Faculty of Education, Department of Primary Education, Diyarbakır, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

2011; Ögel and Aksoy, 2007) are reviewed, it appears that the age of finishing 4-year primary school process and the first year after finishing elementary school (10-12 years age) is the time posing the most risk in terms of starting substance use for the first time. Furthermore, it is easily inferred how important it is in this period to gain preventive strategies in terms of substance use, considering that primary school period is when the positive perceptions of the students about the individuals using substances, like cigarettes, are lower than the middle-school students (Yazıcı & Şahin, 2006). Therefore, while the children are still enrolled in elementary school, they need to be equipped with some abilities that will prevent them from starting substance use. That being said, one of the abilities children need to gain in order to protect themselves about substance use is the ability to say "no" to both their peers and adults (Tokur Kesgin, 2012). Gaining this ability, especially in primary school or in the early years of adolescence can make it easier entirely to reject offers to use substance in the coming years (Belgrave, Reed, Plybon, & Corneille, 2004).

Sexual abuse is defined as violent sexual acts committed by an adult, a teenager, or a peer to a child (Rind, Tromovitch & Bauserman, 1998). One of the studies conducted within the context of sexual abuse children suffer at the international level indicates that 7.9% of males and 19.7% of females were exposed to sexual abuse before the age of 18 (Pereda, Guilera, Forns, & Gómez-Benito, 2009). A meta-analysis study carried out at the international level also demonstrates that at least 164 out of every 1000 girls and at least 66 out of every 1000 boys suffered from sexual abuse (Stoltenborgh, Ijzendoorn, Euser & Bakermans-Kranenburg, 2011). These studies reveal that the sexual abuse children are exposed to is a serious problem that should not be ignored anywhere in the world; also, it should be addressed gravely and needs to be studied in our country, as well (Aktepe, Işık, Kocaman & Eroğlu, 2013; Bilginer, Hesapçıoğlu & Kandil, 2013; Taner, Çetin, Işık & İşeri, 2015). The problem of sexual abuse in Turkey requires some strategies to be adopted in order to prevent sexual abuse and to protect children against sexual abuse, just as in other countries. One of these strategies is the demonstration of the ability to say "no" against sexual abuse (Özkan, 2011). With regards to the sexual abuse children are exposed to, it can be argued whether or not sexual abuse can be perceived by the child and whether or not a child can say "no" even if he/she perceives it. Within the scope of this discussion, Polat (2001) asserts that even younger children can easily understand intuitively what behaviors can be covered by sexual abuse. Sanderson (2010) argues that children are not completely passive victims of sexual abuse and they can protect themselves if they are informed of the threats of sexual abuse, are allowed to resist them, and are given the necessary information to resist acts of sexual abuse.

Time management refers to effective use of time when performing actions for specific purposes (Claessens, Eerde, Rutte & Roe, 2007). Behaviors related to time management should start to be gained along with the understanding that time is a fluid and perpetual structure. According to Passig (2004), the most appropriate stage in this respect is the late childhood stage when the individual begins to establish connections between events, develops a sense of continuity in time, and discovers the recurring patterns of time and the sequence of certain events. Therefore, it can be said that the process of teaching time-management related behaviors should start during the primary school years. In this respect, it can be argued first that there is a need to remove the obstacles in front of the ability to exhibit time-management related behaviors. One of the main obstacles in front of exhibiting time management behaviors is the lack of ability to say "no" (Mackenzie & Nickerson, 2009). Time management does not actually refer to managing time; it refers to managing priorities (Manktelow & Anand, 2008). Accordingly, individuals who do not believe that they have the right to set priorities as one of the basic assertiveness rights and who prioritize requests for the needs of others rather than their needs because they cannot say "no" to others cannot succeed in time management.

The ability to say "no" should also be exhibited in the cases with moral problems. Such cases may involve non-ethical proposals (Szpalski, Gunzburg and De Kleuver, 2003) as well as some social-moral problems (Leming, 1997). Therefore, to say "no" in the face of such situations is the ability to act in line with moral aspects by leaving aside an instinctive reaction (Calabrese, 1989; Gündoğan, 2009). The presence of the moral aspects in the ability to say "no" suggests that this ability is related to the moral character (Yılmaz & Ersoy, 2016). Considering the importance of character development in childhood (O'Sullivan, 2004), it can be put forward that the ability to say "no" in the face of moral problems must be learned in primary school years.

The recent development in the information and communication technologies has led to the transfer of face-to-face communication and interaction patterns in the home, at school, at the market, at the bazaar or on the streets, to virtual or online environments. Such environments now satisfy individuals' needs such as chat, entertainment, education, and shopping. Taking these needs into online environments means that some of the negative daily situations are also being taken to online environments. In particular, children can experience adverse circumstances such as sharing their personal information, sending pornographic pictures or making financial commitments (Livingstone, Bober and Helsper, 2005; Stahl and Fritz, 2002), receiving face-to-face meeting requests from strangers (Liau, Khoo, and Ang, 2005), and exposure to violence (Dehue, Bolman and Völlink, 2008) in online environments. As such, children, as individuals, need to take on responsibilities against risky content and requests inevitably encountered in online environments (Salifu, 2008) and gain some abilities as part of these responsibilities. One of these abilities is the ability to say "no" to inappropriate and unreliable requests and content (Bal & Kahraman, 2015).

The ability to say "no," which is important in terms of the above-mentioned situations, has been examined under the 2005 life studies curriculum (Ministry of National Education [MNE], 2009) within the scope of the skills of "providing security and protection." In this context, it is aimed to teach students to say "no" without feeling guilty, to say "no" and to explain why they say "no", and to suggest an alternative option after refusing and explaining the reason. It is also aimed that students will be able to say "no" effectively in the context of self-confidence, one of the personal qualities tried to be gained by students. Therefore, the 2005 life studies curriculum has covered a comprehensive, open and direct approach in terms of the ability to say "no." However, instead of focusing on the ability to say "no," the "self-protection ability", regarded as a superior ability in terms of conceptual hierarchies, was addressed in the life studies curricula published in 2015 and 2017 (MNE, 2015; 2017). In these curricula, the ability to say "no" is limited only to indirect gains in that refusals should be made within the framework of courtesy rules. From this standpoint, the relevant curricula have shown a limited, indirect and implicit approach. This change may indicate that the relevant ability is not considered too much in the life studies curriculum. Yet, scientific reasoning is required in order for such a thought to be justified. In fact, there is no research in the literature on whether or not the ability has been gained by students, so there is no scientific ground for this change. Therefore, the present study aims to reveal whether or not the ability to say "no" has been gained by children in terms of various situations in so as to allow discussing the change about this ability on scientific grounds.

2. METHOD

Research Model

The survey model was utilized in this study. The survey model is a research model in which the opinions of a large group of people about a particular issue or topic are sought (Fraenkel & Wallen, 2006).

Study Group

This study was conducted with 4th grade students who took the course of life studies in primary school 1st, 2nd and 3rd grades and therefore are expected to have gained the ability to say "no". Within this scope, the application was carried out with a total of 275 students in the 4th grades of six primary schools determined to be able to reach 200-300 students by paying attention to represent the districts of Diyarbakır province (Bağlar, Sur, Yenişehir, Kayapınar) at least once. However, some of the data collection tools obtained from 275 students was excluded from the evaluation due to illegible or improper filling. Therefore, 266 data collection tools were examined within the scope of the research. Considering the Many-Facet Rasch Model, since this number is over 200, it can be said that the accuracy of the estimates obtained through the relevant data collection tool is sufficiently high (DeMars, 2010).

Data Collection Tool

In order to collect data in the scope of the present research, The Ability to Say "No"-Specific Cases Inventory (NSCI) developed by the researchers was utilized. NSCI consists of five open-ended questions regarding sample cases to measure students' ability to say "no" about substance use, sexual abuse, time management, moral problems, and online environments that students may encounter in real life. Through the NSCI rubric, students' answers to these questions are examined in terms of their performances to say "no", to make an explanation, and perform alternative behaviors. However, a total success score is not calculated for student success by evaluating their performances to say "no", to make an explanation, and perform alternative behaviors together. The reason behind this is that saying "no" is considered as an assertiveness right in this study and it is not accepted as an obligation to make an explanation or perform alternative behaviors after saying "no". The students' ability to say "no" in the face of all the cases included in the NSCI is regarded as a fundamental performance and students are expected to exhibit this performance. However, making an explanation about the reasons for saying "no" or performing alternative behaviors after saying "no" is not considered a necessary behavior for the students; such behaviors can only be examined in terms of determining the potential saying "no" attitudes of students, interpreted contextually, or assessed within the framework of assertiveness rights.

The Davis (1992) technique was employed to ensure the content validity of the cases prepared in line with the purposes of the research and the literature. According to the expert opinions obtained in this technical framework, all the validity indices of the five cases in NSCI vary between 88% and 100% and are above 80%. These indices can be interpreted as an indication that all cases will remain in the inventory. Although validity indices are quantitative evidence for the validity of the relevant case studies, experts' opinions of "quite relevant" and "somewhat relevant" on the cases have also been taken into consideration in the data collection tool.

In order to be able to evaluate the responses given to the cases in the NSCI, the "NSCI Rubric" was developed. In this direction, initially, the NSCI Rubric was designed as a general rubric to evaluate all the five cases specified in the NSCI in the framework of common criteria. Second, an analytical structure was established in the scope of sub-dimensions to measure three sub-performances determined within the scope of ability to say "no" in the 2005 life studies curriculum (MNE, 2009). These three sub-performances can be described as saying "no" without the feeling of guilt; saying "no" and explaining the reason of saying "no", and proposing another option after refusing and explaining the reason. In the framework of these sub-performances, the performance dimensions regulated in the NSCI Rubric include saying "no," making an explanation and performing alternative behaviors.

For the purpose of providing validity and reliability in this study (Kutlu, Doğan & Karakaya, 2010), opinions were obtained from eight experts first, and then from two classroom teachers regarding NSCI Rubric; thus, preliminary application of NSCI was conducted. In this context, taking into consideration the data obtained from the classroom teachers and experts, both the cases in the NSCI and the criteria in the NSCI Rubric have been made clearer and more comprehensible. Opinions from classroom teachers made it possible to think that the relevant rubric measures the ability to say "no" in the face of selected cases. The experts commented that the criteria contained in the NSCI Rubric were not overlapping, the descriptive explanations for the criteria were sufficient, the descriptive explanations of rating descriptions accurately reflected the ratings, and the rating numbers reflected the success differences. Taking into account the expert opinions, the students were asked "What do you tell him/her?" and "Why do you say this?" following the case studies. Later, pre-application was carried out with 49 students in two 4th grade classrooms in a primary school. Based on the pre-application, the cases in NSCI were made clearer and more comprehensible, and small contextual changes were made to these cases. In this respect, the performance dimension called "proposing another option" was called "performing alternative behavior." The descriptive definition of "proposing negative options" under this performance dimension was changed to "proposing negative options or performing negative alternative behaviors." The descriptive definition of "proposing positive options" was organized as "proposing positive options or performing positive alternative behaviors." Finally, a table of explanations was added just below the rubric, with explanations of the criteria and descriptive definitions, to ensure convenience for those who will rate answers to the NSCI using the NSCI Rubric. In this table of explanations, there are some literature and pre-application-based definitions and potential answers identified in literature and pre-application.

After the pre-application, the final scale was applied to the students in the actual application and the students' reactions to the related case studies were investigated. Students' reactions to the NSCI were evaluated by three raters using the NSCI Rubric. The measurement reports for the rater facet obtained by many-facet Rasch analysis are presented in Table 1.

Table 1. *The Measurement Reports for the Rater Facet in the Many-Facet Rasch Analysis*

Rater	Logit	Standard Error	Infit MnSq	Outfit MnSq
3	,06	,02	,97	,97
2	-,02	,02	1,01	1,02
1	-,05	,02	1,01	1,04
Mean	,00	,02	1,00	1,01
S.D. (Population)	,04	,00	,02	,03
S.D. (Sample)	,06	,00	,02	,03
Model, Sample: RMSE: ,02; Standard Deviation.: ,05; Separation: 2,66; Reliability: ,88				
Model, Fixed (all same) chi-square: 16,1; d.f.: 2; p=,00				
Exact agreements between raters: 85,5%				
Expected agreement between raters: 35,7%				
Kappa statistics related to inter rater reliability: ,77				

According to the measurement reports for the rater facet in the many-facet Rasch analysis shown in Table 1, the separation ratio is 2.66; and the reliability index is 0.88. Determined in terms of "model, sample," this reliability coefficient reflects the difference between the severity and leniency levels of the raters. The value of 0.88 indicates that the raters differ from each other in terms of severity and leniency. The p (0.00) value obtained for the "model, fixed (all same) chi-square" is interpreted that the variation in these ratings is significant. The essential statistics that should be examined regarding how well the raters make consistent, in other words reliable, ratings are the exact agreement among raters statistic and kappa statistic. The exact agreement between raters statistic indicates how three different raters make a consistent evaluation related

to the responses given to the cases included in the NSCI. This value should be at least 75%; but it is expected to approach 90%. Therefore, it can be said that the reliability level (85.5%) among the raters in this study is high. The fact that the Kappa statistic (0.77) is between 0.61 and 0.81 enables us to make a comment that there is an agreement among the raters (Graham, Milanowski & Miller, 2012). At the same time, the fact that infit and outfit statistics are around 1.0 means that the raters performed their ratings as independent as possible and that they didn't experience too much disagreement (Linacre, 2014). These results indicate that the measurements for evaluating the data collected with NSCI were performed reliably by using NSCI Rubric.

One of the basic facets examined by the many-facet Rasch analysis in this research is the "examinee" facet. The data for this facet signifies that the logit value of the examinee with the highest performance in terms of the ability to say "no" in the face of specific cases is 1.95; the logit value of the examinee with the lowest performance in this case is -4.02. These values imply that the ability of students to say "no" in the face of specific cases has been evaluated over a wide range. Since the logit values for all of the 266 examinees covered in this study cannot be given, only the mean logit values, standard errors, infit and outfit statistics of all individuals are given in Table 2.

Table 2. *The Measurement Reports for the Examinee Facet in the Many-Facet Rasch Analysis*

	Logit	Standard Error	Infit MnSq	Outfit MnSq
Mean	,49	,19	1,05	1,01
S.D. (Population)	,74	,03	,72	,80
S.D. (Sample)	,74	,03	,72	,80
Model, Sample: RMSE: ,19; Standard Deviation.: ,71; Separation: 3,81; Reliability: ,94				
Model, Fixed (all same) chi-square: 2956,0; d.f.: 265; p=,00				

According to the measurement reports for the examinee facet in the many-facet Rasch analysis given in Table 2, infit and outfit statistics are 1.05 and 1.01, respectively. The fact that these values are close to 1 suggests that there is very little distortion in the measurement system. The fact that the average outfit statistic is close to 1 proposes that the data obtained fit the model (Brentari and Golia, 2008). The separation ratio of 3.81 and the reliability index of 0.94 indicate that the measurements related to the NSCI made with the NSCI Rubric can distinguish examinees in terms of their ability to say "no" in the face of specific cases. Moreover, "the model, fixed (all same) chi-square" statistic of 2956.0 and the p value of 0.00 can be interpreted as a significant difference in the students' performance to say "no" in the face of specific cases (İlhan, 2015).

The ultimate facet that can be examined in the many-facet Rasch analysis is the item facet. The measurement reports obtained with the multi-facet Rasch analysis for the item facet is given in Table 3.

A measurement report should have a separation ratio of at least 2 (Linacre, 2012) and a reliability index of at least 0.70 (Walker, Engelhard, and Thompson, 2012). Table 3 reveals that the separation ratio for the items in NSCI is 19.19 and the reliability index is 1.0. These statistics signify that the items in the NSCI differ from each other in terms of the level of difficulty (İlhan, 2015). This is considered to be a significant difference due to p value (<0.05). Unfit and outfit statistics between 0.5 and 1.5 indicate that the model-data fit is gained (Brinthaupt and Kang, 2014).

Table 3. *The Measurement Reports for the Item Facet in the Many-Facet Rasch Analysis*

Madde	Logit	Standard Error	Infit MnSq	Outfit MnSq
3	1,31	,05	,51	,55
15	,75	,04	,72	,72
14	,67	,04	,84	,87
12	,63	,04	,78	,84
2	,59	,04	1,09	1,20
6	,49	,04	,88	,97
9	,44	,04	1,04	1,27
5	,39	,04	,90	,98
1	,26	,04	,83	,98
8	-,22	,04	1,22	1,29
1	-,56	,04	1,44	1,26
4	-,73	,04	,92	,88
13	-1,24	,05	1,33	1,03
7	-1,32	,06	1,47	1,38
10	-1,46	,06	1,35	,95
Mean	,00	,04	1,02	1,01
S.D. (Population)	,83	,01	,28	,22
S.D. (Sample)	,86	,01	,29	,23

Model, Sample: RMSE: ,04; Standard Deviation: ,86; Separation: 19,19; Reliability: 1,0 Model, Fixed (all same) chi-square: 4214,7; d.f.: 14; p=,00

Within the many-facet Rasch analysis, category statistics are given after the measurement reports for each facet. These statistics provide evidence whether the 4-point rating scale (0-3) in the NSCI Rubric works well. The category statistics obtained for this study are shown in Table 4.

Tablo 4. *Category Statistics for NSCI Rubric in the Many-Facet Rasch Analysis*

Scale category	f	%	Cumulative %	Average Measure	Expected Measure	Outfit MnSq
0	1604	13	13	-,58	-,68	1,4
1	4845	40	54	-,03	,05	,5
2	912	8	61	,97	,65	1,2
3	4609	39	100	1,33	1,34	1,2

As it can be inferred from Table 4, there are more than 10 observations in each scale category (0-3), as needed. In addition, as the category level of the rating scale increases (from 0 to 3), the average measures (from -58 to 1.33) also increase. Outfit statistics are distributed between 0.5 and 1.5. These three findings can be regarded as evidence that the rating scale in the NSCI Rubric works well (Linacre, 2014).

Data Analysis

The data obtained from the NSCI were analyzed according to the many-facet Rasch analysis using the FACETS packet program. Rasch is a logistic model with one parameter based on item-response theory, indicating the relationship between the ability or property measured by a measuring tool and the response given to an item (DeMars, 2010). Since this model is based on item-response theory, it gives more weight to distinctive and reliable items, taking into account item difficulty or item discrimination during scoring. Thus, more reliable results are obtained compared to the raw scores (Baker, 2016). In this study, examinee, item, and rater facets were processed in the analysis performed according to the many-facet Rasch analysis. In order to carry out the analysis, the model-data fit hypothesis was tested first. For this purpose, the measurement reports for the item facet in the many-facet Rasch analysis (see Table 3) were

examined. By reason of the fact that the infit and outfit statistics given in this report range from 0.5 to 1.5, it is considered that the model-data fit assumption required for the analysis is provided (Brinthaupt & Kang, 2014). Since ensuring model-data fit connotes that the assumptions of unidimensionality and local independence required for the many-facet Rasch analysis are also met, the findings from the analysis are presented.

3. FINDINGS

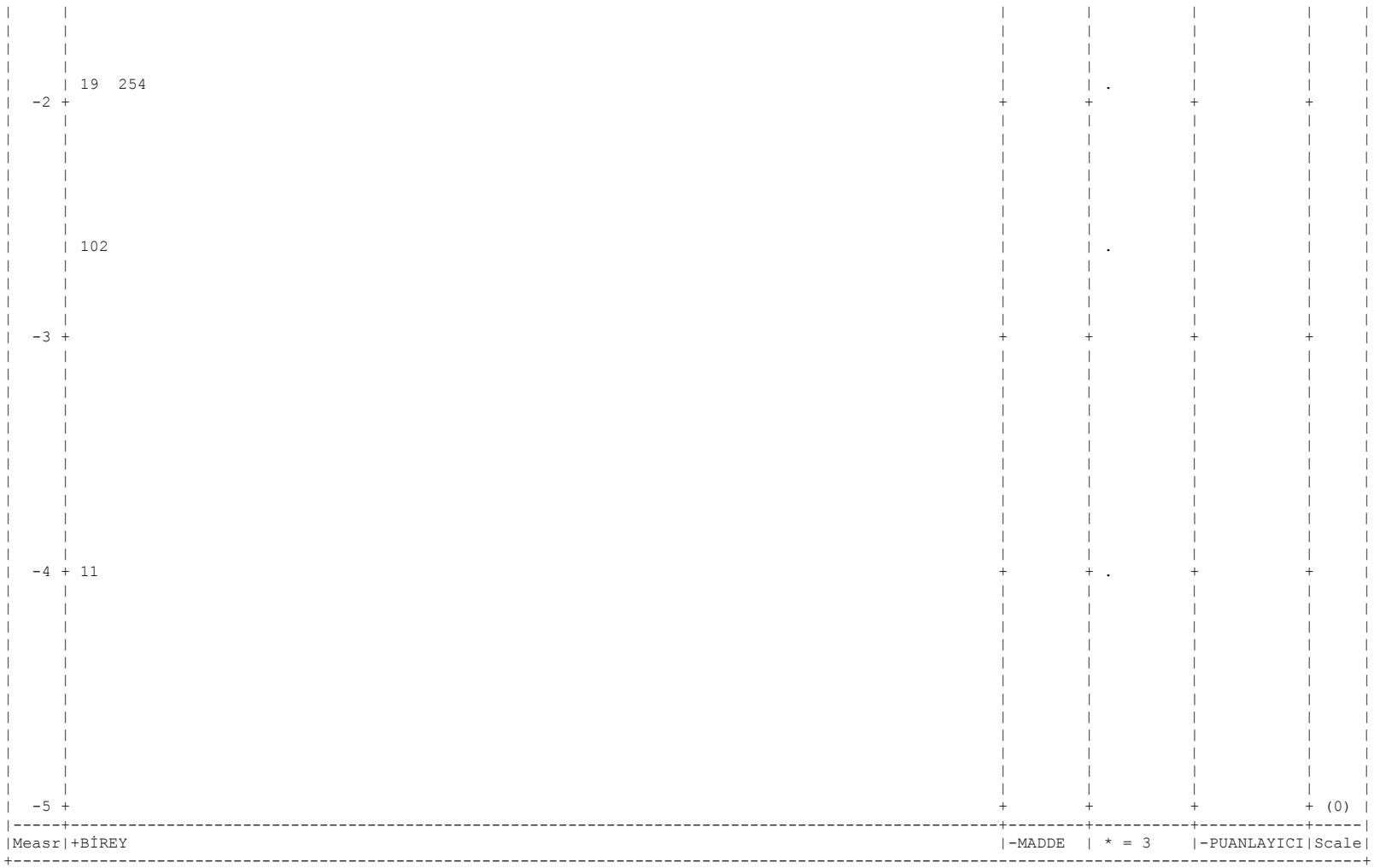
Within the scope of this study, five different cases (substance use, sexual abuse, time management, moral problems, and online environment requests) were presented to primary school 4th graders to find about their abilities to say "no" to specific cases. The responses of the students related to the reactions they will exhibit in the event of these five cases were rated by three independent raters using NSCI Rubric. These ratings were analyzed according to the many-facet Rasch analysis. The variable map created by the many-facet Rasch analysis is given in Table 5.

The first column of Table 5 contains the logit scale and the last column contains the rating scale. The other columns include item, examinee, and rater facets. In the examinee column, the level of students' performance and the degree of difficulty of the item increase from bottom to top. In order to make better comments on this issue, the logit scale needs to be examined. Examinees with more positive values on the Logit scale refer to the examinees with higher performance levels, while items corresponding to positive values refer to more difficult items (Bond & Fox, 2001). Therefore, negative values in this respect indicate individuals more unsuccessful in relation to related performance in terms of examinee facet, while these values indicate performance tasks realized more easily in terms of item facet. When the many-facet Rasch model obtained in this study is examined, it is understood that most of the examinees included in the study correspond to positive values in terms of logit scale. This indicates that the majority of the 4th graders participating in this study had a high ability to say "no" in the face of specific cases. The students with the highest ability to say "no" in the face of specific cases are the students 40 and 151, while the students with the lowest performance in this respect are the students 102 and 11. When the item facet of the Rasch model is examined, it is seen that the highest performance of the students is seen for the item 10 (being able to say "no" in the face of moral problems), followed by the 7th (being able to say "no" in terms of time management), 13th (being able to say "no" in terms of online environments), 4th (being able to say "no" in terms of sexual abuse), 1st (being able to say "no" in terms of substance use), and 8th item (making an explanation in terms of time management), respectively. The other items are the ones that students have difficulty to perform well because they are positive in terms of logit values. Among these items, the students have the poorest performance for item 3 (performing alternative behaviors in terms of substance use), followed by 15th (performing alternative behaviors in terms of online environments), 14th (making an explanation in terms of online environments), 12th and 2nd (performing alternative behaviors in terms of moral problems and making an explanation in terms of substance use), 6th and 9th (performing alternative behaviors in terms of sexual abuse and time management), 5th (making an explanation in terms of sexual abuse), and 11th item (making an explanation in terms of moral problems), respectively. These findings suggest that the students may be able to demonstrate the basic ability to say "no" expected of them in specific cases; but they will not focus too much on making an explanation about the reasons why they have said "no" and performing alternative behaviors after saying "no."

Table 5. *The Variable Map of the Ability to Say "No" to Specific Cases*

```
+-------------------------------------------------------------------------------------------------------------------------------------------------------+
|Measr|+BİREY                                                                                                  |-MADDE  |+BİREY    |-PUANLAYICI|Scale|
|-----+--------------------------------------------------------------------------------------------------------+--------+----------+-----------+-----|
|   2 +                                                                                                        +        +          +           + (3) |
|     | 40  151                                                                                                |        | .        |           |     |
|     |                                                                                                        |        |          |           |     |
|     |                                                                                                        |        |          |           |     |
|     | 28  29                                                                                                 |        | .        |           |     |
|     | 20  32  136 253                                                                                        |        | *.       |           |     |
|     | 71  150 165 166 237 255                                                                                |        | **       |           |     |
|     | 88  105 145                                                                                            |        | *        |           |     |
|     | 35  73  75  76                                                                                         |        | *.       |           | --- |
|     | 77  127 128 169                                                                                        | 3      | *.       |           |     |
|     | 23  27  56  129 134 144 168 199 250                                                                    |        | ***      |           |     |
|     | 22  37  55  74  94  99  193 226 256 266                                                                |        | ***.     |           |     |
|     | 49  80  100 101 104 162 188 194 197 198 245 257                                                        |        | ****     |           |     |
|   1 + 66  84  85  131 147 164 173 190 259                                                                  +        + ***      +           +     |
|     | 24  41  53  135 171 181 212 244                                                                        |        | **.      |           |     |
|     | 33  42  58  59  64  79  89  92  95  98  112 139 161 176 177 179 219 220 232 246 247 261                |        | *******. |           |     |
|     | 17  18  38  44  48  69  82  180 206                                                                    | 15     | ***      |           |  2  |
|     | 9   14  45  51  72  86  103 123 125 126 137 138 167 183 200 213 249 258                                | 14     | ******   |           |     |
|     | 3   4   25  26  47  81  96  122 146 172 178 191 209                                                    | 12  2  | ****.    |           |     |
|     | 1   21  34  36  43  106 109 121 149 182 186 192 201                                                    |        | ****.    |           |     |
|     | 2   7   30  50  52  54  60  62  78  87  107 153 158 160 174 184 187 189 210 230 239 251 262 265        | 6   9  | ******** |           |     |
|     | 46  63  67  93  152 248 252 263                                                                        | 5      | **.      |           |     |
|     | 10  57  68  70  83  140 207 222 238                                                                    |        | ***      |           | --- |
|     | 12  133 163 240                                                                                        | 11     | *.       |           |     |
|     | 108 124 132 154 175 204 217 218 224 233                                                                |        | ***.     |           |     |
|     | 90  91  110 117 130 170 205 211 225 264                                                                |        | ***.     | 3         |     |
*   0 * 5   115 118 120 156 185 215 228 229 241 260                                                          *        * ***.     * 2         *     *
|     | 116 208 214                                                                                            |        | *        | 1         |     |
|     | 61  141 159 195                                                                                        |        | *.       |           |     |
|     | 16                                                                                                     | 8      | .        |           |     |
|     | 111 143 223 227 242                                                                                    |        | *.       |           |  1  |
|     | 114 142 235 243                                                                                        |        | *.       |           |     |
|     | 113 221                                                                                                |        | .        |           |     |
|     | 39  97  203                                                                                            | 1      | *        |           |     |
|     | 31  148 202 234                                                                                        |        | *.       |           |     |
|     | 6   65  157                                                                                            |        | *        |           |     |
|     | 13                                                                                                     | 4      | .        |           |     |
|     | 155                                                                                                    |        | .        |           |     |
|     |                                                                                                        |        |          |           |     |
|  -1 + 15  236                                                                                              +        + .        +           +     |
|     |                                                                                                        |        |          |           |     |
|     |                                                                                                        |        |          |           |     |
|     | 8   119 196                                                                                            | 13     | *        |           |     |
|     |                                                                                                        | 7      |          |           |     |
|     | 231                                                                                                    |        | .        |           |     |
|     |                                                                                                        | 10     |          |           |     |
|     | 216                                                                                                    |        | .        |           | --- |
```

Table 5. Continue



4. DISCUSSION, RESULTS AND RECOMMENDATIONS

The written responses of the 4th graders to the specific cases examined by NSCI imply that they may be highly capable of saying "no" in various cases. When this performance of the students is examined in terms of the behaviors planned to be taught as part of the ability to say "no" in the 2005 life studies curriculum (MNE, 2009), there is no serious problem in terms of the behavior of saying "no". Given the students' responses to the specific cases included in the study, it can be inferred that they can easily say "no" in the face of inappropriate situations.

The study results indicate that the students' ability to say "no" is high in the face of events containing moral problems, time management, and online environments, respectively. However, when it comes to sexual abuse or substance use, they show relatively low performance. The fact that the students exhibited the highest performance for the ability to say "no" in the face of a case with a moral problem in the context of violence is consistent with the fact that Astor (1994) found that all children condemned the violence. The fact that students can say "no" in the face of a moral problem means that they have an instinct to conduct the behaviors regarded to be right according to their moral common sense. This is interpreted as the possibility that they may have developed a moral character at a certain level (Lickona, 1996). Considering the fact that moral character development is largely shaped during childhood, when children are most expected to feel guilty and develop perspectives and self-control as a meta-moral characteristic (Berkowitz, 2002), this result becomes more of an issue. In this study, students responded that they could say "no" in the context of violence and aggression; thus it is concluded that they have moral information and feeling that they should exhibit self-control in a meta-moral sense by blocking their aggressive feelings. The extent to which moral information and feelings direct real-life moral behaviors cannot be interpreted clearly within the limitations of this study; but it is expected to be so.

The results also suggest that a majority of primary school 4th grade students participating in this study can say "no" in the face of the case related to time management as well as the case with a moral problem, indicating that the students have great self-management skills (Güçlü, 2001) and time control perceptions (Macan, Shahani, Dipboye & Philips, 1990). Considering the positive relationship between time management and academic achievement (Forsyth, 2009), this result reflects a desirable situation. Glenn (2003) states that one of the skills that must be possessed in terms of time management is the ability to "set priorities". The results of this study demonstrate that students can set priorities in terms of actions such as "doing their homework" or "satisfying others by fulfilling their non-urgent requests" when considering the content of the time management-related case presented to the students. Therefore, it is inferred that the students participating in this research will not feel bombarded and overloaded (Lovely and Smith, 2004) if they can convert their thoughts about the ability to say "no" to real-life behaviors. According to a study conducted by Livingstone, Haddon, Görzig, and Ólafsson (2011), 14% of children who made online contact with people they did not know, shared their photo/video with these people and 15% of them shared their personal information with these people. Notwithstanding the fact that the majority of the children are not involved in such activities, these ratios still demand urgent intervention.

Likewise, this research has made it clear that the majority of the students can say "no" in the face of the risky situations they may encounter including sharing images and personal information with strangers in online environments. Considering that the vast majority of primary school students use the internet to help them with their homework both in Turkey (Yolcu, 2007) and in other countries across the world (La Ferle, Edwards, and Lee, 2000), this result can be interpreted as the awareness about the proper use of the internet. Since this study does not include an external control element such as parents or secure internet software in the case measuring being able to say "no" in the online environments, it is also thought that the

students have an internal control of sharing personal information and using video chat. When the possibility of the lack of external controls (Berson, Berson and Ralston, 1999) is taken into consideration, this internal control can express a more effective and more important quality when it comes to protecting oneself from the risks of the internet. Although all these conclusions and interpretations support the finding that the tendency of insecure use of the internet has diminished over time, as suggested by the longitudinal studies of Valcke, De Wever, Van Keer and Schellens (2011), they do not imply that all individuals use the internet in a safe manner. Still, there are individual differences in the use of insecure internet. As a matter of fact, this study has revealed that there are students who are inclined to share their personal information with strangers, or talk with them to get help for their homework.

The studies carried out by Leclerc, Wortley, and Smallbone's (2011) indicate that children apply the saying "no" strategy mostly, after the strategy of stating that they do not want it in the face of sexual abuse. The results of this study also indicate that a remarkable majority of participants may say "no" to sexual abuse. This strategy is important in that it creates a fear (Elliott, Browne, and Kilcoyne, 1995) in the individual who intends to exploit; thus, it may lead to a backwards step. However, in this study, it is understood that the results for the cases including sexual abuse reflect a relatively low level of "saying no" compared to the cases including moral problems, time management, and online environments. The underlying reason for this is mainly thought to be the fact that some of the students may not have perceived the "peeking" phenomenon discussed in the relevant case as a sexual abuse, or that they do not anticipate that there is such a possibility within the scenario of the case. However, given that one of the reasons for children's exposure to sexual abuse is deception (Ceylan et al., 2009) and that they cannot perceive the scenario of helping, showing interest, and love (Polat, 2001), it is necessary to point out that some students in this study are at least exposed to the risk of sexual abuse at the level of peeking.

According to Bektaş (2009), the use of substances by individuals can be reduced through the acquisition of social skills such as the ability to say "no." This study shows that primary school 4th grade students have the foreknowledge and awareness that they can use their ability to say "no" to reject the substance use. Therefore, it is estimated that the rate of substance use among children can be reduced if the foreknowledge and awareness are converted to behavior by the students. Although the context of the sample case including substance use in this study was designed to make students curious, it is understood that students will not be affected by this emotion that may be effective in substance use (Kamışlı, Karatay, Terzioğlu, and Kublay, 2008). All these results can be regarded positive results in terms of saying "no" to substance use. However, despite these positive results, it is noteworthy that the rate of saying "no" to the substance use is lower than the rate of saying "no" to all the other cases. It is thought that this situation is caused by the fact that the related case is given in "play" scenario. Therefore, in terms of substance use, it is possible to deduce that playmates and gaming environments may be decisive for at least some children.

According to the results obtained from the research, the students are at a desired level in terms of being able to say "no". However, although they make explanations related to why they have said "no" in the cases regarding time management, they do not reflect such a tendency toward other specific cases. At the same time, the explanations made by the students suggest that they will not perform an alternative behavior after saying "no" regardless of the specific case concerned. These results can be perceived completely negative only when the goals of the 2005 (MNE, 2009) life studies curriculum related to the ability to say "no" are to be considered. This is because, according to these results, most of the students state that they will not make an explanation about the reasons for saying "no" and they will not perform alternative behaviors even if they say "no" in the face of specific cases. However, it should be noted that there is no

obligation in this regard when considering the context of the cases presented to them and assertiveness rights (Kemp, 2006). In this context, there is a need to discuss the reactions that students have stated in response to each case.

The case related to the substance use in this study can be accepted as a case in which saying "no," making an explanation about the reasons for saying "no" and performing an alternative behavior are necessary. Therefore, the students participating in this study were expected to make an explanation in order to inform their friends of the reasons for saying "no" in response to the offer of playing game presented in the form of smelling a substance they are not familiar with. Besides, suggesting alternatives such as playing another game to rescue themselves and their friends from the risky situation they are in, or performing positive alternative behaviors such as leaving the scene against the risk of not being able to persuade their friends, can also be considered as logical approaches.

Considering in the context of real life, the case related to the sexual abuse is approached as a case in which saying "no" and performing an alternative behavior are necessary and making an explanation is preferable. In this case, in which students have already said "no" to protect the private parts of their body, making an explanation to the other side depends entirely on how qualified the student can make this explanation. An explanation of the students that "they can only bath in their own home" may not harm them or be regarded as a necessity. However, the students must perform alternative behaviors such as getting away from the scene and sharing this with their family because their clothes are contaminated and private parts of their body can be peeped out.

In the current study, the case measuring the ability to say "no" in the case regarding time management can be considered as a case in which saying "no" and making an explanation about the reasons for saying "no" are necessary; but performing an alternative behavior is not a necessity in this matter. In this context, in terms of time management, the fact that participants made an explanation about the reasons for saying "no" to their neighbors can be interpreted as a demonstration of kindness and can be accepted as a right behavior. However, considering the assertiveness rights, it is considered that the students do not have the obligation to perform alternative behaviors by postponing the demand of the neighbors because they do not have time.

Likewise, in the case of substance use, the case with a moral problem can be evaluated as a case in which students have to say "no," make an explanation about the reasons for saying "no" and perform alternative behaviors. However, it is seen that students do not prefer to make an explanation about the reasons for saying "no" to violence against an old person to their friends, even if they have said "no" in response to this case. Furthermore, there are no signs indicating that students may perform an alternative behavior in this regard. Yet, it is predicted that it may prevent other friends, as well as themselves, from the same misconduct to make an explanation about why it is wrong to resort to violence against an elderly person or perform alternative behaviors to solve the conflict between the old man and other children.

Students may face a variety of risks in online settings. Therefore, the students are expected to say "no" or behave in a way that could mean "no" in the face of proposals and requests from strangers. In this context, students' efforts to make an explanation about the reasons for saying "no" to a stranger can cause the other party to try to persuade them. However, in such a context, students may be expected to employ exit strategies such as warning, blocking, or reporting the individuals disturbing them in the online environment as alternative behaviors as Tynes (2007) suggests.

In a general sense, the fact that 4th grade students in elementary school gave written answers that they could say "no" to all the cases reflect the desired situation. However, the fact that students have not made any explanations and performed an alternative behavior in the face of the cases related to moral problems and substance use and that they have not performed an

alternative behavior in the face of the cases related to online environments and sexual abuse can be considered as a problem.

The present study shows that 4th grade students in primary school showed a certain level of cognitive and affective readiness to say "no" in specific cases. Nevertheless, a definite comment on the extent to which this level of readiness can lead to the behavior of saying "no" in real life cannot be made depending on the data obtained within the limits of this research. However, it may well be said that both parents and teachers need to evaluate and process this level of readiness of the children with regards to saying "no." In this regard, it is suggested that they should create opportunities that will allow the children to say "no" in the safe environment of the home or school.

The results obtained from the research indicate that the students can say "no" in order to protect themselves; but it is thought that they can perform relatively poorly in this respect when it comes to the cases including substance use or sexual abuse. Therefore, individuals and commissions who will carry out program development studies within the life studies curriculum need to extend the scope of the specific cases of saying "no" to handle high-risk events such as substance use and sexual abuse. Moreover, it is expected that these individuals and commissions should enrich the content of the life studies curriculum with information on how the demands, offers, and behaviors of sexual abuse and substance use can be manipulated.

Notwithstanding the fact that many of the students' performance of saying "no" in specific cases seem satisfactory, their performances in making an explanation and performing an alternative behavior remain below a certain level for many specific cases irrespective of the context. In order to solve this problem, the program development working groups coming together to develop life studies curriculum should determine first, in which specific cases the behaviors of explaining the reasons for saying "no" and suggesting an alternative option after rejecting and explaining the reason can be true, in which cases they can be wrong, and in which cases such behaviors can depend on preferences. Later, students should be motivated to exhibit these behaviors in situations where it is acceptable to exhibit them. As a solution strategy for the situations in which saying "no" is necessary, students should be expected not only to suggest another option but also perform a positive alternative behavior when no other option is possible. In addition, students should be taught what qualifications the explanations they will make should have in the face of a situation in which they have said "no" and which options they will suggest and which behaviors they will perform can solve the problems arising out of the situations in which they have said "no".

ORCID

Ferat Yılmaz https://orcid.org/0000-0002-4947-5416
M. Akif Sözer https://orcid.org/0000-0002-1291-4067

5. REFERENCES

Aktepe, E., Işık, A., Kocaman, O., & Eroğlu, F. Ö. (2013). Demographic and clinical characteristics of children and adolescents examined in a university hospital who are victims of sexual abuse. *New/Yeni Symposium Journal, 51*(2), 115-120.

Aslan, D., & Özcebe, H. (2008). *Eğitim kurumlarında sigarasızlık politikaları* [*No smoking policies in educational institutions]*. Ankara: Klasmat.

Astor, R. A. (1994). Children's moral reasoning about family and peer violence: The role of provocation and retribution. *Child Development, 65*(4), 1054-1067. doi:10.2307/1131304

Aytaç, Ö. (2005). Modern bureaucracies and alienation ethos. *Fırat University Journal of Social Science, 15*(2), 319-348.

Baker, F. B. (2016). *Madde tepki kuramının temelleri [The basics of item response theory]* (M. İlhan, Trans.). Ankara: Pegem Akademi.

Bal, P. N., & Kahraman, S. (2015). The effect of cyber bullying sensibility ımprovement group training program on gifted students. *Journal of Gifted Education Research, 3*(1), 48-57.

Bektaş, M. (2009). *Investigation of effect of developed smoking prevention program toward primary school student*. (Unpublished doctoral dissertation). University of Dokuz Eylül, İzmir.

Belgrave, F. Z., Reed, M. C., Plybon, L. E., & Corneille, M. (2004). The impact of a culturally enhanced drug prevention program on drug and alcohol refusal efficacy among urban African American girls. *J. Drug Education, 34*(3), 267-279. doi:10.2190/H40Y-D098-GCFA-EL74

Berkowitz, M. W. (2002). The science of character education. In W. Damon (Ed.), *Bringing in a new era in character education* (pp. 43-63). Stanford, CA: Hoover Institution.

Berson, M. J., Berson, I. R., & Ralston, M. E. (1999). Threshing out the myths and facts of internet safety. *Social Education, 63*(3), 160-161.

Bilginer, Ç., Hesapçıoğlu, S. T., & Kandil, S. (2013). Sexual abuse in childhood: A multi-dimentional look from the view point of victims and perpetrators. *The Journal of Psychiatry and Neurological Sciences, 26*(1), 55-64. doi:10.5350/DAJPN2013260106

Bolton, R. (1979). *How to assert yourself, listen to others, and resolve conflicts.* New York: Simon & Schuster.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. London: Lawrence Erlbaum.

Boztaş, M. H., & Arısoy, Ö. (2010). Inhalant dependence and its medical consequences. *Current Approaches in Psychiatry, 2*(4), 516-531.

Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement, 9*(1), 45-56.

Brinthaupt, T. M., & Kang, M. (2014). Many-faceted rasch calibration: An example using the self-talk scale. *Assessment, 21*(2) 241−249. doi:10.1177/1073191112446653

Calabrese, R. L. (1989). Effective principals must be ethical: Ethics for principals. *The Education Digest, 54*(7), 16-19.

Ceyhun, B., Oğuztürk, Ö., & Ceyhun, A. G. (2001). The reliabilitiy and validity of substance abuse proclivity scale. *Turkish Journal of Clinical Psychiatry, 4*, 87-93.

Ceylan, A., Tuncer, O., Melek, M., Akgün, C., Gülmehmet, F., & Erden, Ö. (2009). Sexual abuse of children in the Van region. *Van Medical Journal, 16*(4), 131-134.

Claessens, B. J., Eerde, W. v., Rutte, C. G., & Roe, R. A. (2007). A review of the time management literature. *Personnel Review, 36*(2), 255-276. doi:10.1108/00483480710726136

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194-197. doi:10.1016/S0897-1897(05)80008-4

Dehue, F., Bolman, C., & Völlink, T. (2008). Cyberbullying: Youngsters' Experiences and Parental Perception. *CyberPsychology & Behavior, 11*(2), 217-223. doi:10.1089/cpb.2007.0008

DeMars, C. (2010). *Item response theory*. New York: Oxford.

Elliott, M., Browne, K., & Kilcoyne, J. (1995). Child sexual abuse prevention: What offenders tell us. *Child Abuse & Neglect, 19*(5), 579-594. doi:10.1016/0145-2134(95)00017-3

Forsyth, P. (2009). *100 great time management ideas from successful executives and managers around the world.* Sİngapore: Marshall Cavendish.

Glenn, R. E. (2003). Study skills to help kids use time wisely. *The Education Digest, 69*(3), 51-53.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings.* Center for Educator Compensation Reform. Retrieved September 30, 2016, from http://files.eric.ed.gov/fulltext/ED532068.pdf

Güçlü, N. (2001). Time management. *Educational Administration: Theory and Practice*, (25), 87-106.

Güler, N., İlhan, M., Güneyli, A., & Demir, S. (2017). An evaluation of the psychometric properties of three different forms of Daly and Miller's writing apprehension test through rasch analysis. *Educational Sciences: Theory and Practice*, *17*(3), 721-744. doi: 10.12738/estp.2017.3.0051

Gündoğan, A. O. (2009). *Bir isyan ahlakı mümkün müdür? [Is a rebellion morality possible?]* Retrieved March 15, 2015, from http://www.aliosmangundogan.com/PDF/Bildiri/Ali-Osman-Gundogan-Bir-Isyan-Ahlaki-Mumkun-Mudur.pdf

İlhan, M. (2015). *The identification of rater effects on open-ended math questions rated through standard rubrics and rubrics based on the SOLO taxonomy in reference to the many facet Rasch model.* (Unpublished doctoral dissertation). University of Gaziantep, Gaziantep.

Kamışlı, S., Karatay, G., Terzioğlu, F., & Kublay, G. (2008). *Sigara ve ruh sağlığı [Cigarette and mental health]*. Ankara: Klasmat.

Kemp, J. (2006). *Hayır: Basit bir cümle hayatınızı ne kadar değiştirir? [No: How one simple word can transform your life*] (H. Güner, Trans.). İstanbul: GOA.

Korkmaz Ekren, P., Başarık, B., & Özhan, M. H. (2011). Changes in recognition of cigarette brand names and logos among primary school children in two years. *Eurasian Journal of Pulmonology, 13*(2), 86-90.

Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2010). *Öğrenci başarısının değerlendirilmesi: Performansa ve portfolyoya dayalı durum değerlendirme [The evaluation of student success: The assessment based on performance and portfolio]*. Ankara: Pegem Akademi.

La Ferle, C., Edwards, S. M., & Lee, W. N. (2000). Teens' use of traditional media and internet. *Journal of Advertising Research, 40*(3), 1-10. doi:10.2501/JAR-40-3-55-65

Leclerc, B., Wortley, R., & Smallbone, S. (2011). Victim resistance in child sexual abuse: A look into the efficacy of self-protection strategies based on the offender's experience. *Journal of Interpersonal Violence, 26*(9), 1868-1883. doi:10.1177/ 0886260510372941

Leming, J. S. (1997). Whither Goes Character Education? Objectives, Pedagogy, and Research in Education Programs. *Journal of Education, 179*(2), 11-34.

Liau, A. K., Khoo, A., & Ang, P. H. (2005). Factors influencing adolescents engagement in risky internet behavior, *Cyber Psychology & Behavior, 8*(6), 513-520. doi: 10.1089/cpb.2005.8.513

Lickona, T. (1996). Eleven principles of effective character education. *Journal of Moral Education, 25*(1). 93-100. doi:10.1080/0305724960250110

Linacre, J. M. (2012). Many-facet Rasch measurement: Facets tutorial. Retrieved April 24, 2017, from http://www.winsteps.com/a/ftutorial2.pdf

Linacre, J. M. (2014). *A user's guide to FACETS: Rasch-model computer programs*. Retrieved April 24, 2017, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.177.9225

Livingstone, S., Bober, M., & Helsper, E. (2005). *Internet literacy among children and young people: Findings from the UK children go online project*. Retrieved January 04, 2015, from http://eprints.lse.ac.uk/397/

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011) *Risks and safety on the internet: the perspective of European children: full findings and policy implications from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries*. EU Kids Online, Deliverable D4. EU Kids Online Network, London, UK.

Lovely, S., & Smith, S. (2004). Selective abandonment: How and when to say no. *Principal Leadership, 5*(3), 35-38.

Macan, T. H., Shahani, C., Dipboye, R. L., & Philips, A. P. (1990). College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology, 82*(4), 760-768. doi:10.1037/0022-0663.82.4.760

Mackenzie, A., & Nickerson, P. (2009). *The time trap.* New York: American Management Association.

Manktelow, J., & Anand, N. (2008). *Make time for success: Master your workload and live your best life.* Swindon: Mind Tools.

Milli Eğitim Bakanlığı. (2009). *İlköğretim 1, 2 ve 3. sınıflar hayat bilgisi dersi öğretim programı ve kılavuzu [Primary school 1st, 2nd and 3rd grades life studies curriculum and guidance]*. Retrieved February 13, 2016, from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

Milli Eğitim Bakanlığı. (2015). *İlkokul hayat bilgisi dersi (1, 2 ve 3. sınıflar) öğretim programı [Elementary school life studies (1st, 2nd and 3rd grades) curriculum]*. Retrieved February 13, 2016, from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

Milli Eğitim Bakanlığı. (2017). *Hayat bilgisi dersi öğretim programı (İlkokul 1, 2 ve 3. Sınıflar) [Llife studies curriculum (Elementary school 1st, 2nd and 3rd grades)]*. Retrieved September 13, 2017, from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=144

O'Sullivan, S. (2004). Books to live by: Using children's literature for character education. *The Reading Teacher, 57*(7), 640-645.

Ögel, K., & Aksoy, A. (2007). Substance use in deliquent adolescents. *Journal of Dependence, 8*, 11-17.

Özkan, B. (2011). *Ergenlerin cinsel eğitimine farklı bir bakış [A different view of sexual education of adolescents]*. Paper presented at the Conference of Cinsellik ve Cinsel Eğitim [Sexuality and Sexual Education], İstanbul, Turkey.

Özyurt, B., & Dinç, G. (2006). Alcohol drinking prevalance and related factors among school aged children in Manisa. *TAF Preventive Medicine Bulletin,, 5*(2), 61-71.

Passig, D. (2004). Future-time-span as a cognitive skill in future studies. *Futures Research Quarterly, 19*(4), 27-47.

Pereda, N., Guilera, G., Forns, M., & Gómez-Benito, J. (2009). The prevalence of child sexual abuse in community and student samples: A meta-analysis. *Clinical Psychology Review, 29*(4), 328-338. doi:10.1016/j.cpr.2009.02.007

Polat, O. (2001). *Çocuk ve şiddet [Child and violence]*. İstanbul: Der.

Rind, B., Tromovitch, P., & Bauserman, R. (1998). A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin, 124*(1), 22-53. doi:10.1037/0033-2909.124.1.22

Salifu, A. (2008). The impact of internet crime on development. *Journal of Financial Crime, 15*(4), 432-443. doi:10.1108/13590790810907254

Sanderson, C. (2010). *Anne baba ve öğretmenler için çocuğun cinsel eğitimi ve tacizden korunma rehberi [The seduction of children: Empowering parents and teachers to protect children from child sexual abuse]* (Z. Doğan, Trans.). İstanbul: Sistem.

Stahl, C., & Fritz, N. (2002). Internet safety: Adolescents' self-report. *Journal of Adolescent Health, 31*, 7-10. doi:10.1016/S1054-139X(02)00369-5

Stoltenborgh, M., Ijzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M. J. (2011). A global perspective on child sexual abuse: Meta-analysis of prevalence around the world. *Child Maltreatment, 16*(2), 79-101. doi:10.1177/1077559511403920

Szpalski, M., Gunzburg, R., & De Kleuver, M. (2003). Unethical research funding contracts: Just say no! *European Spine Journal, 12*(2), 107-107. doi:10.1007/ s00586-003-0551-x

Taner, H. A., Çetin, F. H., Işık, Y., & İşeri, E. (2015). Psychopathology in abused children and adolescents and related risk factors. *Anatolian Journal of Psychiatry, 16*, 294-300.

Tokur Kesgin, M. (2012). Protection of children from hazards of smoking: Community health nurse and primary responsibilities. *Hacettepe University Faculty of Health Sciences Nursing Journal, 19*(1), 90-96.

Tynes, B. M. (2007). Internet safety gone wild? Sacrificing the educational and psychosocial benefits of online social environments. *Journal of Adolescent Research, 22*(6), 575-584. doi:10.1177/0743558407303979

Ünsal, A., & Sezgin, S. (2009). Cigarette smoking status among students of primary school in Erzurum. *Smokıng and Health, 10*, 75-81.

Valcke, M., De Wever, B., Van Keer, H., & Schellens, T. (2011). Long-term study of safe internet use of young children. *Computers&Edcucation*, 57, 1292-1305. doi:10.1016/j.compedu.2011.01.010

Walker, E. R., Engelhard, G., & Thompson, N. J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure, 21*(6), 437−443. doi:10.1016/j.seizure.2012.04.009

Yazıcı, H., & Şahin, M. (2006). İlköğretim öğrencilerinin sigara içen ve içmeyenleri algılama biçimleri *[The ways primary school students perceive smokers and non-smokers.]*. *Marmara University Atatürk Faculty of Education Journal of Educational Sciences*, (21), 213-224.

Yılmaz, F., & Ersoy, A. (2016). Class teacher candidates skill of saying no in relation to components of moral anatomy. *Eurasian Journal of Educational Research, 16*(62), 161-180. doi:10.14689/ejer.2016.62.10

Yolcu, S. (2007). *Evaluation of homework web sites in terms of education.* (*Unpublished master's thesis).* University of Dokuz Eylül, İzmir.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 461–473

DOI: 10.21449/ijate.442406

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Construction and Validation of Physics Aptitude Test as an Assessment Tool for Senior Secondary School Students

Patrick Uzo Osadebe[1,*], **Chukuka Prince Nwabeze**[1]

[1] Department of Guidance and Counselling, Delta State University, Abraka, Nigeria

Abstract: The purpose of the study was to construct and validate a Physics Aptitude Test (PAT) for senior secondary schools. The test was designed primarily as an assessment tool for senior secondary school students who have aptitude for Physics. Four research questions guided the study. It was an instrumentation study designed to produce an assessment tool for senior secondary school students. A sample of 200 students was randomly selected through stratified random sampling technique. The data collected were analysed to determine the validity of the test, item validity through item analysis and reliability of the Physics Aptitude Test. The findings revealed that the Physics Aptitude Test (PAT) has adequate face and content validities. It is made up of 50 items. The difficulty and Discrimination indices were appropriate because they are within the standard range of indices for the test. Difficulty indices range from 0.30 to 0.70, Discrimination indices range from .30 to 0.44. A reliability coefficient of 0.94 was obtained through Kuder Richardson formula 20 as a measure of internal consistency. Since the Physics Aptitude Test (PAT) was highly valid and reliable, it was recommended that it should always be used as an assessment tool for determining the aptitude of senior secondary school students in Physics.

ARTICLE HISTORY

Received: 26 March 2018
Revised: 13 May 2018
Accepted: 29 May 2018

KEYWORDS

Construction,
Validation,
Physics,
Aptitude Test

1. INTRODUCTION

Education in Nigeria, before the coming of Western Education was mostly oral and informal. Assessment of the informal or traditional education was mainly oral. Written test or formal examination had no place in the evaluation (Ohuche & Akeju, 1988). The advent of Western education in Nigeria gave birth to formalized mode of testing and examination in the educational system. These examinations are in two forms, internal and external examinations. The internal examinations are test items that are designed and administered within the school such as terminal examinations, promotion examinations and so on, while the external examinations are those examinations conducted by public examination bodies which are external to the students for the purpose of certification, placement, and job opportunities. Some of these bodies include the West African Examination Council (WEAC), State Ministry of

CONTACT: Patrick. U. Osadebe ✉ drosadebeuzo@gmail.com Department of Guidance and Counselling, Delta State University Abraka, Nigeria

ISSN-e: 2148-7456 /© IJATE 2018

Education, Joint Admission and Matriculations Board (JAMB), National Examination Council of Nigeria (NECO).

According to Egbule (2002), the term test connotes a common set of tasks or a presentation of a standard set of questions to be answered. Based on the answers provided to the series of questions we obtain a measure that is subject to evaluation. Therefore, a test can be defined as any kind of procedure or device for measuring aptitude, interest, ability, achievement and any other traits or personal attributes.

A test as a set of questions, problems, puzzles, symbols and exercise used to determine a person`s ability, aptitude, knowledge qualification, interest and level of social adjustment. Thus, when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it and you cannot express it in numbers, your knowledge about it is of a meagre and an unsatisfactory kind.

Similarly, a test is a sample measure of behaviour from which inferences may be made to a universe of behaviour. This implies that the purpose of most tests is to measure a sample of behaviour. To attempt to measure the entire behaviour at a time may be difficult and farfetched. It is, therefore, only reasonable to measure and obtain a numerical estimate of a single aspect of performance at a time such as language, mathematics, specific aptitude, artistic performance and so on. Therefore, test enables us to measure and understand a concept, sample behaviour or personal attributes. It serves the purpose of selecting students into educational programmes, classification of students, and certification after graduation from one level of education to another, prediction of future performance or potential, verification and evaluation of academic programmes as well as provision of significant data and information for counselling and diagnosis.

1.1. Review of Literature

Test construction simply implies a systematic process of assembling test items or the preparation of a test by drawing and compiling series of questions which constitute the task for the students. Test Validation on the other hand, is defined by Gronlund (1976) as a procedure for standardising test items by treating them statistically to remove all source of bias in the process of making them valid, reliable, objective and usable.

Itsuokor (1995) suggested an outline for an effective test construction. He noted that a satisfactory sample is most likely to be obtained when test preparation follows a systematic procedure. He listed the following steps as being useful for this purpose.

a) General statement of instructional objectives;
b) Making an outline of the content to be covered;
c) Preparation of a table of a specification and
d) Constructing test items that measure the objectives in the specified table.

Similarly, Okobia (1990) stated that in constructing and validating a test, certain procedures are followed. As such most measurement and evaluation experts have come to agree that test preparation is a procedural and a systematic process. However, construction and validation strategies may vary from one author to the other depending on the nature of the test. Brown (1983) reports that Educational Testing Services (ETS) recommended typical sequences of construction and validation which is stated as follows:

1. Planning the test
2. Writing the items
3. Pretesting the items
4. Preparing the form
5. Collecting the reliability and validity evidence
6. Developing normative and interpretative materials.

However, as noted by Osadebe (2001), principles of test construction are the guidelines for developing a test. Test construction is a special area for differential psychologists, test and measurement experts. Test is a valid and reliable instrument. Teachers should be guided in constructing items/test following the principles stated below;

1. Planning
2. Item writing
3. Item analysis-trial testing
4. Composition of item
5. Reliability
6. Printing and Administration
7. Marking and scoring (measurement)
8. Manual

Therefore, the physics teacher, having undergone a course of training in measurement and evaluation should be able to construct, validate, select, administer and score a test as well as interpret the scores generated from tests. There is a notable short fall of test material standardized or normed in Nigerian culture for use in educational services. This situation creates a problem in educational development of the students, for example in Nigeria; it is unfortunate that many of the inappropriate vocation choices are as a result of ignorance and illiteracy on the parts of parents. They have fixed ideas of the course of study, which they would want their children to study, and also the interest of their children in the said course of study (Zhao, 2006).These led students into choosing courses that are either too low or above their mental capacity.

It should be noted that students' low intellectual ability, poor aptitude disposition coupled with low level of aspiration for particular courses like physics have a devastating implication on the students. These problems are as a result of absence of requisite test materials in physics and also in educational and vocational counselling. It should be noted clearly that many failures recorded in the sciences today especially in physics is as a result of wrong selection of courses to be studied.

Therefore, the need for proper matching of aptitudinal disposition of the student is what prompted this study on the construction and validation of Physics aptitude test. Based on this, it is important to emphasize the meaning and relevance of aptitude test in physics. Aptitude generally implies an innate potential of special ability that is capable of predicting future performance (Bannatyne, 1974).

According to Cronbach (1970), aptitude test is a proficiency test that measures ability to perform some task significant in its right. An aptitude test in this regard, is intended to predict success in some occupation or course of training. Cronbach (1970) reported that there are many types of aptitude, musical aptitude, reading aptitude, clerical aptitude, spatial aptitude and numerical aptitude.

Aptitude testing according to Itsuokor (1995), embraces intelligence and achievement, personality, interest and other skills which predispose one to learning. On this basis, he sees aptitude as an ability to learn, and aptitude test assess knowledge, skills and other characteristics that serve to predict learning and success. According to him, aptitude tests cover wide range of human behaviour among which is mechanical aptitude, clerical aptitude, musical aptitude and artistical aptitude. An example of aptitude test commonly used for school educational and vocational counselling is the Differential Aptitude Tests (DAT) which comprises eight distinct sections – verbal reasoning, space – relations including visualization and perception, clerical spread and accuracy, mechanical reasoning, language, spelling and usage (Berk, 1982).Aptitude test is used to investigate not only what a student can do but also what he does. Therefore, aptitude test is used to forecast success in future assignment or occupation. Physics is a physical

science subject that is concerned mainly with matter as it relates to energy. It has many applications in other fields, for example in chemistry, biology, pharmacy, engineering and other related fields or disciplines. It should be noted that physics is one of the subjects on which modern technology is based. There is the need therefore to lay a solid foundation in the subject by way of developing instrument that can be used to predict or forecast one's success on the subject in future. However, the major reason for this study is to make the students ready for their future. This will be achieved by proper planning and adequate making, and choosing the right so as to properly adjust, or ignore a job. Therefore, the need for proper career planning at entrance into the senior secondary school becomes very important.

Agbola (1990) carried out a study on construction and validation of Mathematics Achievement Test for JSS3 students in Ondo State. He used 50 items in his final test. He involved both rural and urban schools. His calculated t – value was 7.74 which were greater than the 1.96, the critical value at 0.05 level of significance. He found out that there was a significant difference between the scores obtained by the students located in the urban and rural areas in the Ondo state JSC Examinations. Students in the rural areas perform better than the students in the urban areas.

Irighweferhe (2008) carried out a study on construction, validation, and standardization of mathematics achievement test for senior secondary school students in Delta State of Nigeria. The items for the test were drawn from the National curriculum for senior secondary school Mathematics. The items covered the following areas, number and Numeration, Algebraic processes, Plan Geometry, Trigonometry, menstruation and statistics and probability. The constructed test was administered to a sample of 1000 SS3 students in the 2007/2008 academic session selected through stratified random sampling. The result obtained from student's performance was norm using standard score of Z-scores and T-scores. The norm profile of performance of student across gender, location, subject combination and school type indicated that there was a relationship performance of student among the above mention variables. It was therefore recommended that the standardized mathematics achievement test should be used in preparing and assessing the performance of students preparing for the senior secondary school certificate examination and at same time gaining insight into their potentials in mathematics. Test manuals for the Standardized Mathematics Achievement Test (SMAT) was also prepared.

Akaezue (2009) carried out a study on construction and validation physics aptitude test in senior secondary school in Sapele Local Government Area of Delta State. To achieve this purpose four research question were drawn to guide the study. It is an instrumentation study. The instrument developed is called physics aptitude test. A sample of hundred students was used to determine the item difficulty, discriminatory index validity and reliability of the physics aptitude test. The test has content validity; a reliability of 0.75 was obtained through the use of Kuder – Richardson formula 20. The test has a standard error of measurement of 3.15.The difficulty level of the test ranges from 0.325 to 0.855 while the discriminatory level ranges from 0.15 to 0.48. The researcher found out that the test has high content validity, the items of the test are suitable in terms of its difficulty and discriminatory indices and also the test and its entire subsections have high degree of internal consistency with a low standard error of measurement.

The review carried out in this study indicated that some studies have been done in test construction and validation in many subjects such as Chemistry, Biology, Geography, Economics, Mathematics and Integrated Science but not much, known to the researchers, has been done in Physics in Delta State in recent times. This is the gap the study covered.

1.2. Purpose of the Study

The main purpose of the study was to construct and validate Physics Aptitude Test (PAT). Specifically, the study determined the following:

1. Validity of the Physics Aptitude Test
2. Difficulty index of each item of the Physics Aptitude Test
3. Discrimination index of each item of the Physics Aptitude Test
4. Reliability of the Physics Aptitude Test

1.3. Research Questions

The study was guided by the following research questions:

1. What is the validity of Physics Aptitude Test?
2. What is the difficulty index of each item of the Physics Aptitude Test?
3. What is the discrimination index of each item of the Physics Aptitude Test?
4. What is the reliability of Physics Aptitude Test?

2. METHOD

The design of this study is instrumentation. It involves the construction and validation of Physics Aptitude Test (PAT). It is designed for selecting senior secondary school students who have ability for Physics.

The population for this study comprises public Senior Secondary 1(SSI) students in 2014/2015 academic session in Delta North Senatorial District of Delta State. The population consists of 5,760 SSI students from Public Secondary Schools in Delta North Senatorial District of Delta State (Ministry of Education, Delta State. 2014).

From the population, a total of 200 students in SS1 were sampled for the study in Delta North Senatorial District of Delta State. The sampling technique adopted for this study is the stratified sampling and simple random sampling technique.

According to Nworgu (2003), stratified sampling ensures greater representativeness of the sample relative the population and guarantees that minority constituents of the population are represented in the sample. The sample characteristics therefore are better approximation (estimates) of the population characteristics and the associated sampling error is correspondingly reduced. Elements are drawn randomly from each in such a way that the relative proportions of the strata in the resultant sample are the same as in the parent population. The relative contribution of each stratum in the population is exactly its relative contribution in the sample. The sample, therefore, will possess specified characteristics in exactly the same proportion as those characteristics exist in the parent population.

Random sampling is therefore the selection of sample subjects by chance from population and is therefore not influenced by other factors or considerations. The simple random sampling technique ensures that every object, event or individuals in the population has a chance of being drawn for the sample.

The instrument for the research is Physics Aptitude Test (PAT). The instrument is made up of items drawn from SS1 Physics. It was generated in such a way that the cognitive aspects of the students' behaviour were included, and various aspects/component of aptitude test were reflected such as numerical aptitude which is in the first section and it has a total of ten(10) questions, followed by the verbal section which also has a total of ten (10) questions. Next to the verbal aptitude section is the quantitative aptitude section which has a total of 15 questions. The last section of this test is the mechanical aptitude test section which also has a total of fifteen (15) questions.

Summarily, 100 questions were constructed but fifty (50) questions were eliminated from the generated 100 items after item analysis; that is the difficulty level and item discriminative index. After the item analysis and modification of the items, fifty (50) items were selected for the final test. These Fifty (50) questions formed the various sections of the aptitude test. Each item has four (4) options lettered; A, B, C and D. One out of the four (4) options is correct while the other three are treated as distracters.

Validity is one of the most essential psychometric properties of a test. The validity of a test is the extent to which a test measures what it is supposed to measure. It refers to the extent to which the results of an evaluation procedure serve the particular uses for which they are intended (Gronlund, 1976). In order to establish the content validity of the test items, a table of specifications was drawn. This shows the various types of aptitude test considered, and the total number of generated test items. In addition, two specialists in the field of physics and two other specialists in measurement and evaluation were used. The specialists carefully examined the generated test items, made useful suggestions and approved the try out test.

The Kuder- Richardson formula 20 (KR20) was used in establishing a measure of internal consistency of the test. The test was administered to 30 students that were randomly selected from the population, for the purpose of the reliability. A reliability index of 0.94 was obtained from the computation.

To collect the data for the research, the generated test items (Physics Aptitude Test question) was administered to 200 students for the purpose of item analysis. This was done with the assistance of the teachers and cooperation of the management of the selected schools. The instrument was timed and a total of one hour was allowed for the test. The students were stopped at the stipulated time. At the end of the test the researcher collected the students' response sheet. The test items were administered on conducive environment devoid of any form of examination malpractice.

In scoring the test, each item that was correctly answered attracted one (1) mark or one (1) point while wrongly answered items attracted zero (0) mark. The test rates the performance of the testees over hundred (100) points or marks. This was used to select students who are capable of studying physics. It should be noted here that for ease of computation, a multiplying factor of 2 was used to multiply the scores at the end of scoring the testees.

However, the item difficulty and discriminating indices were calculated for the 100 items to select 50 items. The students' sheets were re-arranged from the highest to the lowest. The upper group of 27% and lower group of 27% were selected and used for the item analysis.

Items with difficulty indices ranging from 0.30 to 0.70 were included (accepted) for the fina #+l test while others below 0.30 and above 0.70 were discarded (rejected) as a result of being too difficult or too easy respectively. The item discrimination index (D) ranged from 0.30 to 0.44 and this formed the bench mark for accepting items regarding discrimination indices for the final test.

3. RESULTS

Research Question One: What is the validity of Physics Aptitude Test?

In order to provide answer to the first research question, a table of specifications was constructed. The test blue print was drawn to ascertain the extent of the content validity of the Physics Aptitude Test (PAT). The table is presented below.

Table 1. A Table of Specifications evaluating the extent of the Content Validity of the 50 items of PAT.

	COGNITIVE LEVEL (BEHAVIOURAL OBJECTIVE)						
Content Area	Know 10%	Comp 10%	Application 20%	Analysis 20%	Synthesis 20%	Evaluation 20%	Total 100%
Numerical Aptitude 20%	1	1	2	2	2	2	10
Verbal Aptitude 20%	1	1	2	2	2	2	10
Quantitative Aptitude 30%	1	2	3	3	3	3	15
Mechanical Aptitude 30%	2	1	3	3	3	3	15
Total	5	5	10	10	10	10	50

As seen on Table 1, the table of specifications clearly reflects the various content areas in Physics that were considered in this study which helped to establish a high content validity for the Physics Aptitude Test. Therefore, the Physics Aptitude Test has a high content validity because there was a wide content coverage.

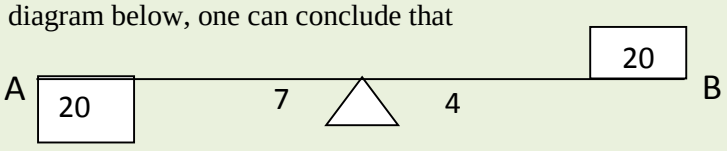
In addition to the table of specifications, the researchers also presented the test items to other experience Physics teachers, project supervisor and two Measurement and Evaluation Experts who also established the correctness, adequateness and appropriateness of the items in the constructed test. The face validity of the test was also established. It describes the appearance of the test. It refers to how real the items of a test are to test takers; the cosmetics to the format for presenting or reporting the test items, the typing and general outlook of the test.

Research Question 2: What is the difficulty index of each item of the Physics Aptitude Test?

The test constructors used the difficulty index to select the suitable items that are appropriate to be included in the final test. These items were selected from the initial items generated for the trial testing. Items with index of 0.30 – 0.70 were selected for the difficulty level

Table 2. Items and their Difficulty indices (P)

S/N	*ITEMS (NUMERICAL APTITUDE TEST)*	*P*
1.	A boy of mass 50kg runs up a set of steps of total height 3.0m.Find the work done against gravity.	*0.55*
2.	A car travels at an average speed of 100kmh^{-1}, what distance does it cover in 5minutes?	*0.54*
3.	An object of mass 5kg is moving at a constant velocity of 15ms^{-1}, calculate its kinetic energy.	*0.60*
4.	A boy of mass 10kg climbs up 10 steps each of height 0.2m in 20 seconds. Calculate the power of the body.	*0.61*
5.	Three resistor of resistance 1 Ω, 2 Ω and 4 Ω are connected in series find the equivalent resistance of the combination.	*0.63*
6.	Find the resistance of a wire of length 0.65m, radius 0.2mm and resistivity 3x10^{-6}ohm-metre.	*0.59*

7.	A force of 0.8N stretches an elastic spring by 2cm. Find the elastic constant of the spring.	*0.62*
8.	A spring is stretches 40mm by a force 15N what is the work done by the force?	*0.55*
9.	A change of 10 coulombs flows round a circuit in 2seconds.what is the current?	*0.59*
10.	A boy moves a load of 10N at a distance of 5metres.calculate the w0sork done.	*0.63*
	ITEMS (VERBAL APTITUDE TEST)	
11.	Which of the following is a fundamental unit?	*0.65*
12.	Which of the units of the following physical quantities are not derived?	*0.62*
13.	What is the best instrument for measuring the diameter of a meter rod?	*0.47*
14.	Which among the following substance is a non- conductor?	*0.59*
15.	Which of the following is not a consequence of a force field?	*0.61*
16.	The slope of a straight line displacement –time graph indicates	*0.52*
17.	Under which of the following conditions is work done	*0.60*
18.	A room is heated by a means of a charcoal fire. A man standing away from the fire is warmed by	*0.62*
19.	On which of the following is the design of the thermostat of an electric iron based?	*0.53*
20.	Which of the following surface will absorb radiant heat energy best?	*0.65*
	ITEMS (QUANTITATIVE APTITUDE TEST)	
21.	An effort of 10N was used in lifting a load of 500N.calculate the mechanical advantage used in lifting the load	*0.49*
22.	A load of 10N was pushed through a height of 2meters.calculate the work done in pushing the load.	*0.55*
23.	A man lifted a bag of cement of mass 50kg through a height of 2meters.If the acceleration due to gravity is $10ms^{-2}$.Calculate the energy used in lifting the load	*0.57*
24.	Express 4000J in kilojoules	*0.53*
25.	3 mega joules is equivalent to ------ joule?	*0.67*
26.	How many minutes do we have in 2hours?	*0.62*
27.	The speed of a bus increases uniformly from $15ms^{-1}$ to $60ms^{-1}$ in 20 seconds. Calculate the average speed.	*0.54*
28.	A man of mass 50kg ascends a flight of stairs 5m high in 5 seconds. If acceleration due to gravity is $10ms^{-2}$, the power expended is?	*0.59*
29.	A body of mass 10kg accelerate at $2ms^{-2}$ what is the force in Newton (take; F=ma)	*0.63*
30.	Two bodies x and y, undergo a distance of 30.25km and 50.75km respectively. What is the total distance made by x and y	*0.57*
31.	Convert 50cm to metres	*0.58*
32.	A metre rule was measured to be 100.02cm.What is the error in the measurement?	*0.49*
33.	A body has a mass 40kg and velocity 5m/s, calculate the energy possessed by this body?	*0.63*
34.	A girl lift a load of 2kg through a height 10m.what is the work done against gravity	*0.61*
35.	A machine has a mechanical advantage 0.2 and load of 5000N.Calculate the effort of the machine	*0.59*
	ITEMS (MECHANICAL APTITUDE TEST)	
36.	A machine has an efficiency of 60%.If the machine is required to overcome a load of 30N with a force of 20N, calculate its velocity ratio	*0.61*
37.	Which among the following is not a consequence and application of expansion?	*0.53*
38.	The bimetallic strip has its application on the following appliance except	*0.50*
39.	A screw that is rotated forward through one revolution will move a distance equal to	*0.58*
40.	From the diagram below, one can conclude that 	*0.61*
41.	On which of the following is the design of the thermostat of an electric iron based?	*0.56*

42.	If a load of 5000N could be lifted up by an effort of 100N, what is the mechanical advantage of the machine?	*0.53*
43.	A building can be adequately protected from lighting by ……	*0.55*
44.	A rocket is burning fuel at the rate of $200gs^{-1}$ and ejecting all the gas in one direction at the rate of $400ms^{-1}$.what is the maximum weight the rocket can have if it is going to move vertically upwards?	*0.59*
45.	Efficiency of a machine is defined by ……?	*0.57*
46.	Which principle does the lever operates?	*0.53*
47.	A machine has a velocity ratio 5 and is 80% efficient. What effort would be needed to lift a load of 200N with the aid of this machine?	*0.50*
48.	The velocity ratio of inclined plane is given by	*0.54*
49.	Which of the following is not an example of levers of the first order?	*0.57*
50.	The ideal method of reducing friction in a machine …….?	*0.55*

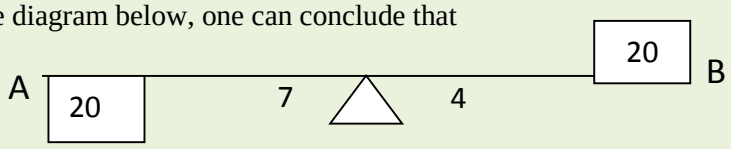
Table 2 shows the difficulty indices of 50 items for the various component of the Physics aptitude test. The acceptable indices ranged from 0.30 to 0.70. The Numerical Physics Aptitude Test items ranged from 0.54 to 0.63, Verbal Physics Aptitude test items ranged from 0.47 to 0.65, Quantitative Aptitude test items ranged from 0.49 to 0.67 and Mechanical Physics Aptitude test items ranged from 0.50 to 0.61. The indices were established during item analysis which helped to ensure high item validity for each Physics test item.

Research Question Three: What is the discrimination index of each item of the Physics Aptitude Test?

This research question was answered using the item analysis. The discriminating indices help to refine test items. The indices help to measure the extent to which items discriminate between high and low achievers (students).

Table 3. Items and their Discriminatory level (D)

S/N	***ITEMS (NUMERICAL APTITUDE TEST)***	***D***
1.	A boy of mass 50kg runs up a set of steps of total height 3.0m.Find the work done against gravity.	*0.41*
2.	A car travels at an average speed of $100kmh^{-1}$, what distance does itCover in 5minutes?	*0.44*
3.	An object of mass 5kg is moving at a constant velocity of $15ms^{-1}$, calculate its kinetic energy.	*0.41*
4.	A boy of mass 10kg climbs up 10 steps each of height 0.2m in 20 seconds. Calculate the power of the body.	*0.33*
5.	Three resistor of resistance 1 Ω, 2 Ω and 4 Ω are connected in series find the equivalent resistance of the combination.	*0.41*
6.	Find the resistance of a wire of length 0.65m, radius 0.2mm and resistivity $3x10^{-6}$ ohm-metre.	*0.44*
7.	A force of 0.8N stretches an elastic spring by 2cm. Find the elastic constant of the spring.	*0.35*
8.	A spring is stretches 40mm by a force 15N what is the work done by the force?	*0.32*
9.	A change of 10 coulombs flows round a circuit in 2seconds.what is the current?	*0.37*
10.	A boy moves a load of 10N at a distance of 5metres.calculate the work done.	*0.33*
	ITEMS (VERBAL APTITUDE TEST)	
11.	Which of the following is a fundamental unit?	*0.30*
12.	Which of the units of the following physical quantities are not derived?	*0.39*
13.	What is the best instrument for measuring the diameter of a meter rod?	*0.35*
14.	Which among the following substance is a non- conductor?	*0.37*
15.	Which of the following is not a consequence of a force field?	*0.41*
16.	The slope of a straight line displacement –time graph indicates	*0.33*

17.	Under which of the following conditions is work done	*0.32*
18.	A room is heated by a means of a charcoal fire. A man standing away from the fire is warmed by	*0.35*
19.	On which of the following is the design of the thermostat of an electric iron based?	*0.39*
20.	Which of the following surface will absorb radiant heat energy best?	*0.37*
	ITEMS (QUANTITATIVE APTITUDE TEST)	
21.	An effort of 10N was used in lifting a load of 500N.calculate the mechanical advantage used in lifting the load	*0.37*
22.	A load of 10N was pushed through a height of 2meters.calculate the work done in pushing the load.	*0.39*
23.	A man lifted a bag of cement of mass 50kg through a height of 2meters.If the acceleration due to gravity is $10ms^{-2}$.Calculate the energy used in lifting the load	*0.33*
24.	Express 4000J in kilo joule	*0.39*
25.	3 mega joules is equivalent to ------ joule?	*0.41*
26.	How many minutes do we have in 2hours?	*0.43*
27.	The speed of a bus increases uniformly from $15ms^{-1}$ to $60ms^{-1}$ in 20 secoSnds. Calculate the average speed.	*0.33*
28.	A man of mass 50kg ascends a flight of stairs 5m high in 5 seconds. If acceleration due to gravity is $10ms^{-2}$, the power expended is?	*0.37*
29.	A body of mass 10kg accelerate at $2ms^{-2}$ what is the force in Newton (take; F=ma)	*0.41*
30.	Two bodies x and y, undergo a distance of 30.25km and 50.75km respectively. What is the total distance made by x and y	*0.33*
31.	Convert 50cm to meters	*0.39*
32.	A metre rule was measured to be 100.02cm.What is the error in the measurement?	*0.35*
33.	A body has a mass 40kg and velocity 5m/s, calculate the energy possessed by this body?	*0.33*
34.	A girl lift a load of 2kg through a height 10m.what is the work done against gravity	*0.37*
35.	A machine has a mechanical advantage 0.2 and load of 5000N.Calculate the effort of the machine	*0.44*
	ITEMS (MECHANICAL APTITUDE TEST)	
36.	A machine has an efficiency of 60%.If the machine is required to overcome a load of 30N with a force of 20N, calculate its velocity ratio	*0.41*
37.	Which among the following is not consequence and application of expansion?	*0.39*
38.	The bimetallic strip has its application on the following appliance except	*0.37*
39.	A screw that is rotated forward through one revolution will move a distance equal to	*0.35*
40.	From the diagram below, one can conclude that 	*0.30*
41.	On which of the following is the design of the thermostat of an electric iron based?	*0.41*
42.	If a load of 5000N could be lifted up by an effort of 100N, what is the mechanical advantage of the machine?	*0.35*
43.	A building can be adequately protected from lighting by ……	*0.32*
44.	A rocket is burning fuel at the rate of $200gs^{-1}$ and ejecting all the gas in one direction at the rate of $400ms^{-1}$.what is the maximum weight the rocket can have if it is going to move vertically upwards?	*0.41*
45.	Efficiency of a machine is defined by ……?	*0.33*
46.	Which principle does the lever operates?	*0.39*
47.	A machine has a velocity ratio 5 and is 80% efficient. What effort would be needed to lift a load of 200N with the aid of this machine?	*0.41*
48.	The velocity ratio of inclined plane is given by	*0.37*
49.	Which of the following is not an example of levers of the first order?	*0.33*

50.	The ideal method of reducing friction in a machine …….?	*0.32*

Table 3 indicates the discrimination indices of 50 items for the various components or characteristics of Physics aptitude Test. The acceptable indices during item analysis ranged from 0.30 to 0.44. The discrimination indices of the Numerical Aptitude test item ranged from 0.32 to 0.44, Verbal Aptitude 0.30 to 0.41, Quantitative Aptitude 0.33 to 0.44, and Mechanical Aptitude 0.30 to 0.41. The item indices were accepted and appropriate for the Physic Aptitude Test (PAT).

Research Question Four: What is the reliability of Physics Aptitude Test?

To give answer to research question four, the reliability of the PAT was estimated using Kuder Richardson formula 20. The formula is stated below:

Table 4. Analysis of Reliability of Physics Aptitude Test.

No of students	*No of items*	*Σpq*	$\bar{X}$	*SD*	*SD2*	*r*	*r2*	*r %*	*SEM*	*Decision at .05*
30	*50*	*9.1801*	*31*	*3.2*	*9.933*	*0.94*	*0.8836*	*94*	*0.78*	*High*

As can be seen from Table 4, an estimate of the PAT reliability using Kuder Richardson formula 20 was established. This approach becomes pertinent because the PAT is a multiple choice objective test with expected response of either pass (1) or fail (0). However, a reliability index of 0.94 was obtained. The result is significant at .05 level of significance which made the instrument (PAT) very suitable for the study. A standard error of measurement (SEM) of 0.78 was obtained. The SEM is low and this confirms that PAT is highly reliable.

4. DISCUSSION

The discussion is based on the main findings after answering the research questions. The instrument was validated. Validity is one of the pertinent psychometric properties of an instrument. It refers to the extent to which an instrument measures what it is designed to measure. In establishing the content validity of the instrument, two approaches were adopted. First the use of table of specification was employed. This approach is similar to Osadebe (2001), Irighweferhe (2008), Akpoguma (2008), Akaezue (2009) and Osiobe (2012). The second approach adopted was the use of experts' judgement. The items were presented to experienced physics teachers and measurement and evaluation experts. This provide for the correctness, adequateness and appropriateness of the test. To establish the reliability of the Physics Aptitude Test, Kuder – Richard formula 20 was employed. The use of Kuder – Richard formula 20 was as a result of the fact that the Physics Aptitude Test is a multiple choice objective test with expected response of either pass (1) or fail (0). A reliability coefficient of 0.94 was obtained at 0.05 unlike Oloya (2005) and Onoyumolo (2005), who used split – half method in establishing their reliability.

Furthermore, Irighweferhe (2008) agreed that reliability coefficient of 0.69 is high and adequate. Akazue (2009) in his study, reported reliability coefficient of 0.75 which he judged to be significant for a test. This study has found out that the new instrument (PAT) has a higher reliability of 0.94 which is higher than the above reported ones. The instrument yielded very high internal consistency of scores.

Classical test theory was used in the construction of the Physics Aptitude Test (PAT). This approach is similar to Egbule (1998), Osadebe (2001), Irighweferhe (2008), Akpogumu (2008), Akaezue (2009) and Osiobe (2012). In addition to the classical test theory, a conceptual model was also designed to enhance the quality of the instrument.

The items that made up the instrument (PAT) were selected through the item analysis. Their difficulty indices and discriminating indices were computed. In terms of difficulty indices, experts in measurement and evaluation such as Nworgu (2003) reported that an ideal item should have facility index of 0.5 but in real life situation it will range from 0.30 to 0.70. All items in the instrument are within the range of 0.30 and 0.70 making them very appropriate, suitable and effective. This is similar to Akpoguma (2008) and Osiobe (2012).

The discriminating indices that measure the extent items discriminate between the bright and dull students were also computed. Discriminating index of an item varies from 0.00 to 0.01. Negative indices are abnormal because they penalised more of the bright students than the dull students; hence they were rejected. Nworgu (2003) agreed that an ideal item should possess discriminating indices of +1 but realistically it should range from + 0.03 to 1.00. In order to include only high-quality items, the researcher used a realistic range of discriminating indices from 0.30 to 1.00 to select the items included in the instrument. This is similar to Akpoguma (2008) and Osiobe (2012).

Conclusively, the Physics Aptitude Test developed by the researchers is a test with high psychometric properties. As such, the test could be used for the selection of secondary school students who have the desire to study physics in their senior secondary schools as well as an assessment tool for the evaluation of learning outcomes.

The items of the test are suitable and appropriate in terms of difficulty and discrimination indices. The test has a high degree of internal consistency with a low standard error of measurement.

ORCID

Patrick Osadebe https://orcid.org/0000-0002-6365-4285

5. REFERENCES

Agbola, F. (1990). *Construction and Validation of Mathematics Achievement Test for J.S.S 3 Students*. Unpublished M. Ed. Dissertation, University of Benin, Benin City, Nigeria.

Akaezue, N.N. (2009). *Construction and Validation of Physics Aptitude Test for Senior Secondary School Student in Delta State*. Unpublished M.Ed. project, Delta State University, Abraka, Delta State.

Akpoguma, J. (2008). *Construction and Validation of Mathematics Achievement Test for Senior Secondary School Students in Delta State*. Unpublished M.Ed. Project Delta State University, Abraka.

Bannatyne, A. (1974). Diagnosis: A Note on the Recategorization of the Wisc Scaled Scores. *Journal of Learning Disorders, 7*, 272–273.

Berk, R.A. (1982). *Handbook for Detecting Test Bias*. Baltimore: John Hopkins University Press.

Brown, F.G. (1983). Principles of Education and Psychological Testing (3rd Edition). Oxford University Press.

Cronbach, L. J. (1970). *Essential of Psychological Testing*: New York: Steal Press.

Educational Testing Service (1992). *The Origin of Educational Testing Service*. Princeton, NJ: ETS. En.wikipedia.org/wiki/Educational-Testing-Service.

Egbule, J. F. (2002). *Principles of Psychological Testing*. Owerri: Whyte and Whyte Publishers.

Egbule, J.F. (1998). *Construction and Validation of Differential Aptitude Test (DAT) for Educational and Vocational Counselling*. An Unpublished Ph. D thesis, Delta State University Abraka.

Gronlund, E. N. (1976). *Measurement and Evaluation in Psychology and Education*. New York: Macmillan Publishing Company.

Irighweferhe, S. U. (2008). *Construction, Validation and Standardization of Mathematics Achievement Test for Senior Sate Secondary School Student in Delta State of Nigeria*. An Unpublished M.Ed project, Delta State University Abraka, Abraka, Delta State.

Itsuokor, D. E. (1995). *Essentials of Test and Measurement*. Illorin: Woye and Sons (Nig) Ltd.

Nworgu, B.G. (2003). *Educational Measurement and Evaluation*, Theory and Practices. Nsukka; University Trust Publishers.

Ohuche, R.O. &Akeju, S.A. (1988). *Continuous Assessment for every Learner*. Onitsha: Africana FEP publishers.

Okobia, D.O. (1990). *Construction and Validation of Social Studies Achievement Test (SSAT) for SS Students in Selected Secondary Schools in Bendel State*. An unpublished M. Ed project, University of Benin, Benin City, Nigeria.

Onoyumolo, L.A. (2005). *Construction, Validation of achievement test for Physics for SSS Student in Warri South L.G.A. of Delta State*. An unpublished M. Ed project Delta State University, Abraka.

Osadebe, P.U. (2001). *Construction and Standardization of an Economics achievement test for senior secondary school students*. Unpublished Doctoral thesis, University of Port Harcourt.

Osiobe, G.A. (2012). *Construction and Standardization of Geography Objective Test for Senior Secondary Schools in Delta State*. Unpublished Ph.D thesis, Delta State University Abraka.

Zhao, Y.Y. (2006). Motivation in Education. *International Journal of Engineering Education, 22*, 1281–1286.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 474–490

DOI: 10.21449/ijate.435507

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Predicting Achievement with Artificial Neural Networks: The Case of Anadolu University Open Education System

Hilal Seda Yıldız Aybek[1, *], **Muhammet Recep Okur**[2]

[1]Center for Open and Distance Learning, Yasar University, Turkey
[2]Faculty of Open Education, Anadolu University, Turkey

Abstract: This study aims to predict the final exam scores and pass/fail rates of the students taking the Basic Information Technologies – 1 (BIL101U) course in 2014-2015 and 2015-2016 academic years in the Open Education System of Anadolu University, through Artificial Neural Networks (ANN). In this research, data about the demographics, educational background, BIL101U course mid-term, final and success scores of 626,478 students was collected and purged. Data of 195,584 students, obtained after this process was analysed through Multilayer Perception (MLP) and Radial Basis Function (RBF) models. Sixteen different networks attained through the combination of ANN parameters were used to predict the final exam scores and pass/fail rates of the students. As a result of the analyses, it was found out that networks established through MLPs make more exact predictions. In the prediction of the final exam scores, it was determined that there is a low level of correlation between the actual scores and predicted scores. In the analyses for the prediction of pass/fail rates of the students, networks established through MLPs ensured more exact prediction results. Moreover, it was determined that the variables as mid-term exam scores, university entrance scores and secondary school graduation year were of highest importance in explaining the final exam scores and pass/fail rates of the students. It was found out that in the higher institutions serving for Open and Distance Learning, pass/fail state of the students can be predicted through ANN under favour of variables of students which have been found as most the important predictors.

ARTICLE HISTORY

Received: 05 February 2018
Revised: 13 May 2018
Accepted: 29 May 2018

KEYWORDS

Prediction of Student Achievement,
Achievement in the Higher Education,
Open and Distance Learning,
Artificial Neural Networks

1. INTRODUCTION

Considering the great number of students registered in the Open Education System (OES) of Anadolu University, which is one of the mega universities of the world, it is necessary to analyse diverse features of this group of students in detail, with the aim of ensuring various student support services effectively and efficiently (da Silva, de Fátima Nunes, Santos, Queiroz & Leles, 2012; Kose & Arslan, 2017). Particularly, identification and presentation of the variables explaining the achievement levels of the students may provide key information to the

CONTACT: Hilal Seda YILDIZ-AYBEK ✉ hilal.aybek@yasar.edu.tr Center for Open and Distance Learning, Yasar University

ISSN-e: 2148-7456 /© IJATE 2018

institutions for the planned student support services. Identification of the variables explaining the achievement level of the present students may help to make predictions about the future achievement of the students. Machine learning algorithms such as Artificial Neural Networks (ANN) and various statistical models are used to predict the achievement of the student. It is considered that prediction of student's achievement in advance may be beneficial in various terms for both higher education institutions and students.

It has been observed in the related literature that there is a limited number of studies on the variables explaining the academic achievement of students registered in mega universities such as Anadolu University. Therefore, this study was carried out to develop intelligent systems and applications based on the prediction of student achievement.

1.1. Artificial Neural Networks

Core component of the nervous system of biological organism is neurons, i.e. neural cells (Mangels, 2003). It is estimated that the nervous system of a human has 100 billion neurons in average (Mastin, 2010). These cells do work together. As in each system, in biological systems, source (stimulus), receptor (collecting information), processing system (neural network or brain), effector (that turns brain's signals into movements or responses) and response (feedback) processes interoperate (Siegelbaum & Hudspeth, 2000).

Neurons are generally composed of soma, dendrites, axon and synapses (Finger, & Tansey, 1994; Guillery, 2005; Kandel, Schwartz, & Jessell, 2000). Stimuli (signals) from the external environment are transferred to axon from the dendrite in the nerve cell. During this transfer, nonlinear, complex processes take place. Information (signals) transferred to synapse following these processes are transferred to other neurons by means of synapse (Rojas, 2013).

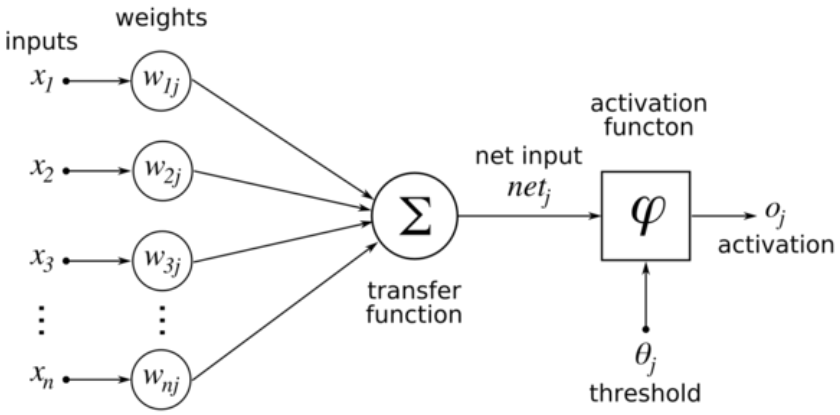


Figure 1. Artificial Neural Network
(Source:https://en.wikibooks.org/wiki/File:ArtificialNeuronModel_english.png)

In engineering sciences, artificial neurons are also called as "processing unit". In Figure 1, x_n represents input while w_{nj} represents weight coefficients. Dendrite in biological nerve cell acts similar to the combining function of artificial neural network; cell body to transfer function; axons to element output and synapses to weights (Bullinaria, 2015). Functioning of an artificial neural network depends on the threshold value; neuron is activated when the result of multiplying input from external environment or another cell ($w_{nj}x_n$) is higher than the threshold value of the cell (Basheer & Hajmeer, 2000; Kotsiantis, Pierrakeas & Pintelas, 2003).

ANNs are systems learning from their experiences, making inferences based on the prior learning, and taking decisions, in a way similar to people (Öztemel, 2012). They have non-linear fields of application such as image processing, image classification, verification, speech

analysis, optimization problems, robot navigation, processing of incomplete or indefinite data, quality assurance, stock market prediction and simulation (Lippmann, 1987).

As seen in Figure 2, ANNs are basically composed of input layer, hidden layer(s) and output layer. ANNs learning is based on iterative loops. In the first stage, network output is produced from the input in the training set. Input used for training is the data enabling ANNs to learn, i.e. to gain experience (Çayıroğlu, 2013). Later on, weights calculated for the network ties are changed according to the accuracy level of the output. Determination of the network output and changes in the weights are realized variously depending on the learning rule and learning algorithm (Öztemel, 2012).

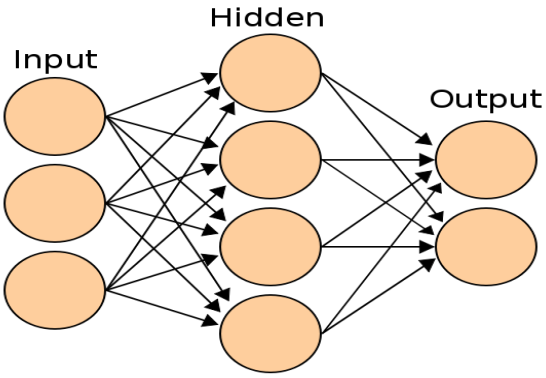


Figure 2. Single Layer Artificial Neural Network Model (Burnett, 2006)

In this study, Multilayer Perception (MLP) and Radial Basis Function (RBF) networks were utilized. MLPs are one of the feed-forward ANN models. They utilize Back Propagation Algorithm (BPA). BPAs are used to minimize the error rate in the network output (Yılmaz, Yavuz & Erkmen, 2013).

RBF operated and popularized by Moody & Darken (1989) are another type of feed-forward BPAs. RBFs can solve non-linear problems but it has been observed that they are poor in determining independent variables explaining dependent variables (Akbilgiç, 2011).

According to Koca (2006), RBFs learn faster than MLPs, and can make classifications and generalisations. In comparison to MLPs, RBFs have simpler architectures (Yu, Xie, Paszczynski & Wilamowski, 2011).

1.2. Achievement and Factors Explaining Achievement

Achievement of an individual in his/her school life is called as academic achievement and today, may represent almost the whole achievement of an individual throughout his/her life. Academic achievement may be defined as an indicator of achieving certain learning objectives (Choi, 2005). Criteria such as grade-point averages, cumulative grade-point averages and course notes are described as academic achievement (Astin, 1991; Snyder, et al., 2002).

A certain part of the studies on academic achievement concerns the examination of the variables explaining the academic achievement of the student. Variables addressed in these studies include *gender* (Amro, Mundy & Kupczynski, 2015; Collins, McLeod & Kenway, 2000; Hajovsky & Kaufman, 2015; Pike, Schroeder & Berry, 1997; Scheiber, Reynolds, Mlambo, 2012); *attitude* (Brown, et al., 2015; Odom & Bell, 2015), *anxiety* (Khalaila, 2015; Macher, Paechter & Papousek, 2015), *socio-economic level* (Jurdak, 2014; Suphi & Yaratan, 2012), *prior learning* (Musso, Kyndt, Cascallar & Dochy, 2013; Power, Robertson & Baker, 1987; Strayhorn, 2006), *self-efficacy* (Valentine, DuBois & Cooper, 2004).

In his study Hattie (2009) examined the factors affecting achievement, by synthesising more than 800 meta-analysis and pointed that gender has a small effect size (d = .12), while

socio- economic level has medium (d = .57), computer assisted training has medium (d =.48), attitude has medium (d = .36) and previous achievement has medium (d = .67) effect sizes. In addition to these, he addressed 138 more factors affecting achievement, in his study. Considering the studies benefiting from machine learning approaches in the prediction of achievement, findings in Table 1 are determined.

Table 1. Machine learning approaches in the prediction of achievement

Study	Approach
Yukselturk, Ozekes & Türel, 2014	*k - Nearest Neighbour*, *Decision Tree*, *Naive Bayes Classifier* and ANN *(3 th rank)*
Turhan et al., 2013	ANN (best result) and regression analyses
Lykourentzou et al., 2009	ANN displayed more effective performance, in comparison to linear regression.
Aydın, 2007	C5.0, Logistic Regression, ANN, C&RT, CHAID and QUEST
Rusli et al., 2008	Prediction through ANN is provided more exact results than decision tree and linear regression.
Naik & Ragothaman, 2004	Prediction through ANN is 93.38% exact.

This study generally aims to predict the final exam scores and pass/fail rates of the students taking Basic Information Technologies – 1 (BIL101U) course in 2014-2015 and 2015-2016 academic years in the Open Education System (OES) of Anadolu University, through Artificial Neural Networks (ANN). On the basis of this main aim, below-given questions were tried to be answered:

1. What are the variables that explain the final exam scores of the students taking BIL101U course?
2. At which level do MLP and RBF networks explain final exam scores?
3. Which one of the MLP and RBF type ANN models provide more exact results when pass/fail rates are determined on the basis of the predicted final exam scores?
4. Which one of the MLP and RBF type ANN models can predict pass/fail state of the student more exactly?

In this study, it was examined whether the final exam scores and pass/fail state of the students registered in OES can be predicted by means of ANN models, and accordingly, variables that explain achievement best were determined. In line with the relations among independent variables, dependent variables were explained. Therefore, correlational research model was utilised.

2. MATERIAL METHOD

2.1. Participatiants

Research population of this is composed of students taking common and compulsory BIL101U course in the fall semester of 2014-2015 (SG-1) and 2015-2016 (SG-2) academic years in different faculties (Open Education Faculty, Faculty of Business Administration and Faculty of Economics) and departments in the OES of Anadolu University. In order to prevent difference among the variables in the dataset, it was preferred to utilise data of two academic years (2014-2015 and 2015-2016).

Number of students in SG-1 was 306.633, while the number of students in SG-2 was 319.845. Total number of students taking BIL101U course in two academic years was 626.478. Following the data cleaning process, as explained in detail under the next title, the

total number of students included in the analysis in two years was decreased to 195.584. 93% of the students taking BIL101U course passed.

2.2. Preparing data for the Analysis

Registration procedures of the students in the OES are carried out at the beginning of each fall semester. During registration, TR identity numbers of the students are identified as unique, and data transfer to OES database is realized according to certain variables in the OSYM (Student Selection and Placement Centre) database.

Table 2 displays the data obtained from the OES and Computer Research Centre (CRC) of Anadolu University, under "demographic", "educational background", "OES" and "other" columns.

Table 2. Variables in the analysis

Demographic	Educational Background	OES	Other
Year of birth, TR identity no, Nationality, Gender, Province	Year of Graduation from the High School, University entrance score (UES), High school type, High school code, High School GPA, Foreign Language, UES Score type, Quota type, University placement ranking	Mid-term exam, final exam, letter grade	Faculty, department, type of registration

OSYM based data was taken from OES and data regarding the variables in Table 3 was taken from CRC; however, some variables were converted during the preparation phase of the analysis. This conversion process is displayed in Figure 3.

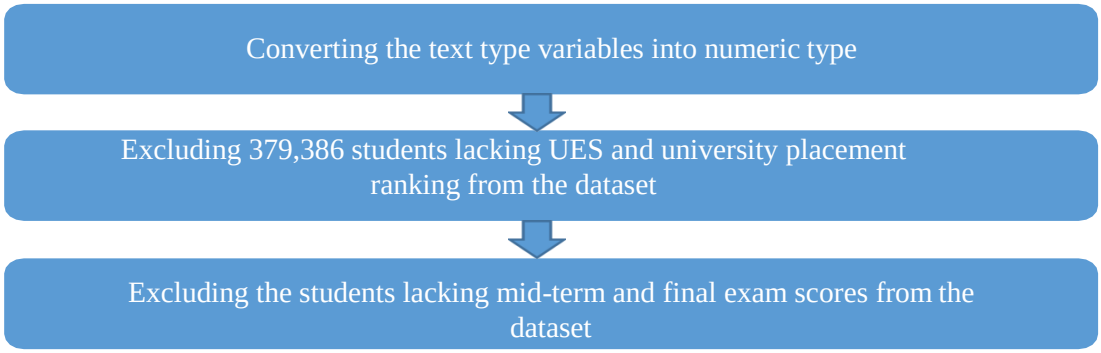


Figure 3. Preparation Process of the Data for the Analysis

Variables of TR identity number, which is used only for identification; and school code, given to each school separately were not used in the data analysis. Variable of university placement ranking was not included in the analysis, as this data was missing for many students. Instead, UES was preferred.

2.3. Data Analysis

In compliance with the aim of the research, ANN was established according to different parameters with MLP and RBF. SG-1 dataset was used as training and test set for the networks; while SG-2 was used as validity set in the determination of the prediction level of the network.

Training set composed 70% of SG-1, while test set composed 30%. In the determination of training and test set, IBM SPSS Statistics v21, *136940* core initial value selected randomly

by the researcher, and *2*rv.bernoulli(0.7)-1* relation were utilized (IBM Knowledge Center, n.d.).

Accordingly, by benefiting from the training, test and control groups, twelve MLPs consisting of various combinations of six parameters, and four RBFs consisting of various combinations of three parameters were established. Characteristics of these networks are displayed in Table 3 and Table 4.

Table 3. Parameters used for MLP analyses

Network Name	Hidden Layer	Min. Unit	Max. Unit	Training	Activation Function	Re-scaling of continuous variables
MLP-A	1	1	50	Batch	Hyperbolic	Standardized
MLP-B	1	1	50	Batch	Hyperbolic	Normalized
MLP-C	1	1	50	Online	Hyperbolic	Standardized
MLP-D	1	1	50	Online	Hyperbolic	Normalized
MLP-E	2	Auto	Auto	Batch	Hyperbolic	Standardized
MLP-F	2	Auto	Auto	Batch	Hyperbolic	Normalized
MLP-G	2	Auto	Auto	Batch	Sigmoid	Standardized
MLP-H	2	Auto	Auto	Batch	Sigmoid	Normalized
MLP-I	2	Auto	Auto	Online	Hyperbolic	Standardized
MLP-J	2	Auto	Auto	Online	Hyperbolic	Normalized
MLP-K	2	Auto	Auto	Online	Sigmoid	Standardized
MLP-L	2	Auto	Auto	Online	Sigmoid	Normalized

As seen in Table 3, 12 MLP networks were established with different parameters, which are explained below:

- Hidden Layer: is a place which is composed of one or more layers and where the network realizes learning process (Nabiyev, 2005). IBM SPSS Statistics v21 software supports one or two layers option for MLP.
- Min - Max Unit: is the number of processing units within the hidden layers. The number varies between 1 and 50.
- Training: Batch training, which is one of the parameters utilized in this study, minimizes the total error, and changes weight values after all data is known (Shalev- Shwartz, 2011). Online training is preferred in comprehensive datasets and changes weight value of each data during learning.
- Activation Function: Under IBM SPSS Statistics v21 software, there is Hyperbolic Tangent which provides output between [-1,1] by subjecting input to tangent function, and Sigmoid function which enables input to be converted into values between [0,1].
- Re-scaling of Continuous Variables: In order to develop the training of the network, continuous variables may be re-scaled. Under this study, parameters of Standardized, which converts data to a range between [0, 1] and Normalized, which converts data to a range between [-1, 1] were tested.

Table 4. Parameters used for RBF Analyses

Network	Hidden Layer	Activation Function	Re-scaling of Continuous Variables
A	Auto	Normalized	Standardized
B	Auto	Normalized	Normalized
C	Auto	Ordinary	Standardized
D	Auto	Ordinary	Normalized

In addition to MLP parameters displayed in Table 4 and explained later on, activation function type: "simple" was used for RBF-C and RBF-D among RBF networks. In simple RBF selected for activation function, basically exponential activation function was used to ensure the usual distribution of hidden layers (Matignon, 2005).

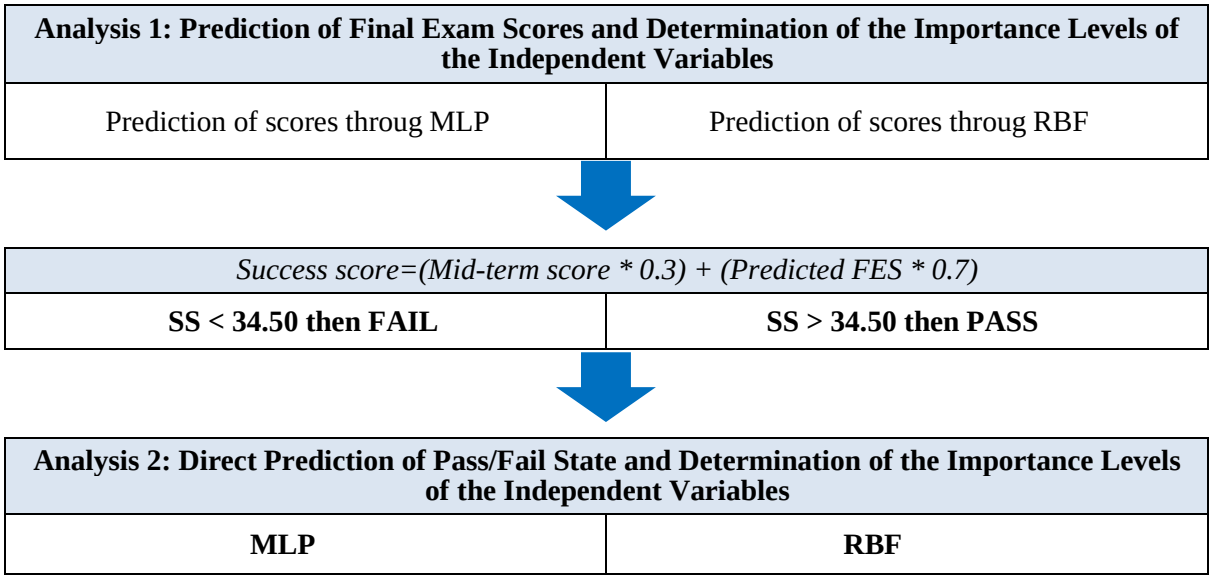


Figure 4. Data analysis process

16 different networks presented in Table 3 and Table 4 was operated twice, one to predict the final exam scores of the students and one to predict the pass/fail state of the students. 32 analyses were carried out in total.

Correlation between the final exam score predicted as a result of the first analysis and the actual score of the student was determined through Pearson Product-Moment Correlation Coefficient. When pass/fail state of the students were determined based on the predicted final exam scores, inconsistencies were examined through crosstabs. During the second analyses, in the networks predicting pass/fail state of the students, consistency between the predicted and actual state was tried to be determined through crosstabs.

3. FINDINGS

3.1. Analysis 1: Prediction of Achievement in Final Exam through MLP and RBF

When the correlation between the results of twelve MLP analyses realized to predict the final exam scores of the students and the actual final exam scores of the students is examined, findings displayed in Table 5 were obtained. Moreover, variables used in each network to explain the final exam scores of the students, and their importance levels were determined.

Table 5. Importance levels, correlation and determination coefficients of the independent variables, obtained as a result of the MLP analyses

Network Name	r_{gk}	r_{gk}^2	Importance Level	Network Name	r_{gk}	r_{gk}^2	Importance Level
A	.482**	.232	mes (100), ues (72.4)	G	.462**	.213	mes (100), ues (55.7)
B	.486**	.236	mes (100), ues (61.6)	H	.445**	.198	mes (100), hst (50.3)
C	.431**	.186	mes (100), ues (71.8)	I	.438**	.192	mes (100), ues (64.2)
D	.311**	.097	mes (100), hst (93.8), ues (79.8)	J	.171**	.029	mes (100), hst (99.6), dept (91.4)
E	.469**	.220	mes (100), ues (71.6)	K	.469**	.220	mes (100), ues (56)
F	.391**	.153	mes (100), prov (67.1), hst (63.0)	L	.369**	.136	mes (100), dept (64.1)

mes: Mid-term exam score, ues: University entrance score, hst: High school type, prov: Province, dept: Department

When Table 5 is examined, it is seen that the network that has highest correlation with the actual final exam scores of the students is MLP-B. MLP-A, MLP-E, MLP-G and MLP-K networks are also among the networks with highest correlation. However, when all the correlation coefficients are examined, medium level correlation coefficients are observed between the predicted and actual scores. Considering the determination coefficients (r^2), it has been found out that the variables in these networks explain the final exam scores of students in a range between 21.3% and 23.6%. Accordingly, it may be asserted that 76.3% and 78.7% of the final exam scores of the students are explained by other variables.

Table 6. Importance levels, correlation and determination coefficients of the independent variables, obtained as a result of the RBF analyses

Network	r_{gk}	r_{gk}^2	Importance Level
A	.224	.05	ues (100), yghs (59.8)
B	.141	.01	tor (100), faculty (87.2)
C	.112	.01	ues (100), faculty (98.8), mes (77.4)
D	.135	.01	tor (100), tor (75.7), fl (68.7)

yghs: Year of Graduation from the High School, tor: Type of registration, fl: Foreign language

Considering the determination coefficients in Table 6, low correlation between the output of RBF-C network and the actual state is remarkable. In this context, it may be claimed that RBF networks do not provide consistent results in explaining final exam scores.

3.2. Benefiting from Final Exam Scores Predicted through MLP and RBF in Pass/Fail Decisions

After the final exam scores of the students were predicted through MLP, their success scores were calculated with mid-term and final exam scores. Following the pass/fail decisions taken on the basis of these scores, consistency between the ANN and actual state was examined through crosstabs. Crosstabs were calculated separately for SG-1 utilised in training and test set; and for SG-2 utilised for control. Accordingly, when pass/fail decisions are taken on the basis of the predicted final exam scores, consistency with the actual state for SG-1 was displayed in percentages in Table 7.

Table 7. Determination of pass/fail state with final exam scores predicted through MLP for SG-1

Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0	Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0
A	76.2	6.9	3.6	13.4	G	75.7	7.3	4.1	13.0
B	76.1	7.1	3.6	13.2	H	76.0	6.7	3.8	13.5
C	76.2	6.3	3.6	13.9	I	75.0	7.7	4.7	12.5
D	75.5	6.0	4.2	14.3	J	79.5	0.8	0.3	19.5
E	76.2	6.7	3.5	13.6	K	74.1	8.7	5.6	11.5
F	77.1	5.1	2.7	15.2	L	76.5	5.4	3.2	14.8

P_M – P_0: State of passing the course both in actual state and according to MLP
F_M – F_0: State of failing the course both in actual state and according to MLP
F_M – P_0: Fail decision according to MLP, when actually the student passes
P_M – F_0: Pass decision according to MLP, when actually the student fails

In interpreting Table 7 and similar tables, P_M-F_O column displaying that ANN decides that the student passes when actually the student fails has been determined as intolerable error by the researchers. Accordingly, it may be asserted that the network "K" leads to lower level of errors. Network "J" provided "pass" decision for 99% of the students.

Table 8. Determination of pass/fail state with final exam scores predicted through MLP for SG-2

Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0	Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0
A	80.9	4.0	11.6	3.5	G	79.5	4.2	13.0	3.3
B	79.8	4.2	12.8	3.3	H	81.4	3.8	11.2	3.7
C	82.5	3.4	10.0	4.0	I	77.0	4.6	15.5	2.9
D	83.1	2.7	9.4	4.8	J	91.6	0.5	0.9	6.9
E	80.9	4.0	11.6	3.5	K	76.2	4.8	16.4	2.7
F	85.3	2.8	7.2	4.6	L	83.5	3.2	9.1	4.3

In the case when SG-2 is used as validity set, it is remarkable that there is a decrease, in the level of errors in all networks, as seen in Table 8. Moreover, it is also seen that network "K" is the network with lowest level of errors.

According to the analysis with RBF, when pass/fail decisions are taken on the basis of predicted final exam scores, consistency with the actual state for 2014-2015 academic year was displayed in percentages in Table 9.

Table 9. Determination of pass/fail state with final exam scores predicted through MLP for SG-1 and SG-2

	SG-1					SG-2			
Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0	Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0
A	79.3	0.9	0.5	19.3	A	91.7	0.5	0.8	7.0
B	79.7	0.3	0.0	20.0	B	92.3	0.2	0.2	7.3
C	79.7	0.0	0.0	20.3	C	92.5	0.0	0.0	7.5
D	79.7	0.3	0.0	20.0	D	92.4	0.2	0.2	7.3

When the data about the students, in the validity set is examined, actual state is consistent with RBF networks for nearly 92% of the students as similar to the results of training and test set, but RBF networks lead to decisions different from the actual state for about 7.5% of the students.

3.3. Analysis 2: Direct Prediction of Pass/Fail State and Determination of the Importance Levels of the Independent Variables

Instead of the final exam scores of the students, pass/fail state of the students were tried to be predicted directly through ANN. During this process, parameters in Table 3 were used for MLP networks, while those in Table 4 were used for RBF networks. In predicting the pass/fail state of the students, *passfail* variable set by the researcher was used as the dependent variable. This variable enables binary prediction with regard to the pass/fail state of the students.

Table 10. In MLP networks, Importance Levels of the Independent Variables Utilised in the Prediction of Pass/Fail State

Network	Importance Levels	Network	Importance Levels
A	mes (100), ues (40.0)	G	mes (100), age (33.2)
B	mes (100), ues (35.3)	H	mes (100), hst (83.4)
C	mes (100), hst (44.6)	I	mes (100), age (57.9)
D	mes (100), yghs (29.7)	J	prov (100), hst (83.5)
E	mes (100), ues (34.2)	K	mes (100), ues (23.0)
F	hst (100), prov (93.7), mes (82.2)	L	hst (100), mes (90.8), prov (81.3)

It was observed that F, H and L networks, among MLP networks are different from the other networks in terms of their breakdown of relative importance levels. When the erroneous prediction rates of these networks in Table 11 are examined, it is seen that these networks have more errors in the training set, in comparison to the other networks, but this error level is observed to decrease in validity set.

Table 11. Decisions of pass/fail state for SG-1 and SG-2 taken through MLP

	SG-1					SG-2			
Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0	Network	P_M-P_0	F_M-F_0	F_M-P_0	P_M-F_0
A	75.2	7.9	4.5	12.3	A	78.0	4.5	14.5	3.0
B	75.1	7.9	4.7	12.4	B	79.2	4.2	13.3	3.3
C	74.9	6.9	4.8	13.4	C	79.3	3.9	13.2	3.5
D	75.0	5.3	4.8	15.0	D	83.7	2.5	8.9	5.0
E	74.7	7.9	5.0	12.3	E	78.5	4.2	14.0	3.3
F	79.7	0.0	0.0	20.2	F	92.5	0.0	0.0	7.5
G	75.3	7.5	4.5	12.8	G	79.4	4.1	13.1	3.4
H	78.4	2.0	1.4	18.3	H	89.6	1.1	2.9	6.4
I	74.4	7.8	5.3	12.4	I	78.6	4.1	13.9	3.4
J	76.1	3.2	3.7	17.1	J	86.9	1.2	5.6	6.3
K	77.5	5.0	2.2	15.3	K	82.5	3.0	7.4	4.4
L	79.7	0.1	0.0	20.2	L	92.5	0.0	0.0	7.5

It is seen in Table 11 that the error ratios of the networks in SG-1 are higher than that of SG-2. Erroneous prediction rates of MLP networks in training, test and validity sets are presented in Table 12. As seen in this table, while the error rate of MLP-F, MLP-H and MLP-L networks is close to 20% in training and test sets, their error rate in validity set is below 10%.

Table 12. Erroneous prediction rates of MLP networks

Network	Training	Test	Validity	Network	Training	Test	Validity
A	16.8	17.0	17.5	G	17.3	17.3	16.5
B	17.1	17.1	16.6	H	19.5	20.0	9.3
C	18.3	18.0	16.7	I	17.7	17.9	17.3
D	19.7	19.7	13.9	J	20.6	21.2	11.9
E	17.4	17.2	17.3	K	17.5	17.5	11.8
F	20.1	20.6	7.5	L	20.1	20.6	7.5

In RBF analyses where pass/fail states instead of the final exam scores of the students are predicted through ANN, pass/fail variables and parameters in Table 13 were used, as in MLP analyses.

Table 13. Pass/fail decisions for SG-1 and SG-2 taken through RBF

	SG-1					SG-2			
Network	P_R-P_0	F_R-F_0	F_R-P_0	P_R-F_0	Network	P_R-P_0	F_R-F_0	F_R-P_0	P_R-F_0
A	79.3	0.9	0.5	19.3	A	91.7	0.5	0.8	7.0
B	79.7	0.3	0.0	20.0	B	92.3	0.2	0.2	7.3
C	79.7	0.0	0.0	20.3	C	92.5	0.0	0.0	7.5
D	79.7	0.3	0.0	20.0	D	92.4	0.2	0.2	7.3

In Table 13, it is seen that although 92.5% percentage of passing, which is same for all the networks is the closest result to the actual state as 93%, these networks provided "pass" decision for all students. In this context, it was determined that RBF networks are not successful in classifying pass/fail states or the students.

Table 14 displays the erroneous prediction rates in training, test and validity sets when pass/fail states of the students are predicted through RBF instead of predicting the final exam scores of the students through ANN.

Table 14. Erroneous prediction rates of RBF networks

Network	Training	Testing	Holdout
A	20.1	20.6	7.5
B	20.1	20.6	7.5
C	20.1	20.6	7.5
D	20.1	20.6	7.5

Consistency between the predictions and the actual state was examined by means of crosstabs and the results are presented in Table 15.

Table 15. Pass/Fail Decisions taken through RBF for SG-1 and SG-2

	SG-1					SG-2			
Network	P_R-P_0	F_R-F_0	F_R-P_0	P_R-F_0	Network	P_R-P_0	F_R-F_0	F_R-P_0	P_R-F_0
A	79.7	-	-	20.3	A	92.5	-	-	7.5
B	79.7	-	-	20.3	B	92.5	-	-	7.5
C	79.7	-	-	20.3	C	92.5	-	-	7.5
D	79.7	-	-	20.3	D	92.5	-	-	7.5

When Table 15 is examined, it is seen that the erroneous decision rates in validity sets decreased to 7.5%. However, it was found out that RBF networks provided "pass" decision for

all students. These findings are similar to the inconsistencies in the results of RBF networks, through which final exam scores of the students were predicted.

4. DISCUSSION

According to MLP analyses, final exam scores of the students taking BIL101U course and included in SG-2 are mostly explained by the variables of *Mid-term Score, Year of Graduation from the Secondary School, Placement Score, Type of High School and Province of the Address.* According to RBF analyses, the variables that mostly explain final exam scores are the variables of *Registration Type*, *Placement Score, Type of High School* and *Faculty*.

When the variables affecting the student achievement are examined, it is seen that the variable of gender is important (Amro, Mundy & Kupczynski, 2015; Mlambo, 2012; Collins, McLeod & Kenway, 2000; Pike, Schroeder & Berry, 1997; Scheiber, Reynolds, Hajovsky & Kaufman, 2015; Zheng, 2002). Considering all the analyses carried out within the framework of this study, it has been found out that the variable of gender has low level of importance within SG-1 and SG-2. This study asserting that the gender is not an important explanatory variable of achievement does not coincide with the previous researches in terms of this variable.

With regard to the courses at the level of bachelor's degree, concerning information technologies, it was found out that gender is not a important explanatory variable of achievement (Fan, 1998; Werth, 1986; Wilson, 2002; Wilson & Shrock, 2001). Considering the literature on the achievement in courses related to information technologies, the results of this study coincide with the studies claiming that gender is not a significant explanatory variable.

Another factor determined to explain student achievement is the educational background (Power, Robertson & Baker, 1987; Strayhorn, 2006; Zheng, 2002). In his study on the examination of explanatory variables of student achievement, Wolfe (1995) found out that high school grade-point average is the first explanatory variable while score in SAT, which is an exam taken by American students for placement in university is the third explanatory variable of achievement. In this study, variables of score (SSPE score) required in registration to OES, year of graduation from the secondary school, type of high school being graduated were addressed under the educational background category. It may be asserted that these variables have high importance for the overall MLP networks, are significant in terms of explaining student achievement, and thus, the results of this study are similar to the results of previous studies.

In general, maths competence of the students acquired during high school years explain their achievement in information technologies courses at the level of bachelor's degree, (Oman, 1986; Wilson & Shrock, 2001; Wilson, 2002). This study did not include a variable of maths competence, but it is observed in the related literature that the SAT score addressed in the studies, which is similar to SSPE score in Turkey, is an explanatory variable with low significance (Ventura, 2005). Considering that SSPE score is an important explanatory variable in this study, it may be said that this study is not in parallel with the previous studies in this regard.

It was found out that predictions through MLP provided more exact results than predictions through RBF, with a difference of nearly 4% in error rate. In their study, Huang and Fang (2013) compared MLP and RBFs, and as in this study, they asserted that MLP network provided more exact predictions than RBF network.

When the studies on the prediction of student achievement are examined, it is seen that in general, different mathematical models such as Regression, Artificial Neural Networks,

Decision support Systems, Decision Trees and Baynes were compared and that ANN displayed a better performance than others (Herzog, 2006; Lykourentzou et al., 2009; Naik and Ragothaman, 2004; Schumacher, Olinsky, Quinn, & Smith, 2010; Şen, Uçar & Delen, 2012; Rusli, Ibrahim & Janor, 2008; Turhan et al., 2013). As distinct from these results, a practice with ANN was ranked the second in the study carried out by Aydın (2007) with an accuracy rate of 77.80%; and ANN analysis was ranked the third in the study carried out by Yükseltürk et al. with a classification rate of 79.7%. Although comparison of ANN models with other mathematical models is not aimed or practiced in this study, the accurate prediction rate ranging between 85% and 87%, obtained through different parameters should be taken into account.

It was observed that certain parameters led to low performance in predictions as each network was established on different parameters. When MLP-D and MLP-J out of MLP networks established to predict the final exam scores of the students are examined, it was found out that they have a low correlation with the actual state. These networks were established with *online* training method, *Hyperbolic Tangent* function and *Normalized* scaling parameters. Therefore, utilisation of the combination of these parameters in the prediction of scores does not ensure a good performance.

When interpreting the findings obtained from the second application set in which predictions are made directly on pass/fail state of the students, error rates resulting from the "failed" decision of MLP although the student passes in the actual state were defined as *negligible errors* in this study. With an error rate below 3%, MLP-K and MLP-I networks were identified as the networks with the fewest error. Both networks were established with online training and Standardized scaling parameters. When the parameters off MLP-A, MLP-B, MLP-C, MLP-E, MLP-G and MLP-H networks, whose error rate is below 4% are examined, it is seen that Hyperbolic Tangent function was used in these networks, in general. Similarly, the study carried out by Özkan & Erbek (2003) pointed that in classification problems, Hyperbolic Tangent function displays better performance than the sigmoid function; and thus, this result complies with the results of this study.

When the findings of MLP and RBF analyses are examined with an overall perspective, it is seen that in comparison to RBF networks, MLP networks provided more exact results in the prediction of both final exam scores and pass/fail states. Moreover, considering the importance levels of the independent variables in RBF analyses, there is no covariance among the networks. According to Akbilgiç (2011), there are problems in determining the independent variables affecting the dependent variable in RBF and hybrid networks.

Final exam score and pass/fail state predictions, where SG-1 was used as training and test set and SG-2 as validity set provided more exact results than the cases where SG-1 was used singly with a division into training and test sets. Accordingly, it may be claimed that when ANN is used in the prediction of student achievement in a course, obtaining training and test sets from the data sets of previous years may ensure more exact results.

Suggestions for the future studies and practices to be carried out within the scope of the findings and results of this research are as follows:

- In this research aiming at the prediction of the student achievement, the student scores only in BIL101U course were utilized. As the content of BIL101U course is related with information technologies, variables such as math competence and educational background play an important role in terms of explaining achievement in this course or similar courses. In this context, data concerning the variables such as right/wrong rates in SSPE, which is stated to be a good explanatory variable of achievement, experience of a computer course, student's expectations related to the course may also be collected for analysis.

- Considering that each course has different aims and objectives and thus different learning outcomes, explanatory variables of achievement may not be same. Therefore, carrying out similar analyses with similar or different variables for other courses may provide a scientific conclusion in this topic.
- As a result of the analyses with ANN, it has been found out that the variable with the highest effect on explaining achievement is mid-term score. Accordingly, data obtained from the pilot tests and similar practices in the "Anadolum eKampüs" application of Anadolu University may be included in ANN analyses to examine whether these process evaluation variables create a difference in the prediction of achievement.
- In line with the findings of the analyses carried out, pass/fail states of the students may be predicted. Most suitable and goal-oriented model may be selected and integrated into web-based learning environments. This integrated system may be considered as an early warning system for the students. When the student is registered, data obtained from the student himself/herself or from the institution-based systems may be processed through ANN, and predictions regarding the pass/fail state of the students in a course or courses may be made. These predictions should not be communicated to the student through notifications such as "you have passed" or "you have failed" but they may receive suggested topics of study, course materials or pilot test warnings. By this means, "fail" prediction about a student who actually passes – which was considered as a "negligible error" in this study – may be reflected as an additional support to the student, in this process.
- In order to predict the student achievement and to integrate these predictions with the systems used in the learning process, exactness level of these predictions is important. Exact predictions are basically ensured through clear and organized data. Accordingly, this research is important in terms of encouraging educational institutions and organizations to identify the explanatory variables of achievement and to collect data about these or similar variables.
- SSPE score is also one of the most important variables in predicting the student achievement. This variable provides quantitative data about the educational background of the student. Similar studies on the prediction of student achievement (Oman, 1986; Wilson & Shrock, 2001; Wilson, 2002) provide data about the courses taken students during high school education. It is considered that a similar study may be carried out with the OES students to achieve more extensive results, and more exact predictions.

Acknowledgements

This research was produced from master's thesis titled "Predicting Achievement with Artificial Neural Networks: The Case of Anadolu University Open Education System" by Hilal S. Yıldız Aybek in supervising of M. Recep Okur, and supported by Anadolu University Scientific Research Projects Commission.

ORCID

Hilal Seda Yıldız Aybek https://orcid.org/0000-0001-9128-734X
Muhammet Recep Okur https://orcid.org/0000-0003-2639-4987

5. REFERENCES

Akbilgiç, O. (2011). *Hybrid radial basis function networks and variable selection and estimation: an application for securities investment decisions.* (Doctoral dissertation). Istanbul University. Istanbul.

Amro, H. J., Mundy, M., & Kupczynski, L. (2015). The effects of age and gender on student achievement in face-to-face and online college algebra classes. *Research in Higher Education Journal,* 27.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and*

evaluation in higher education. New York: Macmillian.

Aydın, S. (2007). *Data mining and an application in Anadolu University Open Education System.* (Doctoral dissertation). Anadolu University Graduate School of Social Sciences. Eskisehir.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods, 43*(1), 3-31.

Bower, B. L., & Hardy, K. P. (2004). From correspondence to cyberspace: Changes and challenges in distance education. *New Directions for Community Colleges, 2004*(128), 5-12.

Brown, S. J., White, S., Sharma, B., Wakeling, L., Naiker, M., Chandra, S., & Bilimoria, V. (2015). Attitude to the study of chemistry and its relationship with achievement in an introductory undergraduate course. *Journal of the Scholarship of Teaching and Learning, 15*(2), 33-41.

Bullinaria, J. A. (2015) Biological neurons and neural networks, artificial neurons: neural computation: Lecture 2. Retrieved April 20, 2016, from http://www.cs.bham.ac.uk/~jxb/INC/l2.pdf

Burnett, C. M. (2006, December 27). Artificial neural network. Retrieved April 20, 2016, from https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg#/media/Fi

Collins, C. W., McLeod, J., & Kenway, J. (2000). *Factors influencing the educational performance of males and females in school and their initial destinations after leaving school.* Canberra: Department of Education, Training and Youth Affairs.

da Silva, E. T., de Fátima Nunes, M., Santos, L. B., Queiroz, M. G., & Leles, C. R. (2012). Identifying student profiles and their impact on academic performance in a Brazilian undergraduate student sample. *European Journal of Dental Education, 16*(1), 27-32.

Finger, S., & Tansey, T. (1994). Origins of neuroscience: a history of explorations into brain function. *Trends in Neurosciences, 17*(7), 310-310.

Guillery, R. W. (2005). Observations of synaptic structures: origins of the neuron doctrine and its current status. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1458), 1281-1307.

Hattie, J.C. (2009). *Visible learning. a synthesis of over 800 meta-anaylses relating to achievement.* New York: Routledge.

Herzog, S. (2006). Estimating student retention and degree□completion time: Decision trees and neural networks vis□à□vis regression. *New Directions for Institutional Research, 2006*(131), 17-33.

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education, 61*, 133-145.

Ibrahim, Z., & Rusli, D. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In *21st Annual SAS Malaysia Forum*, 5th September.

IBM Knowledge Center. (n.d.). Preparing the Data for Analysis. Retrieved April 21, 2016, from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/tutorials/mlp_bankloan_dataprep.html

Jurdak, M. (2014). Socio-economic and cultural mediators of mathematics achievement and between-school equity in mathematics education at the global level. *Zdm, (7),* 1025.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (2000). Principles of neural science*, (4),* 1227-1246. New York: McGraw-hill.

Khalaila, R. (2015). The relationship between academic self-concept, intrinsic motivation, test

anxiety, and academic achievement among nursing students: *Mediating and moderating effects. Nurse Education Today, (3)*, 432. doi:10.1016/j.nedt.2014.11.001

Koca, Z. (2006). *Speed control with artificial neural networks with vector control of three phase asynchronous motors* (Master's thesis). Kahramanmaras Surcu Imam University Graduate School of Natural and Applied Sciences. Kahramanmaras.

Kose, U., & Arslan, A. (2017). Optimization of self□learning in Computer Engineering courses: An intelligent software system supported by Artificial Neural Network and Vortex Optimization Algorithm. *Computer Applications in Engineering Education, 25*(1), 142-156.

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2003). Preventing student dropout in distance learning systems using machine learning techniques. In *Knowledge -Based Intelligent Information and Engineering Systems,* (pp. 267–274).

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine IEEE ASSP Mag.*, 4-22.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology, 60*(2), 372-380.

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., & Arendasy, M. (2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *British Journal of Educational Psychology, 83*(4), 535-549.

Mangels, J. (2003). Cells of the Nervous System. Retrieved May 3, 2016, from http://www.columbia.edu/cu/psychology/courses/1010/mangels/neuro/neurocells/neurocells.html.

Mastin, L. (2010). Neurons & Synapses. Retrieved from May, 4, 2016, from http://www.human-memory.net/brain_neurons.html.

Matignon, R. (2005). *Neural network modeling using SAS enterprise miner.* AuthorHouse.

Mlambo, V. (2012). An analysis of some factors affecting student academic performance in an introductory biochemistry course at the University of the West Indies. *The Caribbean Teaching Scholar, 1*(2).

Moore, M. G., & Kearsley, G. (2011). *Distance education: A systems view of online learning*. Cengage Learning.

Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1*(1), 42-71.

Naik, B., & Ragothaman, S. (2004). Using neural networks to predict MBA student success. *College Student Journal, 38*(1), 143.

Odom, A. L., & Bell, C. V. (2015). Associations of middle school student science achievement and attitudes about science with student-reported frequency of teacher lecture demonstrations and student-centered learning. *International Journal of Environmental and Science Education, 10*(1), 87-97.

Özkan, C., & Erbek, F. S. (2003). The comparison of activation functions for multispectral Landsat TM image classification. *Photogrammetric Engineering & Remote Sensing, 69*(11), 1225-1234.

Öztemel, E. (2012). *Yapay Sinir Ağları [Artificial Neural Networks]*. İstanbul: Papatya Yayıncılık.

Pike, G. R., Schroeder, C. C., & Berry, T. R. (1997). Enhancing the educational impact of residence halls: The relationship between residential learning communities and first-year college experiences and persistence. *Journal of College Student Development, 38*(6), 609-

621.

Power, C., Robertson, F., & Baker, M. (1987). *Success in higher education* (No. 94). National Institute of Labour Studies.

Rojas, R. (2013). *Neural Networks: A Systematic Introduction*. Springer Science & Business Media.

Rusli, N. M., Ibrahim, Z., & Janor, R. M. (2008, August). Predicting students' academic achievement: Comparison between logistic regression, artificial neural network, and Neuro-fuzzy. In *Information Technology, 2008. ITSim 2008. International Symposium on (1)*, 1-6. IEEE.

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools, 52*(4), 335-348.

Schumacher, P., Olinsky, A., Quinn, J., & Smith, R. (2010). A comparison of logistic regression, neural networks, and classification trees predicting success of actuarial students. *Journal of Education for Business, 85*(5), 258-263.

Shalev-Shwartz, S. (2011). Online learning and online convex optimization. *Foundations and Trends in Machine Learning, 4*(2), 107-194.

Siegelbaum, S. A., & Hudspeth, A. J. (2000). *Principles of neural science* (Vol. 4, pp. 1227-1246). E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.). New York: McGraw- hill.

Simonson, M., Smaldino, S., Albright, M. & Zvacek, S. (2003). *Teaching and Learing at a Distance.* Ohio: Columbus.

Snyder, C. R., Shorey, H. S., Cheavens, J., Pulvers, K. M., Adams, V. H. III, & Wiklund, C. (2002). Hope and academic success in college. *Journal of Educational Psychology, 94*(4), 820-826.

Strayhorn, T. L. (2006). Factors influencing the academic achievement of first-generation college students. *NASPA Journal, 43*(4), 82-111.

Suphi, N., & Yaratan, H. (2012). Effects of learning approaches, locus of control, socio-economic status and self-efficacy on academic achievement: a Turkish perspective. *Educational Studies, 38*(4), 419-431.

Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications, 39*(10), 9468-9476.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Estimation of Student Success with Artificial Neural Networks. *Education and Science, 38*(170).

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133.

Ventura, P. R. (2005). Identifying predictors of success for an objects-first CS1. *Computer Science Education, 15*(3).

Yu, H., Xie, T., Paszczynski, S., & Wilamowski, B. M. (2011). Advantages of radial basis function networks for dynamic system design. *Industrial Electronics, IEEE Transactions on, 58*(12), 5438-5450.

Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and e-Learning, 17*(1), 118-133.

Zheng, J. L., Saunders, K. P., Shelley, I. I., Mack, C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 43*(2), 267.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 491–509

DOI: 10.21449/ijate.444073

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Modeling Course Achievements of Elementary Education Teacher Candidates with Artificial Neural Networks

Ergün Akgün[1,*], **Metin Demir**[2]

[1] The Department of Elementary Education, Usak University, Usak, Turkey
[2] The Department of Elementary Education, Dumlupınar University, Kütahya, Turkey

Abstract: In this study, it was aimed to predict elementary education teacher candidates' achievements in "Science and Technology Education I and II" courses by using artificial neural networks. It was also aimed to show the independent variables importance in the prediction. In the data set used in this study, variables of gender, type of education, field of study in high school and transcript information of 14 courses including end-of-term letter grades were collected. The fact that the artificial neural network performance in this study was R=0.84 for the Science and Technology Education I course, and R=0.84 for the Science and Technology Education II course shows that the network performance overlaps with the findings obtained from the related studies.

ARTICLE HISTORY

Received: 16 April 2018
Revised: 20 May 2018
Accepted: 03 July 2018

KEYWORDS

Elementary Education, Science and Technology Teaching, Data Mining, Artificial Neural Networks

1. INTRODUCTION

Although the key study on artificial neural networks (ANN) was based on the models of McCulloch and Pitts (1943), which they called "A logical calculus of the ideas" it is possible to say that the use of artificial neural networks constructed by means of computers dates back to the 1950s (Heaton, 2008). Historically, basic network architectures were first announced by Frank Rosenblatt as "The Perceptron." Then, chronologically, the progress has been as follows (Graupe, 2007):

- "The Artron" by R. Lee in 1950,
- "The Adaline" by B. Widrow in 1960, and
- "The Madaline" by, again, B. Widrow in 1988.

The ANN models developed in the subsequent years are based on the working principles of these four models in general. The use of this method of analysis that can be regarded as new is rapidly increasing with many artificial neural network architectures being developed. Today, it is possible to find use of artificial neural networks in many different fields including brain and cognition (Gupta, Molfese, & Tammana, 1995), scientific and technical information (Polanco, Francois, & Keim, 1998), environmental planning, design and architecture (Raju,

CONTACT: Ergün Akgün ✉ ea.ergunakgun@gmail.com Usak University Education Faculty 1 Eylul Campus 64100 Usak, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

Sikdar, & Dhingra, 1996; Wyatt, 1996), geographic information systems (Foody, 1995), grammer (Vokey & Higham, 2004), business and economics (Selim, 2009; Suzuki, 2001; Tang, 2009), industrial engineering (Azadeh, Saberi, & Anvari, 2011), energy (Geem, 2010), ergonomics (Nussbaum & Chaffin, 1996), ethology (Snyder, 1998), weather forecast (Ghiassi, Saidane, & Zimbra, 2005), air pollution (Cai, Yin, & Xie, 2009), human behavior and computer (Stevens, Ikeda, Casillas, Palacio-Cayetano, & Clyman, 1999), job security (de Haen, 2009), paleontology (Anemone, Emerson, & Conroy, 2011), psychiatry (Cohen, 1994), psychology (Verhagen & Scott, 2004), psycho-sociological (Dowman & Ben-Avraham, 2008), health (Alam & Briggs, 2011), telematics and informatics (Sim, Tan, Wong, Ooi, & Hew, 2014), prevent and analyze traffic accidents (Wei & Lee, 2007), tourism management (Palmer, Montano, & Sese, 2006), expert systems (Ahn, Cho, & Kim, 2000) and education and training (Demir, 2015).

When the successful applications in these fields are examined, it can be said that artificial neural networks are used especially when there are non-linear, multidimensional, incomplete, imperfect and error-prone data, and where there is no mathematical model for solving a problem (Çırak & Çokluk, 2013). In order to better understand this method, which is trying to imitate the functioning of the human brain, its structure and its basic components need to be examined.

The structure of this method is based on the functions of the human brain. The cells in the brain provide humans with the ability to use and practice their thinking, reasoning, and experience (González & DesJardins, 2002). Artificial neural networks aim to make use of these abilities of the human brain to automatically generate new information through the features of learning, discovery and construction without any help (Çırak, 2012; Yavuz & Deveci, 2012). It is important to know how a neural network works biologically to be able to better understand the working principle of artificial neural networks and the elements of a network.

The biological nervous system in the human body consists of a three-layered structure that includes receiving data, interpreting them, and making decisions (Kuyucu, 2012). The brain is in the center of this system shown as the "Nervous System Block Diagram" in Figure 1. The brain receives information, makes sense of it and makes a proper decision. Arrows from left to right convey the information-bearing signals into the system through feedforward, and the arrows from right to left through feedback (Haykin, 2009). Receptors transmit stimuli from the human body and environment to the neural network by turning them into electrical impulses; the effectors turn these impulses into understandable reactions as an output of the system (Haykin, 2009).

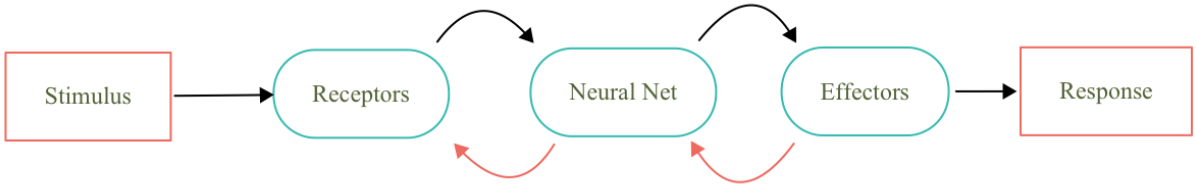


Figure 1. Nervous system block diagram (Haykin, 2009).

The basis of the biological neural networks is the nerve cells. The number of nerve cells in the cortex of the human brain is estimated to be about 10 billion (Cuhadar, 2006). A nerve cell is composed of a cell body, a dendrite and an axon. At the macro level, in nerve cells, which work in a way similar to the working principle of the nervous system, incoming stimuli are transmitted to the cell body via dendrites. Outputs generated after operations in the cell body are transmitted to other nerve cells via axons.

The similarity of the working principle of the artificial neural networks and the elements of the network with the biological neural cells is shown in Figure 2. The connections between the cells correspond to axons and dendrites. The weight factors (W_k) correspond to synapses. In an artificial nerve cell, stimuli coming to the cell (X_1, X_2, ... X_m) based on the effect of the weight factor (Wk_1, Wk_2, ... Wk_m) are converted into stimuli as outputs in response to a nonlinear activation function by taking into account the state or grade of intracellular synaptic weights (Koç, Balas, & Arslan, 2004).

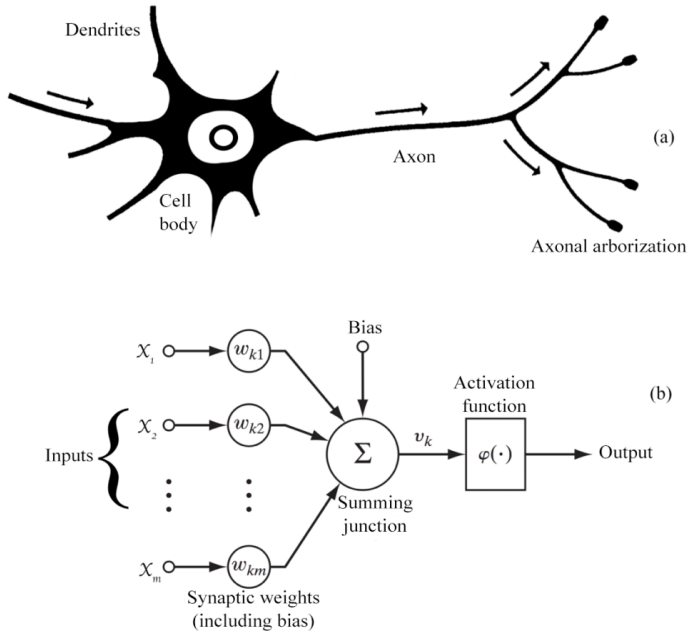


Figure 2. Similarity between biological and artificial neural networks (Arbib, 2003a; Haykin, 2009b).

The basic components of an artificial neural network consist of six different components: input layer, weights, hidden layer, summing function, activation function and output layer. Biologically, a neuron corresponds to a processor element in artificial neural networks. Dendrites are used as inputs, and the cell body is used as a substitute for transfer functions used in the network architecture. Weights used in artificial neural networks fulfill the function of synapses, while the axons represent the neuronal output of the artificial neural network. The basic components of the artificial neural network can be explained as follows.

Input Layer: Input signals (X_1, X_2, ... X_m) sent to the input layer are transmitted to the next layer without any statistical processing. The only function of this layer is to transmit the data to the next layer (Yurdakul, 2014). In addition, the input signals on this layer can be more than one. The input signal sent to the input layer may be any of the texture, mathematical value, audio signal, or image processing elements according to the type of the network.

Synaptic Weights: Synaptic weights (Wk_1, Wk_2, ... Wk_m) are the statistical coefficients indicating the importance of the input data to the hidden layer and its effect on the learning of the network. A positive or negative weight is generated for each input signal (Kuyucu, 2012). All links that provide the relationship between input signals and other layers have different weight values. This variant assignment of values ensures that synaptic weights are effective on all processing elements (Yurdakul, 2014). In artificial neural networks, the display of information is provided by these weights. Therefore, synaptic weights are an important variable affecting the design and performance of a network (Emir, 2013). In the determination of this variable, it is assumed that the input signals have some statistical distributions.

Hidden Layer: Intermediate layers providing information exchange between the input and output layers are called hidden layers (Şengür & Tekin, 2013). More than one hidden layer can be found in an artificial neural network. However, depending on the problem situation, if there are not enough hidden layers, the network fails to learn. If there is more hidden layer than necessary, it will cause the network to memorize the current situation and make the learning situation impossible (Yurdakul, 2014).

Summing Junction: This junction, also called merging, calculates the net input to the cell (Adıyaman, 2007; Yavuz & Deveci, 2012). It makes this calculation by using the linear combination of the weights of the respective input values for each input value (Emir, 2013). The addition functions can change in different models depending on the structure of the network architecture. In some models, input values are important, while, in some other models, the number of inputs may be an important variable. Taking this difference into account, the most appropriate addition function is determined by trial and error (Çuhadar, 2006; Kuyucu, 2012).

Activation Function: The activation function, also called the transfer function, determines the output that the cell produces in response to the input by processing the net input obtained from the addition function (Gülçe, 2010; Kayıkçı, 2014). Separate activation functions are applied to all of the nerve cells in the network, and the output value can be calculated by means of the statistical value obtained after the function (Çırak, 2012). As with the addition function, one can benefit from different functions in the use of this function.

Output Layer: In this layer, the information to which an appropriate activation function is applied is processed to produce the output required for the input data given in the first stage to the network. The output layer consists of a single layer, where the generated data are transmitted to the outside world.

Although the components forming an artificial neural network are mainly composed of these components, artificial neural networks vary in many different classifications according to their intended use. In this study, it was aimed to predict elementary education teacher candidates' achievements in "Science and Technology Education I and II" courses by using artificial neural networks. It was also aimed to show the importance levels of the independent variables in the prediction. The features and the construction of the network architecture created in this framework are explained in the Method section.

2. METHOD

2.1. Sample

The data in this study were obtained from the transcripts of elementary education teacher candidates graduating from four different state universities and the demographic information without personal information found in student information systems. Graduates who have not graduated withinfour years were not accepted into the sample. Moreover, the data of students who attended and left the current programs with a lateral or vertical transfer via some programs implemented by the Council of Higher Education of Turkey (such as Mevlana or Erasmus Student Mobility) were removed and not included in the study. After these procedures, the data of a total of 865 graduates were analyzed within the scope of the study.

2.2. Data Analysis

The data in this study were obtained from the transcripts of elementary education teacher candidates graduating from four different state universities and the demographic information without personal information found in student information systems.

In the data set used in this study, variables of gender, type of education, field of study in high school and transcript information of 14 courses including end-of-term letter grades were collected and coded as categorical data. In the data analysis stage, these data were converted

into numerical data using the methods described below. Categorical variables such as gender (male, female), type of education (daytime education, evening education) and field of study in high school (Turkish, Mathematics, Science, other) were converted into numerical data with 1-of-N encoding method such as 'male' = [1 0] and 'female' = [0 1]. In addition, the placement score given by the Assessment Selection and Placement Center of Turkey (ASPC), which is another variable, was also included in the data set the same way as it was used by the ASPC during placement.

When the end-of-term letter grades containing the transcript information of the 14 courses that constitute the second part of the data set were examined, it was seen that these letter grades represented numerically different grade ranges among the universities. For this reason, it was found which score ranges the letter grades represented at each university by checking the regulations for education and examinations valid for years in which each sample graduated. Then, each university was separately transformed into the same standards at intervals given in the table of grades of the Higher Education Council, by using the Higher Education Council Grade Transformation Table.

Once the data were converted, the dataset was divided into two parts to construct the network and test the problem situations that constituted the purpose of the study. The main purpose of this separation was to (1) determine the best network performance and then to (2) evaluate the performance of the prediction data that the constructed artificial neural network would generate in the presence of the new data set that it would encounter for the first time. In other words, this separation would allow the comparison of two things: (1) the output data that the artificial neural network would produce when it takes the independent variables as input data, assuming that the artificial neural network has learned the relationship between dependent and independent variables, and (2) the actual data of the dependent variable in the data set of the study.

In this context, the graphical representation of all operations on the dataset is presented below and each step is explained in detail (Fig 3).

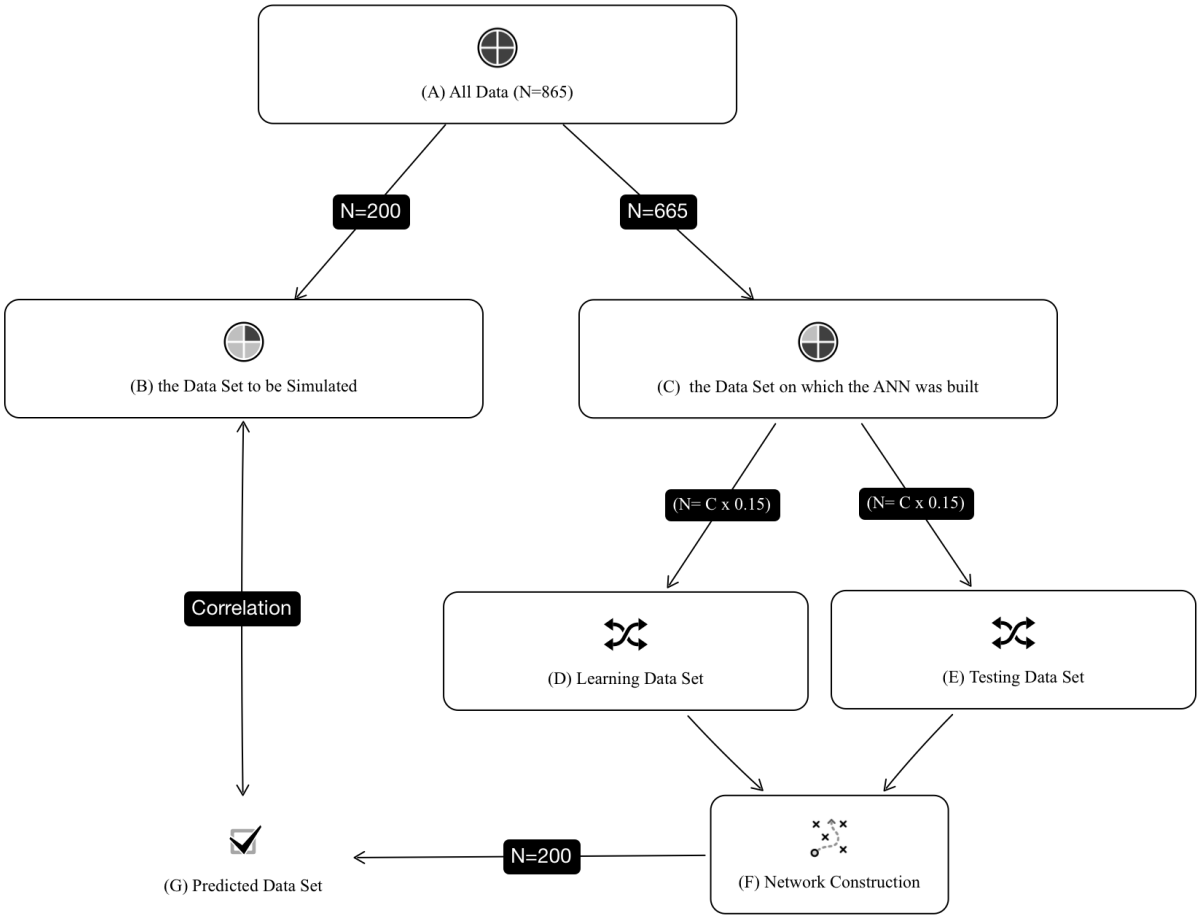


Figure 3. Data set operations for network construction.

Within the scope of this study, (A) all data of N=865 people who constituted the sample were randomly divided into two parts as (B) "the Data Set to be Predicted" and (C) "the Data Set on which the ANN was built" to construct a network using the Matlab Neural Network Toolbox.

(B) "The Data Set to be Simulated" consisted of 200 individuals randomly selected from the 865 individuals constituting the sample. In general, at least 15% of all datasets are used as simulation data in studies (e.g., Bahadır 2013; Başman 2014; Demir 2015), although there is no method or standard for determining this number in the literature. In the context of this study, 23% (N=200) of the entire data set was reserved for comparing the simulation data with the actual data after constructing network. The data set was not seen by the system during the time the network learned the relation between its dependent and independent variables. In other words, the constructed artificial neural network was formed by the data of the remaining 665 people.

(C) "The Data Set on which the ANN was built" consisted of 665 people, except for the data of 200 people that were excluded to be simulated later on. This data set included the following data, which were thought to be predictive of the achievement in the Science and Technology Education course: 665 participants' gender, field of study in high school, ASPC placement score, and year of graduation as well as quantified data of the letter grades of General Biology, Introduction to Educational Science, General Chemistry, Educational Psychology, General Physics, Science and Technology Laboratory Applications I, Instructional Principles and Methods, Science and Technology Laboratory Applications II, and Environmental Education courses. When the network learns within this data set, it tries to learn the relationships between dependent and independent variables by using the learning functions

determined by the researcher, by determining random selections from the data in itself. It, then, produces simulation data by testing the performance of the network it has constructed. The learning and testing stages are described in detail below.

(D) In the "Learning Data Set", the system tests the user-specified learning functions during the network learning by using 15% of the participants with random selections from the data set given to it. This 15% rate can be increased or decreased by the researcher. Within the scope of this study, 15% was used as the default setting for network learning. At this stage, the network tries to discover the dependent and independent variable relations in the dataset with the determined 15% parts to construct the expected output values.

(E) A "Testing Data Set" is constructed to test the relations systematic— established in the network learning stage, which is the preceding stage — in sets of 15% of data. As in the learning data set, the network does this randomly in order to test the performance of the network.

(F) In the "Network Construction" stage, the network — having learned and tested the relationships between dependent and independent variables — predicts the possible outputs of the data to be inputted by the user. Here, the input data of 200 people reserved in the first stage (B) were fed to the network and the outputs were predicted by the network.

(G) A statistically significant difference was sought by comparing the output data generated in the previous stage to the actual data reserved in stage B.

To summarize all these steps, the entire dataset was divided into two parts: the network construction and the comparison of the actual data to the data to be simulated by the constructed network. In this separation process, the data set to be predicted was not included in the network in order to prevent the network from memorizing the outputs. The data set on which the network was constructed was randomly separated by the system to test the network learning and the conditions it has learned, and a network architecture was constructed. It was ensured that the constructed network architecture predicted the outputs by loading only the input data of the dataset reserved in the first stage. Finally, the data that were modeled and predicted by the network were compared with their actual values.

All of these operations in the dataset were done to create the best network architecture. However, the most important thing to know here is that there is more than one network architecture in artificial neural networks. Within the scope of this study, the procedures for creating the best network architecture considering the properties of the problem state and variables is explained in detail in the next section.

2.3. Construction of the Network Architecture

There are hundreds of options for creating network architectures used in modeling and predictioning with artificial neural networks. This feature, which provides the researcher with flexibility in selecting the components to be used in network construction (such as network type, learning algorithm, and transfer function), sometimes causes the researcher to make too many attempts at finding the best combination of components for a proper network selection. In these attempts, the goal is to establish a network structure that learns the desired output values when it encounters a new data set, or in other words, learns relationships between the dependent and independent variables of the research problem in the best way possible. A network architecture that solves the research problem in the best way possible can be described as the architecture that best learns the relationship between existing inputs and outputs. In order to construct the most suitable network architecture, it is necessary — during the selection of components stage — to consider many variables such as the hardware characteristics of the computer in which the network is tried, the type of variables used, and the characteristics of the

desired output data. Given this diversity, many network architectures should be tried to do any modeling, and the architecture that gives the best result should be preferred.

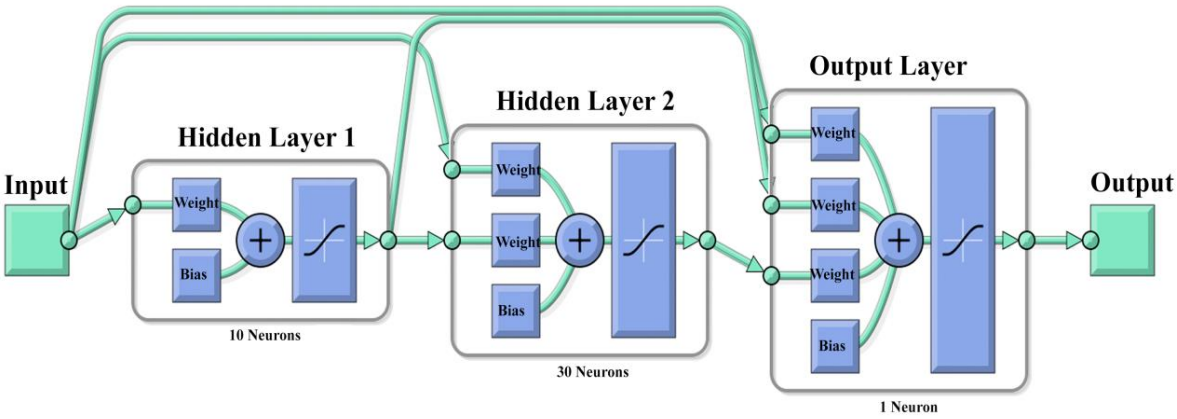


Figure 4. Constructed Network Architecture

Within the scope of this study, 146 different network architectures were constructed to obtain the most suitable network architecture. Each constructed architecture was tested 10 times. Among these 10 trials, the highest performance verification value was recorded together with the R values of the trial, verification and test processes that were created during the network learning stage. Therefore, within the scope of this study, when selecting the most suitable network architecture for the problem situation, 1460 different network architectures were investigated. The network architecture with the lowest Mean-Square Error (MSE) value was taken as the criterion for the most suitable network selection.

The table below shows the characteristics of the network with the best performance value after the trials (Table 1).

Table 1. Characteristics of the Network Architecture (see Fig 4).

Network Type	Adaptation Learning Function	Performance Function	Transfer Function	Number of Hidden Layers	Learning Function	Number of Neurons	
						Hidden Layer 1	Hidden Layer 2
Cascade-forward Backpropagation	Momentum Weights Gradient Descent	Mean-Square Error	Hyperbolic Tangent Sigmoid	2	Levenberg-Marquardt	10	30

After setting the network type, the stage for the determination of the learning function began. At this stage, Levenberg-Marquardt (TRAINLM), the gradient descent (TRAINGD), and the Powell-Beale restart conjugate gradient descent learning algorithm (TRAINCGB), which yield fast results for nonlinear problems, were tested separately in the architectures.

3. FINDINGS

3.1. Modeling and Estimation of Achievement in Science and Technology Education I and II Courses using Artificial Neural Networks

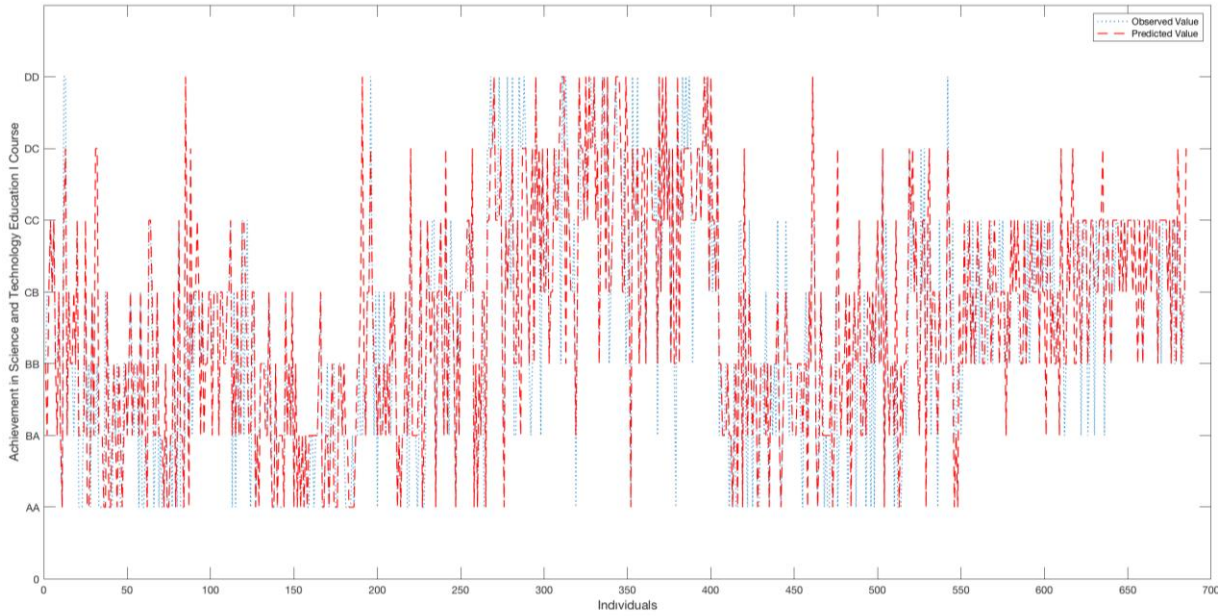
When the network performances of the architectures — prepared to predict achievements in the Science and Technology Education I (Network 1) and II (Network 2) courses and to ensure learning of the network — were examined, it was seen that the mean-

square error values were $MSE_{Network\ 1}$=0.478 and $MSE_{Network\ 2}$=0.427. The table showing the regression coefficients for learning, validation and testing of the artificial neural network is given below (Table 2).

Table 2. Network Performance and Regression Values

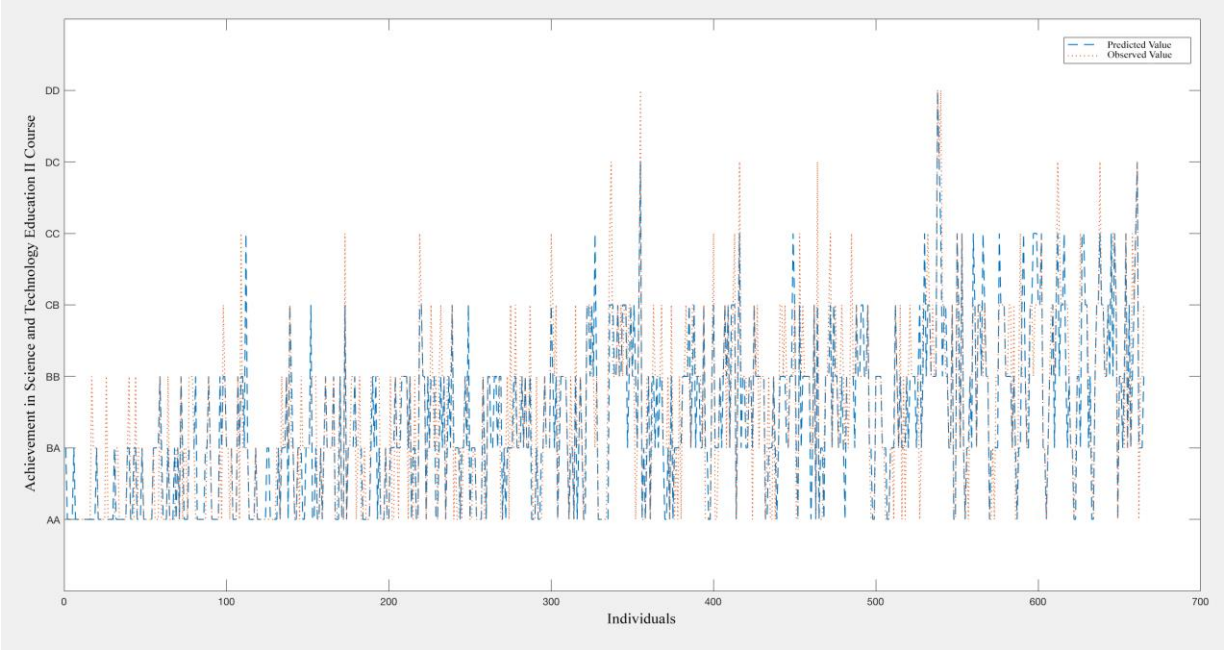
Network	MSE	R			
		Learning	Verification	Test	Total
1	0.47754	0.81632	0.90097	0.87808	0.83774
2	0.42740	0.81917	0.89856	0.83345	0.83679

When the above table is examined, it can be seen that the network regression values were R=0.82 for the learning stage, *R*=0.90 for the verification stage and *R*=0.88 for the test stage in Network 1; and *R*=0.82 for the learning stage, *R*=0.90 for the verification stage and *R*=0.83 for the test stage in Network 2. The total R value, which is the other regression coefficient given above, was obtained by introducing the input values of 665 people to the network after the network learning stage, estimating the output values and comparing these values with the actual values in a correlation. Therefore, it is seen that there was a statistical correlation between the Science and Technology Education I course grades produced by the network and the actual grades (*R*=0.84), and between the Science and Technology Education II course grades produced by the network and the actual grades (*R*=0.84). These values can be interpreted as a high correlation between the data produced by the network and the actual data. Based on this, it can be said that the network accomplished a successful modeling. A graphical comparison of the actual grades of the people with the grades predicted by the network is given below (Fig 5 & Fig 6).



Observed Value: Blue Dots, Predicted Value: Red Dots.

Figure 5. Science and Technology Education I Course Observed and Predicted Values



Observed Value: Red Dots, Predicted Value: Blue Dots.

Figure 6. Science and Technology Education II Course Observed and Predicted Values

After the establishment of the network architecture, the achievement grades in the Science and Technology Education I and II courses were estimated by using the input values of the 200 people, which were reserved at the beginning of the study and were not seen by the network before, as well as the existing network structure and connections. These grades that were predicted were compared with the actual data set, and correlations for Network 1 (r=.68, p<0.001) and for Network 2 (r=.69, p<0.001) were obtained. As it is known, there is no general rule about the evaluation of the correlation coefficient, but it is possible to describe the relationship between .68 and 1 as a high or strong relation according to Taylor (1990). This finding is an indication that the constructed artificial neural network learned the achievement grades in the Science and Technology Education I and II courses, which were the output values, using the input values of the people, at a good level.

3.2. Independent Variables Importance Used to Estimate the Achievement in the Science and Technology Education I and II Courses

The final finding obtained within the scope of this study was to determine the independent variables importance used to predict the dependent variables. In this context, the independent variables importance for the Science and Technology Education I course are given in Table 3. When Table 3 is examined, it is seen that the placement score (100%) was the most important variable affecting the dependent variable. The variables with a normalized importance level greater than 50% were Introduction to Educational Science (71.80%), Science and Technology Laboratory Applications I (68.40%), General Physics (58.30%), Instructional Principles and Methods (53.80%), General Chemistry (53.00%) and Educational Psychology (50.70%). However, variables with a importance level below 50% were Science and Technology Laboratory Applications II (48.20%), Gender (45.90%), General Biology (44.90%), Environmental Education (43.40%), Field of Study in High School (40.20%) and Type of Education (35.80%). A bar graph showing the normalized independent variables importance are given below.

Table 3. Normalized Independent Variables Importance for Science and Technology Education I Course

Independent Variable	Importance	Normalized Importance*
Placement Score	.14	100.00%
Introduction to Educational Science	.101	71.80%
Science and Technology Laboratory Applications I	.096	68.40%
General Physics	.082	58.30%
Instructional Principles and Methods	.075	53.80%
General Chemistry	.074	53.00%
Educational Psychology	.071	50.70%
Science and Technology Laboratory Applications II	.067	48.20%
Gender	.064	45.90%
General Biology	.063	44.90%
Environmental Education	.061	43.40%
Field of Study in High School	.056	40.20%
Type of Education	.05	35.80%

* Importance values divided by the largest importance values and expressed as percentages.

As can be seen above, 13 variables, which were defined as the independent variables for the Science and Technology Education I course, were ranked in descending order according to their normalized importance levels.

The independent variables importance for the Science and Technology Education II course are given in Table 4.

Table 4. Normalized Independent Variables Importance for Science and Technology Education II Course

Independent Variable	Importance	Normalized Importance*
Placement Score	.108	100.00%
Instructional Principles and Methods	.1	92.2%
General Biology	.071	65.8%
General Physics	.071	65.8%
Classroom Management	.07	64.2%
Science and Technology Laboratory Applications I	.066	61.00%
Introduction to Educational Science	.061	56.1%
Measurement and Assessment	.056	51.9%
General Chemistry	.055	51.00%
Educational Psychology	.052	48.2%
Science and Technology Education I	.051	47.3%
Instructional Technologies and Material Design	.051	47.00%
Science and Technology Laboratory Applications II	.044	40.2%
Environmental Education	.041	37.6%
Type of Education	.037	34.5%
Field of Study in High School	.035	32.7%
Gender	.03	27.6%

* Importance values divided by the largest importance values and expressed as percentages.

When Table 4 is examined, it is seen that the most important variable affecting the Science and Technology Education II course was the Placement Score (100%), which was also the most important variable that predicted the Science and Technology Education I course. The

variables with a normalized importance level greater than 50% were the achievement grades of the Instructional Principles and Methods (92.2%), General Biology (65.8%), General Physics (65.8%), Classroom Management (64.2%), Science and Technology Laboratory Applications I (61.00%), Introduction to Educational Science (56.1%), Measurement and Assessment (51.9%) and General Chemistry (51.00%) courses. They were followed by Educational Psychology (48.2%), Science and Technology Education I (47.3%), Instructional Technologies and Material Design (47.00%), Science and Technology Laboratory Applications II (40.2%), Environmental Education (37.6%), Type of Education (34.5%), Field of Study in High School (32.7%) and Gender (27.6%).

4. DISCUSSION AND CONCLUSION

When the literature on the use of artificial neural networks in the field of education and training is examined, it is found that even though there are several studies conducted in the international literature (Ibrahim & Rusli, 2007; Karamouzis & Vrettos, 2008; Oladokun, Adebanjo, & Charles-Owaba, 2008), far less attention has been paid on the prediction and classification of student achievement in Turkey. However, it is possible to come across frequent use of regression analysis types (Açıl, 2010; Bahar, 2011; Baştürk, 2008; Doğan & Şahin, 2009; Kablan, 2010; Köstereliоğlu, Köstereliоğlu, & Kilmen, 2008) in prediction studies related to education and training in our country.

Some of the artificial neural network studies conducted in Turkey have been directed towards the prediction of standard tests such as TEOG (Şen, Uçar, & Delen, 2012), KPSS (Demir, 2015), and PISA (Tepehan, 2011) as well as students' course scores (Turhan, Kurt, & Engin, 2013) and general academic scores (Şengür, & Tekin, 2013). In general, two types of benchmarking studies on artificial neural networks are widely available on the basis of education and training. The first of these is the type of studies by Ayık, Özdemir and Yavuz (2007), Ibrahim and Rusli (2007), Şengür (2013), Şengür and Tekin (2013), Tosun (2007), and Vandamme, Meskens and Superby (2007), which are based on a comparison of prediction performances of decision trees and artificial neural networks, including classification analyzes from data mining models. The second is the type of studies by Bahadır (2013), Çırak (2012), Guo (2010), Tepehan (2011), and Turhan, Kurt and Engin (2013), which include the comparison of prediction performances of regression models with those of artificial neural networks. Other than these, there are also studies on the prediction and classification of student achievement using artificial neural networks without performance comparisons (Demir, 2015; Karamouzis & Vrettos, 2008; Kardan, Sadeghi, Ghidary and Sani, 2013; Naser, Zaqout, Ghosh, Atallah, & Alajrami, 2015; Oancea, Dragoescu, & Ciucu, 2013; Oladokun, Adebanjo, & Charles-Owaba, 2008).

However, it is seen that research mainly focuses on studies to compare the performances of artificial neural networks, decision trees and regression models. It can be argued that this is due to the fact that researchers have been engaged in such studies mainly to measure and demonstrate the effectiveness of the statistical methodology. In this study, the main purpose was the prediction of Science and Technology Education course achievements of elementary education teacher candidates. As a result of the study, it is seen that the most important variable in the prediction of Science and Technology Education I and II courses was the university placement scores of the candidates. In a study conducted by Sitturug (1997), it is seen that scientific process skills are the most effective variable in predicting the science achievement of teacher candidates. The main reason for the differences in these two studies is due to the difference in the type and number of independent variables employed in the prediction of science achievement. In addition, the fact that the number of people in the sample of Sitturug's (1997) study was 80 and that the statistical analysis was carried out through regression analysis can be shown as another factor causing this difference between the two studies. Similarly, in a

study conducted by Anıl (2009) in parallel with the main purpose of this current study, the most important variable in predicting students' science achievement in the PISA test was the father's educational status. Moreover, Ceylan and Berberoğlu (2007) concluded that the most important variable in explaining the science achievement of the students participating in the TIMSS test was the students' perceptions of failure. In this study conducted by Ceylan and Berberoğlu (2007), parental education level was not included as an independent variable. In addition, the reasons for the above-mentioned studies to differ from the findings of the current study include the facts that those studies were carried out on students under the age of 15, they employed regression analyses and structural equation modeling, and they were carried out to predict standard test scores such as PISA and TIMSS.

Numerous studies exist in predicting science achievement of teacher candidates as well as students, showing that variables such as attitudes towards science and technology, parental education level, socio-economic level, and scientific process skills are related to science achievement and that they have an important effect in predicting the science achievement of students (Berberoğlu, Çelebi, Özdemir, Uysal, & Yayan, 2003; Fleming & Malone, 1983; Germann, 1994; Schibeci & Riley, 1986)..

Likewise, in studies where artificial neural networks are used to predict academic achievement, variables that are expected to influence the success of the course or general academic achievement have been incorporated into models by researchers as independent variables by creating different network architectures. This also leads to differentiation of research results. The biggest cause of these differences is the fact that different types of input variables are chosen by researchers as mentioned above. In addition, the number of layers, number of neurons and learning function are increased when creating the best model with artificial neural networks. These increase the number of combinations of trials. This excess number of combinations allows researchers to create a large number of network architectures in their studies. This leads to different analyses and different outcomes in similar studies, as it is the case in the prediction of academic achievement. However, it is seen that the lowest classification prediction and the highest network performance vary between 51.88% and 91.77%, respectively, in the findings of studies (such as linear regression, logistic regression and decision tree) on the comparison of the predictions and the statistical methods with respect to artificial neural networks related to education and training (Bahadır, 2013; Çırak, 2012; Gülçin, Çırak, & Çokluk, 2013; Demir, 2015; Guo, 2010; Ibrahim & Rusli, 2007; Karamouzis & Vrettos, 2008; Kardan, Sadeghi, Ghidary, & Sani, 2013; Moridis & Economides, 2009; Naser, Zaqout, Ghosh, Atallah, & Alajrami, 2015; Oancea, Dragoescu, & Ciucu, 2013; Oladokun, Adebanjo, & Charles-Owaba, 2008; Paliwal & Kumar, 2009; Romero, Ventura, & García, 2008; Rusli, Ibrahim, & Janor, 2008; Şen, Uçar, & Delen, 2012; Şengür, 2013; Şengür & Tekin, 2013; Tepehan, 2011; Tosun, 2007; Turhan, Kurt, & Engin, 2013; Vandamme, Meskens, & Superby, 2007).

The fact that the artificial neural network performance in this study was R=0.84 for the Science and Technology Education I course, and R=0.84 for the Science and Technology Education II course shows that the network performance overlaps with the findings obtained from the above studies. In the prediction of academic achievement, both in the context of this study and in the studies mentioned above, the results obtained using artificial neural networks show that the prediction results can be considered quite good specifically for social sciences.

Acknowledgements

This research was produced from doctoral thesis titled "Modeling of Science and Technology Teaching Course Achievements of Elementary Teacher Candidates with Artificial Neural Networks" by Ergün Akgün in supervising of Metin Demir.

ORCID

Ergün Akgün https://orcid.org/0000-0002-7271-6900

5. REFERENCES

Açıl, Ü. (2010). *Öğretmen adaylarının akademik başarıları ile KPSS puanları arasındaki ilişkinin çeşitli değişkenler açısından incelenmesi [Examination of the relationship between academic success of teacher candidates and civil servant selection examination (KPSS) scores, in terms of different variables (Master's Thesis)]*. Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü, Hatay.

Adıyaman, F. (2007). *Talep tahmininde yapay sinir ağlarının kullanılması [Sales forecasting using artifical neural networks (Master's Thesis)]*. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications, 18*(2), 65-74. doi:10.1016/s0957-4174(99)00053-6

Alam, M., & Briggs, A. (2011). Artificial neural network meta-models in cost-effectiveness analysis of intensive blood-glucose control: A case study applied to the UK prospective diabetes study (Ukpds) individual patient outcome simulation model. *Value in Health, 14*(7), 234-235.

Anemone, R., Emerson, C., & Conroy, G. (2011). Finding fossils in new ways: An artificial neural network approach to predicting the location of productive fossil localities. *Evol Anthropol, 20*(5), 169-180. doi:10.1002/evan.20324

Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler [Factors effecting science achievement of science students in programme for international students' achievement (PISA) in Turkey]. *Eğitim ve Bilim, 34*(152), 87-100.

Arbib, M. A. (1987). *Brains, machines, and mathematics.* New York: Springer-Verlag.

Arbib, M. A. (2003). *The handbook of brain theory and neural networks.* Cambridge, Massachusetts, London, England: MIT press.

Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2007). Lise türü ve lise mezuniyet başarısının kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi [An analysis of the correlation between the type of high school and high school graduation score and university placement examination using data mining technique.]. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10*(2), 441-454.

Azadeh, A., Saberi, M., & Anvari, M. (2011). An integrated artificial neural network fuzzy C-means-normalization algorithm for performance assessment of decision-making units: The cases of auto industry and power plant. *Computers & Industrial Engineering, 60*(2), 328-340. doi:10.1016/j.cie.2010.11.016

Bahadır, E. (2013). *Yapay sinir ağları ve lojistik regresyon analizi yaklaşımları ile öğretmen adaylarının akademik başarılarının tahmini [Prediction of student teachers' academic success with logistic regression analysis and artificial neural networks methods (Doctoral Thesis)]*. Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul.

Bahar, H. H. (2011). ÖSS puanı ve lisans mezuniyet notunun KPSS 10 puanını yordama gücü [KPSS 10 score prediction power of bachelor graduation mark and OSS Score]. *Education and Science, 36*(162), 168-181.

Baştürk, R. (2008). Fen ve teknoloji alanı öğretmen adaylarının kamu personeli seçme sınavı başarılarının yordanması [Predictive validity of the science and technology pre-service teachers' civil servant selection examination]. *İlköğretim Online, 7*, 323-332.

Berberoğlu, G., Çelebi, Ö., Özdemir, E., Uysal, E. ve Yayan, B. (2003). Üçüncü uluslararası matematik ve fen çalışmasında Türk öğrencilerinin başarı düzeylerini etkileyen etmenler [Factors effecting achievement level of Turkish students in the third international mathematics and science study (TIMSS)]. *Eğitim Bilimleri ve Uygulama, 2*(3), 3-14.

Cai, M., Yin, Y. F., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D-Transport and Environment, 14*(1), 32-41.

Ceylan, E. ve Berberoğlu, G. (2007). Öğrencilerin fen başarısını açıklayan etmenler: Bir modelleme çalışması [Factors related with students' science achievement: A modeling study]. *Eğitim ve Bilim, 32*(144), 36-48.

Cohen, I. L. (1994). An artificial neural network analogue of learning in autism. *Biological Psychiatry, 36*(1), 5-20. doi:10.1016/0006-3223(94)90057-4

Çırak, G. (2012). *Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması [The usage of artifical neural network and logistic regression methods in the classification of student achievement at higher education (Master's Thesis)]*. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Çırak, G. ve Çokluk, Ö. (2013). Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması [The usage of artifical neural network and logistic regresssion methods in the classification of student achievement in higher education]. *Mediterranean Journal of Humanities, 3*(2), 71-79. doi:10.13114/mjh/201322471

Çuhadar, M. (2006). *Turizm sektöründe talep tahmini için yapay sinir ağları kullanımı ve diğer yöntemlerle karşılaştırmalı analizi (Antalya İli'nin dış turizm talebinde uygulama) [Forecasting tourism demand by artificial neural networks and time series methods (A comparative analysis in inbound tourism demand to Antalya) (Doctoral Thesis)]* Süleymen Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta.

de Haen, C. (2009). Knowledge culture conditioned bounded rationality and human artificial neural network processes (HANNP decision theory): How risks of accidents and environmental impact of a new chemical production process and plant site have entered decisions. *Safety Science, 47*(6), 843-852.

Demir, M. (2015). Predicting pre-service classroom teachers' civil servant recruitment examination's educational sciences test scores using artificial neural networks. *Educational Sciences: Theory & Practice, 5*, 1169-1177.

Doğan, N., & Şahin, A. E. (2009). The variables predicting the appointment of candidate teachers to primary schools. *Inonu University Journal of the Faculty of Education, 10*, 183-199.

Dowman, R., & Ben-Avraham, D. (2008). An artificial neural network model of orienting attention toward threatening somatosensory stimuli. *Psychophysiology, 45*(2), 229-239. doi:10.1111/j.1469-8986.2007.00614.x

Emir, Ş. (2013). *Yapay sinir ağları ve destek vektör makineleri yöntemlerinin sınıflandırma performanslarının karşılaştırılması: Borsa endeks yönünün tahmini üzerine bir uygulama [Classification performance comparison of artificial neural networks andsupport vector machines methods: An empirical study on predicting stockmarket index movement direction (Doctoral Thesis)]*. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Fleming, M. L., & Malone, M. R. (1983). The relationship of student characteristics and student performance in science as viewed by meta-analysis research. *Journal of Research in Science Teaching, 20*, 481-495.

Foody, G. M. (1995). Land cover classification by an artificial neural network with ancillary information. *International Journal of Geographical Information Systems, 9*(5), 527-542. doi:10.1080/02693799508902054

Geem, Z. W. (2010). Response to "Comments on 'Energy demand estimation of South Korea using artificial neural network' by Zong Woo Geem and William E. Roper" by Serhat Kucukali. *Energy Policy, 38*(10), 6381-6382. doi:10.1016/j.enpol.2009.08.062

Germann, P. J. (1994). Testing a model of science process skills acquisition: An interaction with parents' education, preferred language, gender, science attitude, cognitive development, academic ability, and biology knowledge. *Journal of Research in Science Teaching, 31*(7), 749-783.

Ghiassi, M., Saidane, H., & Zimbra, D. K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting, 21*(2), 341-362. doi:10.1016/j.ijforecast.2004.10.008

Gonzalez, J. M. B., & DesJardins, S. L. (2002). Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education, 43*(2), 235-258. doi:Doi 10.1023/A:1014423925000

Graupe, D. (2007). Principles of artificial neural networks. *Advenced Series on Circuits and Systems, 6*.

Guo, W. W. (2010). Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. *Expert Systems with Applications, 37*(4), 3358-3365. doi:10.1016/j.eswa.2009.10.014

Gupta, L., Molfese, D. L., & Tammana, R. (1995). An artificial neural network approach to ERP classification. *Brain Cogn, 27*(3), 311-330. doi:10.1006/brcg.1995.1025

Gülçe, G. (2010). *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma [Developing decision support system for student information system by using data warehouse and data mining technics (Master's Thesis)]*. Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli.

Haykin, S. (2009). *Neural networks and learning machines.* Prentice Hall. New Jersey

Heaton, J. (2008). *Introduction to neural networks for C #.* Heaton Research, Inc.

Ibrahim, Z., & Rusli, D. (2007). *Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression.* Paper presented at the 21st Annual SAS Malaysia Forum, 5th September.

Kablan, Z. (2010). Akademik mezuniyet ortalama puanı ile KPSS başarı puanı arasındaki ilişki [Relationship between undergraduate grade point average and student performance on civil cervant selection exam (KPSS)]. *Uludağ Eğitim Fakültesi Dergisi, 23*, 451-470.

Karamouzis, S. T., & Vrettos, A. (2008). *An artificial neural network for predicting student graduation outcomes.* Paper presented at the Proceedings of the World Congress on Engineering and Computer Science.

Kardan, A. A., Sadeghi, H., Ghidary, S. S., & Sani, M. R. F. (2013). Prediction of student course selection in online higher education institutes using neural network. *Computers & Education, 65*, 1-11. doi:10.1016/j.compedu.2013.01.015

Kayıkçı, Ş. (2014). *Web sayfalarının yapay sinir ağları ile sınıflandırılması [Classification of web pages using neural networks (Doctoral Thesis)]*. Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Koç, M. L., Balas, C. E. ve Arslan, A. (2004). Taş dolgu dalgakıranların yapay sinir ağları ile ön tasarımı [Preliminary design of ruble mound breakwaters by using artificial neural networks]. *İMO Teknik Dergi, 15*(4), 3351-3375.

Köstereliođlu, İ., Köstereliođlu, M. ve Kilmen, S. (2008). Kamu personeli seçme sınavı (KPSS) puanları ile lisans diploma notu arasındaki ilişkilerin çeşitli değişkenlere göre incelenmesi [A study of the relationship between the civil servant selection examination and undergraduate cumulative general point averages according to a number of variables]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 8*, 151-160.

Kuyucu, Y. E. (2012). *Lojistik regresyon analizi (LRA), yapay sinir ağları (YSA) ve sınıflandırma ve regresyon ağaçları (C&RT) yöntemlerinin karşılaştırılması ve tıp alanında bir uygulama [Comparison of logistic regression analysis (LRA), artificial neural networks (ANN) and classfication and regression trees (C&RT) methods and an aplication in medicine (Master's Thesis)]*. Gaziosmanpaşa Üniversitesi Sağlık Bilimleri Enstitüsü, Tokat.

MATLAB and Neural Network Toolbox Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115-133.

Moridis, C. N., & Economides, A. A. (2009). Prediction of student's mood during an online test using formula-based and neural network-based method. *Computers & Education, 53*(3), 644-652. doi:10.1016/j.compedu.2009.04.002

Naser, S. A., Zaqout, I., Ghosh, M. A., Atallah, R., & Alajrami, E. (2015). Predicting student performance using artificial neural network: In the faculty of engineering and information technology. *International Journal of Hybrid Information Technology, 8*(2), 221-228.

Nussbaum, M. A., & Chaffin, N. B. (1996). Evaluation of artificial neural network modelling to predict torso muscle activity. *Ergonomics, 39*(12), 1430-1444. doi:10.1080/00140139608964562

Oancea, B., Dragoescu, R., & Ciucu, S. (2013). *Predicting students' results in higher education using neural networks.* Paper presented at the The 6-th Conference Applied Information and Communication Technology, Faculty of Information Technology, Latvia University of Agriculture, Jelgava, Latvia.

Oladokun, V., Adebanjo, A., & Charles-Owaba, O. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology, 9*(1), 72-79.

Paliwal, M., & Kumar, U. A. (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications, 36*(4), 7865-7872. doi:10.1016/j.eswa.2008.11.003

Palmer, A., José Montaño, J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781-790. doi:10.1016/j.tourman.2005.05.006

Polanco, X., Francois, C., & Keim, J. P. (1998). Artificial neural network technology for the classification and cartography of scientific and technical information. *Scientometrics, 41*(1-2), 69-82.

Raju, K. A., Sikdar, P. K., & Dhingra, S. L. (1996). Modelling mode choice by means of an artificial neural network. *Environment and Planning B-Planning & Design, 23*(6), 677-683.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384. doi:10.1016/j.compedu.2007.05.016

Rusli, N. M., Ibrahim, Z., & Janor, R. M. (2008, 26-28 Aug. 2008). *Predicting students' academic achievement: Comparison between logistic regression, artificial neural network, and Neuro-fuzzy.* Paper presented at the Information Technology, 2008. ITSim 2008. International Symposium on.

Schibeci, R., & Riley, J. (1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching, 23*, 177-187.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications, 36*(2), 2843-2852.

Sim, J. J., Tan, G. W. H., Wong, J. C. J., Ooi, K. B., & Hew, T. S. (2014). Understanding and predicting the motivators of mobile music acceptance - A multi-stage MRA-artificial neural network approach. *Telematics and Informatics, 31*(4), 569-584.

Sittirug, H. (1997). *The predictive value of science process skills, attitude toward science, and cognitive development on achievement in a Thai teacher institution*. Unpublished PhD thesis, University of Missouri, Columbia.

Snyder, M. R. (1998). A functionally equivalent artificial neural network model of the prey orientation behavior of waterstriders (Gerridae). *Ethology, 104*(4), 285-297.

Stevens, R., Ikeda, J., Casillas, A., Palacio-Cayetano, J., & Clyman, S. (1999). Artificial neural network-based performance assessments. *Computers in Human Behavior, 15*(3-4), 295-313.

Suzuki, Y. (2001). An artificial neural network test for structural change with unspecified parametric form. *Japanese Economic Review, 52*(3), 339-365.

Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications, 39*(10), 9468-9476. doi:10.1016/j.eswa.2012.02.112

Şengür, D. (2013). *Öğrencilerin akademik başarılarının veri madenciliği metotları ile tahmini [Prediction of student's academic achievements by using the data mining methods (Master's Thesis)]*. Fırat Üniversitesi Eğitim Bilimleri Enstitüsü, Elazığ.

Şengür, D. ve Tekin, A. (2013). Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini [Prediction of student's grade point average by using the data mining methods]. *Bilişim Teknolojileri Dergisi, 6*(3), 7-16.

Tang, Y. C. (2009). An approach to budget allocation for an aerospace company-Fuzzy analytic hierarchy process and artificial neural network. *Neurocomputing, 72*(16-18), 3477-3489.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6*(1), 35-39.

Tepehan, T. (2011). *Türk öğrencilerinin PISA başarılarının yordanmasında yapay sinir ağı ve lojistik regresyon modeli performanslarının karşılaştırılması [Performance comparison of artificial neural network and logistic regression model in predicting Turkish students' PISA success (Doctoral Thesis)]*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Tosun, S. (2007). *Sınıflandırmada yapay sinir ağları ve karar ağaçları karşılaştırması: Öğrenci başarıları üzerine bir uygulama [Artificial neural networks-decision tree comparison in classification analysis: An application on students' success (Master's Thesis)]*. İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Estimation of student success with artificial neural networks. *Education and Science, 38*(170), 112-120.

Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15*(4), 405-419. doi:10.1080/09645290701409939

Verhagen, J. V., & Scott, T. R. (2004). Artificial neural network analysis of gustatory responses in the thalamic taste relay of the rat. *Physiology & Behavior, 80*(4), 499-513.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition, 13*(3), 565-578. doi:10.1016/j.concog.2004.05.008

Wei, C. H., & Lee, Y. (2007). Sequential forecast of incident duration using artificial neural network models. *Accident Analysis and Prevention, 39*(5), 944-954.

Wyatt, R. (1996). Evaluating strategies by means of an artificial neural network. *Environment and Planning B-Planning & Design, 23*(6), 685-695.

Yavuz, S., & Deveci, M. (2012). İstatistiksel normalizasyon tekniklerinin yapay sinir ağın performansına etkisi [The effect of statistical normalization techniques on the performance of artificial neural network]. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 40*, 167-187.

Yurdakul, E. M. (2014). *Türkiye'de ithalatın gelişimi ve ithalatın yapay sinir ağları yöntemi ile tahmin edilebilirliğine yönelik bir analiz [The evolution of import in Turkey and an analysis abaut predictability of import by artificial neural network method (Doctoral Thesis)]*. Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü, Aydın.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 510–523

DOI: 10.21449/ijate.443222

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Functional Status of Women and Their Partners after Childbirth

Sevgi Özkan [1,*], Sinem Göral Türkcü [1], Özlem Kayhan [1], Pınar Serçekuş Ak [1], Ümran Sevil [2]

[1] Pamukkale University, Faculty of Health Sciences, Department of Obstetrics and Gynecology Nursing, Denizli, Turkey
[2] Ege University, Nursing Faculty, Department of Obstetrics and Gynecology Nursing, İzmir, Turkey

Abstract: The aim of this study is to define the changes within the social dimension of the functional status of women and their partners in the months following the birth of their child. A total of 90 participants were included in the study (45 women and 45 partners). Data were collected in two steps via face-to-face interviews. The first step involved administration of a personal information form and the functional status inventories to the women and their partners between the postpartum sixth and eighth weeks. The second step involved re-administration of the same functional status inventories to the women and their partners in the postpartum sixth month. Results from the examination of the functional status of the women and their partners in the postpartum second and sixth months showed that there was a significant increase in the areas of household activities (p=0,000), social and community activities (p=0,000) and self-care activities (p=0,000) for the women in the sixth month; similarly, for the fathers, household activities (p=0,008), social and community activities (p=0,003) and child care activities (p=0,007) rose in the sixth month. Furthermore, the study found that certain sociodemographic variables had an effect on functional status (p=0,000). It is important that nurses and/or midwives support parents during the pregnancy process, particularly during the transition into their new roles as women and their partners, and help prepare them for the changes to occur as they enter into this new period of life.

ARTICLE HISTORY

Received: 27 April 2018
Revised: 19 June 2018
Accepted: 12 July 2018

KEYWORDS

Functional Status,
Transition to Motherhood,
Transition to Fatherhood

1. INTRODUCTION

In the postpartum period, much of the focus in the studies conducted has tended to center on recovery of the reproductive organs. On the other hand, some topics such as the process of transition to parenthood, increased responsibilities, tiredness, changes in the relationship with the spouses, pre-pregnancy social and professional activities are not considered enough (Fichardt, Van Wyk & Weich, 1994; Gjerdingenm & Chaloner, 1994; Herbert, 1998; Hodnett, 1996). The functional status of women are required to "take on the responsibility of infant care,

CONTACT: Sevgi Özkan ✉ sozkan@pau.edu.tr Pamukkale University, Faculty of Health Sciences, Campus Kınıklı, 20070, Denizli/ Turkey.

ISSN-e: 2148-7456 /© IJATE 2018

self-care, household activities, and social, community, and occupational activities" (Fawcett, Tulman & Myers, 1988). The functional status of fathers, on the other hand, either stays the same or expands to involve more responsibilities, including those related to social, community, occupational, and educational activities, and child and personal care. Furthermore, partners take on the responsibility of taking care of other children within the family during pregnancy and after the childbirth (McVeigh, 2001; Tulman, Fawcett & Weiss, 1993).

In some studies, parents have described the postpartum period as a process that generates major problems for both themselves and their family life. Moreover, research has shown that it requires women a longer amount of time to return to their antenatal functional status than to recover physiologically (Fawcett et al., 1988; Fichardt et al., 1994). Identifying precisely how the functional status of women changes in the postpartum period, as well as the factors that are brought to bear on these changes, is therefore key. During the postpartum period in particular, which is marked by an increasing number of difficulties, it is necessary to focus on solutions, not only to the physical problems women experience, but also to their social and psychological conditions, as these tend to take a longer time to overcome (Apay, Ejder & Pasinlioğlu, 2009; Beji, Coşkun & Yıldırım, 2003; Özkan & Sevil, 2007).

The men's transition to fatherhood is as complicated as the woman's transition to motherhood. The role of fatherhood has been shown to start at the diagnosis of pregnancy, continuing up to the birth of the baby and evolving during the months following the birth (Barclay & Lupon, 1999; Başbakkal, 1999; Sevil & Özkan, 2010). This period requires that the father takes on more responsibility, remains committed, and allocates sufficient time for the infant. In this period of transition, men experience a number of changes in their lifestyle, including having to face the facts about pregnancy and child birth, making efforts to accept fatherhood, and adapting to a new role (Anderson, 1996). "Expectant" fathers may experience anxiety, conflicting emotions, weakness and happiness, all of which enable them to develop a new set of skills necessary for the roles related to fatherhood (Barclay, Donovan, & Genovese, 1996; Henderson & Brouse, 1991). The functional status of fathers is a multidimensional process, and the transition to motherhood and fatherhood can be a complicated period for both women and their partners. In this process, men are expected to not only fulfil their previous duties but also to get accustomed to their new and challenging role of being father (Fawcett et al., 1988; McVeigh, 2000a; McVeigh, 2000b; Sevil & Özkan, 2010; Tulman et al., 1993).

Most studies have largely concentrated on the womens' functional status before and after the birth. There is not enough information about the functional status of both women and their partners in the postpartum period (Apay et al., 2009; Beji et al., 2003; Fichardt et al., 1994; McVeigh & Chaboyer, 2002; Özkan & Sevil, 2007; Posmontier, 2008; Şanlı & Öncel, 2014; Tulman, Fawcett, Groblewski & Silverman, 1990). For this reason, this study has aimed to determine the changes that occur within the social sphere, in terms of the postpartum functional status and recoveries of women and their partners, in the months following the birth, and also to determine the correlation between socio-demographic factors and postpartum functional status.

2. MATERIALS AND METHODS

2.1. Data and test subject selection

This descriptive study was conducted in order to identify the postpartum functional status of women and their partners. This study was conducted in three Family Health Centers which were low, moderate, and high socio-economic standings. The Family Health Centers were selected from 51 Family Health Centers located in the city center of Denizli, Turkey. The study involved women and their partners in the postpartum sixth to eighth weeks. Using a sampling size calculation formula, applicable for when the test subject is known, the number of

participants was calculated to be 41, at a confidence interval of 95% and a sampling error of 5%; however, to preempt possible loss of participants, the study started out with 58 women and their partners. The study ended up being completed with 45 women and 45 partners (in total 90 participants).

2.2. Inclusion criteria

To participate in the study, the women had to have had a full-term vaginal delivery or caesarean section and be in the postpartum sixth to eighth weeks, while both mother and father had to have no mental retardation, chronic illness or disability, be able to speak and understand Turkish, and agree to participate in the study. All the infants in the study were full term without any health problems.

2.3. Data collection tool

The tools used in the study included a personal information form, which was developed based on corresponding literature, to determine the socio-demographic characteristics of the test subject; the Inventory of Functional Status After Childbirth (IFSAC), to evaluate the postpartum functional status of women; and the Inventory of Functional Status-Fathers (IFS-F), to evaluate the functional status of fathers. Data collection involved a two-step process, which was conducted through face-to-face interviews. The first step included administration of the personal information form and the functional status inventories (IFSAC, IFS-F) to the women and their partners in the postpartum sixth to eighth weeks. The second step included re-administration of the same functional status inventories to the women and their partners in the postpartum sixth month.

2.3.1. Personal information form

Prepared on the basis of the corresponding literature, the questionnaire that ten questions included features questions about the socio-demographic characteristics and income status of participants, as well as the number of pregnancies and number of living children, delivery method, method of feeding their children, occupation, and the state of the support they give in infant care and household activities (Beji et al., 2003; Fawcett et al., 1988; McVeigh, 2000a; McVeigh, 2000b; Özkan & Sevil, 2007).

2.3.2. IFSAC: Inventory of functional status after childbirth

The Inventory of Functional Status After Childbirth (IFSAC), which is based on the role function model of the Roy Adaptation Model, is used to measure functional status (Fawcett et al., 1988; Fawcett & Tulman, 1990). The IFSAC questionnaire allows for the assessment of primary, secondary and tertiary roles in the postpartum period. There are five dimensional subscales of the inventory: infant care, self-care, household activities, social, community activities, and occupational activities. The validity and reliability were conducted by Özkan and Sevil (2007) in Turkey. The questionnaire consists of 36 items. Higher mean scores indicate greater functional status.

2.3.3. IFS-F: The Inventory of functional status-fathers

Developed to evaluate the functional status of fathers during the last trimester of pregnancy and the postpartum period (Tulman et al., 1993), this inventory was created on the basis of the role function model of the Roy Adaptation Model and features seven subscales: occupational activities, household activities, personal care activities, social and community activities, child care responsibilities, educational activities and baby care activities The validity and reliability were conducted by Sevil and Özkan (2010) in Turkey. The questionnaire consists of 51 items. Higher scores indicate a high functional status (McVeigh, 2001; Sevil & Özkan, 2010; Tulman et al., 1993).

2.4. Ethical aspect of the study

In order to conduct the study, permission was obtained from Denizli Provincial Directorate of Health and Non-Invasive Researches Ethics Committee of Pamukkale University. Also permission was obtained from the institution where the research was conducted. The women and their partners to be included in the study were informed about the purpose of the study and verbal and written consents were received from those who agreed to participate in the study.

2.5. Data Analysis

The data were analyzed using number values, percentage, mean, paired t-test, Mann-Whitney U Test, Wilcoxon Signed Ranks Test, and Kruskal Wallis analyses. Statistical significance was defined as $p<0.05$.

3. FINDINGS

Table 1 shows the socio-demographic data of the women and their partners who participated in the study. In terms of the functional status of the women, a significant increase was found in the mean scores of household activities ($p=0.000$), social and community activities ($p=0.000$) and self-care activities ($p=0.000$) in the postpartum sixth month. No significant difference, however, was determined between the months studied in terms of the mean scores obtained in the areas of infant care activities and occupational activities ($p>0.05$) (Table 2).

Table 1. Socio-demographic characteristics of women and their partners

Socio-Demographic Characteristics	Mother	Father
	n(%)	*n*(%)
Residential area		
Low	15(33.3)	15(33.3)
Moderate	15(33.3)	15(33.3)
High	15(33.3)	15(33.3)
Age		
18-23 years	4(8.9)	0(0)
24-29 years	17(37.8)	9(20.0)
30-35 years	16(35.6)	21(46.7)
35 years and above	8(17.7)	15(33.3)
Educational Background		
Illiterate	1(2.2)	0(0)
Primary school graduate	12(26.7)	14(31.1)
Secondary school graduate	4(8.9)	5(11.1)
High school graduate	10(22.2)	9(20.0)
Associate degree//Bachelor's degree	18(40.0)	17(37.8)
Occupation		
Unemployed	24(53.3)	1(2.2)
Civil servant	6(13.3)	4(8.9)
Worker	5(11.1)	17(37.8)
Self-employed	4(8.9)	15(33.3)
Other	6(13.4)	8(17.8)
Economic condition		
Income higher than expense	11(24.4)	10(22.2)
Income equal to expense	26(57.8)	22(48.9)
Income lower than expense	8(17.8)	13(28.9)

Social security		
Yes	43(95.6)	42(93.3)
No	2(4.4)	3(6.7)
State of intending pregnancy		
Intended and planned	36(80.0)	36(80.0)
Intended but unplanned	9(20.0)	9(20.0)
Common characteristics of Mother-Father	n(%)	
Method of feeding the infant		
Breast milk	35(77.8)	
Formula	1(2.2)	
Breast milk and formula	9(20.0)	
Contribution to infant care from the outside		
Yes	25(55.6)	
No	20(44.4)	
Contribution to infant care (*n*=25)		
Maternal/Grandmother-Paternal Grandmother	16(35.6)	
Father	6(13.3)	
Nanny	3(6.7)	
Contribution to household chores from the outside		
Yes	24(53.3)	
No	21(46.7)	
Contribution to household chores (*n*=24)		
Maternal Grandmother/Paternal Grandmother	12(26.7)	
Father	9(20.0)	
Nanny	3(6.7)	
Duration of marriage		
1-3 years	17(37.7)	
4-6 years	11(24.4)	
7-10 years	10(22.2)	
10 years and more	7(15.7)	
Delivery method		
Normal birth	11(24.4)	
Caesarean section	34(75.6)	
Number of pregnancies		
1	16(35.6)	
2 and more	29(64.4)	
Live birth		
1	17(37.8)	
2 and more	28(62.2)	

Table 2. Functional status mean scores of women according to months

Subscales	2nd month (*n* = 45)	6th month (*n* = 45)	*P*
	Mean±*SD*	Mean±*SD*	
Household activities	2.59±0.49	3.02±0.44	**.000***
Social and community activities	1.81±0.39	2.13±0.51	**.000***
Infant care activities	3.75±0.26	3.83±0.24	.113*
Self-care activities	2.70±0.43	3.08±0.28	**.000***
Occupational activities (*n* = 2)	2.40±0.00	2.86±0.43	.180**

*Paired t Test ** Wilcoxon Signed Ranks Test

The examination of the fathers' functional status in the postpartum second and sixth months showed there to be a significant increase in the mean scores of household activities (p=0.008), social and community activities (p=0.003) and child care activities (p=0.007) in the sixth month, while no significant difference was found in terms of the months studied in the mean scores of infant care, self-care, and occupational and educational activities (p>0.05) (Table 3).

The effect of socio-demographic variables on the functional status of fathers was studied and analyzed as well. Fathers with a higher educational level had statistically significant mean scores in the area of infant care responsibilities in the postpartum eighth week (p=0.024), and in the area of child care activities in the postpartum eighth week and the postpartum sixth month (p=0.023; p=0.030). No significant difference was detected between the other subscales and educational background (p>0.05). Likewise, no significant difference was found between age, economic condition, number of children and the functional status in the postpartum second and sixth months (p>0.05) (Table 4).

In examining the effect of socio-demographic variables on the functional status of women, it was discovered that women with a higher educational level had higher mean scores in the area of household activities in the postpartum sixth month (p=0.043). Furthermore, women who had a caesarean section were found to have stable mean scores in the area of household activities in the postpartum eighth week and higher mean scores in the postpartum sixth month, as compared to women's who gave normal birth (0.042). The other socio-demographic variables showed no significant influence on the women' functional status (p<0.05) (Table 5a, 5b).

Table 3. Functional status mean scores of partners according to months

Subscales	2nd month (*n* = 45)	6th month (*n* = 45)	*p*
	Mean±*SD*	Mean±*SD*	
Household activities	2.34±0.94	2.75±0.74	**.008***
Social and community activities	2.08±0.66	2.47±0.61	**.003***
Infant care activities	2.24±0.96	2.51±0.73	.055*
Child care activities (*n* = 29)	2.51±0.84	2.80±0.72	**.007****
Self-care activities	2.73±0.53	2.60±0.38	.200*
Occupational activities	3.00±0.40	3.03±0.20	.697*
Educational activities (*n* = 4)	2.31±0.81	3.00±0.00	.180**

*Paired t Test ** Wilcoxon Signed Ranks Test

Table 4. The effect of some socio-demographic variables on the functional status of fathers

Variables	Household responsibilities		Social and community activities		Infant care responsibilities		Personal care activities		Occupational activities		Child care activities		Educational activities
	2nd month	6th month	2nd month	6th month	2nd month	6th month	2nd month	6th month	2nd month	6th month	2nd month	6th month	2nd month
Age *													
29 year and younger (n:9)	2.26±0.78	2.86±0.75	1.95±0.68	2.31±0.74	2.37±1.08	2.44±0.73	2.74±0.67	2.61±0.47	3.00±0.50	3.00±0.10	1.95±0.06	2.40±1.09	
30 year and over (n:36)	2.36±0.98	2.72±0.74	2.12±0.66	2.51±0.58	2.21±094	2.53±0.74	2.72±0.50	2.60±0.37	3.00±0.38	3.03±0.22	2.55±0.86	2.83±0.71	
p value	0.875	0.809	0.305	0.359	0.694	0.875	0.919	0.919	0.546	0.755	0.227	0.630	
Education*													
Secondary school and below (n:19)	2.20±0.92	2.59±0.76	1.96±0.53	2.32±0.64	1.87±0.97	2.34±0.77	2.78±0.45	2.51±0.43	3.03±0.34	2.97±0.14	2.19±0.86	2.50±0.76	2.32±0.93
High school and above (n:26)	2.44±0.95	2.87±0.72	2.17±0.74	2.57±0.59	2.51±0.87	2.64±0.69	2.69±0.59	2.67±0.34	2.98±0.45	3.06±0.23	2.85±0.71	3.12±0.54	2.30±0.77
p value	0.474	0.294	0.309	0.116	**0.024**	0.084	0.769	0.264	0.666	0.130	**0.023**	**0.030**	0.607
Economic condition**													
Income higher than expense (n:11)	2.68±1.18	2.81±0.76	1.98±0.64	2.28±0.65	2.53±0.72	2.76±052	2.60±063	2.54±0.47	2.82±0.43	2.96±0.15	2.68±0.82	3.00±0.71	2.20±0.91
Income equal to expense (n:26)	2.12±0.76	2.81±0.79	2.15±0.64	2.60±0.57	2.22±0.90	2.53±0.79	2.77±0.47	2.64±0.33	3.01±0.44	3.08±0.22	2.21±0.69	2.59±0.67	2.33±0.83
Income lower than expense (n:8)	2.44±0.98	2.60±0.67	2.06±0.74	2.40±0.65	2.06±1.21	2.29±0.75	2.75±0.58	2.58±0.42	3.12±0.27	3.00±0.20	2.74±0.98	2.91±0.79	2.40±1.03
p value	0.447	0.737	0.876	0.334	0.508	0.275	0.929	0.794	0.199	0.437	0.249	0.385	0.853
Number of children*													
1 (n:17)	2.44±0.96	2.90±0.85	2.14±0.70	2.50±0.65	2.55±0.83	2.56±0.75	2.75±0.54	2.72±0.36	2.89±0.49	3.07±0.23			
2 and more (n:28)	2.28±0.93	2.66±0.67	2.05±0.64	2.45±0.60	2.05±1.00	2.48±0.73	2.71±0.54	2.53±0.38	3.07±0.33	3.00±0.19			
p value	0.548	0.459	0.933	0.733	0.090	0.546	0.783	0.213	0.113	0.449			

Note: 9 of the fathers who participated in the study started education and 4 continued. As no father continued education in the 2nd and 6th months, the section of educational activities was not analyzed.

*Mann Whitney U Test **Kruskal Wallis Test

Table 5a. The effect of some socio-demographic variables on the functional status of mothers

Variables	Household responsibilities		Social and community activities		Infant care responsibilities		Personal care activities		Occupational activities
	2nd month	6th month	2nd month	6th month	2nd month	6th month	2nd month	6th month	6th month
Age *									
29 year and younger (n:9)	2.55±0.57	3.00±0.35	1.86±0.44	2.06±0.47	3.74±0.28	3.87±0.18	2.71±0.34	3.11±0.32	2.80±0.60
30 year and over (n:36)	2.62±0.42	3.03±0.52	1.76±0.35	2.19±0.54	3.76±0.25	3.80±0.28	2.69±0.50	3.06±0.25	2.88±0.39
p value	0.882	0.954	0.514	0.674	0.916	0.642	0.991	0.720	0.833
Education*									
Secondary school and below (n:17)	2.70±0.54	2.84±0.44	1.77±0.33	2.00±0.32	3.71±0.26	3.81±0.28	2.71±0.42	3.11±0.29	-
High school and above (n:28)	2.52±0.46	3.12±0.42	1.83±0.43	2.21±0.59	3.77±0.26	3.85±0.22	2.70±0.44	3.06±0.78	2.86±0.45
p value	0.385	**0.043**	0.728	0.166	0.339	0.737	0.787	0.739	0.859
Economic condition**									
Income higher than expense (n:11)	2.57±0.44	3.14±0.47	1.89±0.51	2.14±0.59	3.66±0.35	3.77±0.27	2.71±0.44	3.07±0.30	2.84±0.45
Income equal to expense (n:26)	2.54±0.45	3.07±0.40	1.81±0.37	2.14±0.55	3.78±0.24	3.85±0.19	2.75±0.39	3.08±0.30	2.88±0.46
Income lower than expense (n:8)	2.49±0.68	2.66±0.39	1.70±0.28	2.07±0.18	3.79±0.14	3.85±0.35	2.53±054	3.10±0.21	
p value	0.615	0.051	0.685	0.919	0.626	0.408	0.542	0.988	0.915
Number of children*									
1 (n:17)	2.50±0.47	3.11±0.36	1.78±0.46	2.10±0.45	3.71±0.28	3.87±0.18	2.61±0.42	3.03±0.29	2.85±0.64
2 and more (n:28)	2.65±0.50	2.96±0.48	1.82±0.35	2.15±0.55	3.77±0.25	3.81±0.27	2.76±0.43	3.11±0.28	2.86±0.30
p value	0.425	0.230	0.517	0.980	0.303	0.796	0.188	0.418	0.913

*Mann Whitney U Test **Kruskal Wallis Test

Note: As no woman started working in the 2nd month, the section of occupational activities was not analyzed.

Table 5b. The effect of some socio-demographic variables on the functional status of mothers

Variables	Household responsibilities		Social and community activities		Infant care responsibilities		Personal care activities		Occupational activities
	2. month	6. month	2. month	6. month	2. month	6. month	2. month	6. month	6. month
Delivery method									
Normal birth (n:11)	2.75±0.57	2.79±0.27	1.89±0.36	1.96±0.41	3.72±0.29	3.80±0.24	2.81±0.33	3.05±0.23	-
Caesarean -section (n:34)	2.54±0.46	3.09±0.47	1.78±0.41	2.18±0.53	3.76±0.25	3.84±0.24	2.66±0.46	3.09±0.30	2.86±0.43
p value	0.153	**0.042**	0.372	0.350	0.712	0.569	0.307	0.629	-
State of intending pregnancy									
Intended and planned (n:36)	2.63±0.47	3.03±0.43	1.83±0.39	2.16±0.55	3.75±0.27	3.82±0.22	2.72±0.41	3.09±0.31	2.86±0.45
Intended but unplanned (n:9)	2.41±0.58	2.98±0.51	1.71±0.40	2.02±0.30	0.75±0.22	3.87±0.33	2.61±0.52	3.04±0.13	2.80±0.40
p value	0.159	0.932	0.416	0.559	0.837	0.193	0.679	0.471	0.859
Contribution to infant care from the outside									
Yes (n:25)	2.50±0.47	3.08±0.47	1.79±0.42	2.24±0.60	3.76±0.26	3.84±0.24	2.69±0.35	3.10±0.31	2.87±0.36
No (n:20)	2.71±0.50	2.94±0.40	1.84±0.37	2.00±0.34	3.75±0.27	3.83±0.25	2.72±0.52	3.06±0.26	2.80±0.84
p value	0.192	0.301	0.543	0.247	0.962	0.890	0.705	0.676	0.894
Contribution to household chores from the outside									
Yes (n:24)	2.48±0.53	3.05±0.45	1.87±0.40	2.12±0.52	3.74±0.27	3.77±0.29	2.62±0.36	3.10±0.27	2.96±0.36
No (n:21)	2.71±0.42	2.98±0.44	1.74±0.39	2.14±0.51	3.76±0.25	3.91±0.13	2.79±0.49	3.05±0.30	2.70±0.52
p value	0.218	0.522	0.333	0.885	0.698	0.105	0.160	0.398	0.383

Mann Whitney U Test **Kruskal Wallis Test

Note: As no woman started working in the 2nd month, the section of occupational activities was not analyzed.

4. DISCUSSION, CONCLUSION AND SUGGESTIONS

Preparing both women and their partners for the challenges of parenthood before, during and after the birth is very important in terms of maternal, child, and family health. The preparation of fathers to fatherhood, as well as mothers for motherhood, positively affects family health (Apay Ejder & Pasinlioğlu, 2009; Beji et al., 2003; Özkan & Sevil, 2007; Sevil & Özkan, 2010).

According to the results of this study, positive changes developed in the functional status of women and their partners within the months marking the postpartum period. Women were found to have higher mean scores in the areas of household activities, social and community activities and self-care activities in the postpartum sixth month, compared to those of the postpartum eighth week. However, there was no significant difference between the postpartum second and sixth months in terms of infant care activities and occupational activities. According to some studies in the academic literature, the postpartum functional status of women shows significant development in the months following the birth (McVeigh, 1997; McVeigh & Chaboyer, 2002; Özkan & Sevil, 2007; Şanlı & Öncel, 2014; Tulman et al., 1990). While postpartum physical recovery generally takes six weeks to complete, adapting to the increasing responsibilities related the postpartum period, like parenthood and infant care, tends to require a longer time (Beji et al., 2003). Apay and Pasinlioğlu (2009) reported in their study that after childbirth, the scores obtained on the subscale of social and community activities increased. In the study conducted by Şanlı and Öncel (2014) they found that women become capable of fulfilling their responsibilities, primarily those related to infant care, followed by housework, within the first sixth months following childbirth.

In this study, a significant increase was observed in the household activities, social and community activities and child care activities of fathers in the postpartum sixth month, as compared to the postpartum eighth week. On the other hand, no significant difference was discovered between the months in terms of the mean scores obtained in the areas of infant care activities, self-care activities, occupational activities, and educational activities. In one study on the functional status of fathers, it was noted that following childbirth, for the most part, a majority of fathers maintained their regular participation level in household and intrafamilial activities, while for some, there was a very slight increase in their responsibility (McVeigh, 2001). Other studies have reported that husband's whose wives supported their involvement in infant care had more infant care responsibilities. An examination of the related expectations of couples within the first three weeks after the birth revealed that fathers felt unsupported in infant care when it came to actions (St John, Cameron, & McVeigh, 2005; Sevil & Özkan, 2010). There are only a very limited number of studies in Turkey regarding the functional status of fathers. In one of them, Sevil and Özkan (2010) found there to be no difference in the functional status of fathers at the time their wives were pregnant and during the postpartum period. In Turkey, men are not expected to participate in trainings with their wives before and after the birth, and there are only limited opportunities available for men to prepare them for the role of fatherhood. These drawbacks are believed to be related to the assumption that there is no difference between the functional status of "expectant" fathers and new fathers.

The present study found that age had no impact on the functional status of women and their partners. In contrast, Apay and Pasinlioğlu (2009) in their study, determined that with the increase of age, functional status scores decreased, as individuals avoided doing certain activities or did a lower rate of activities, possibly as a result of age-related physical inability and/or health disorders. In the study by Sevil and Özkan (2010), it was reported that compared to younger age groups, elderly fathers were better at occupational activities. The same study also found there to be a negative correlation between the age of fathers and social-community and self-care activities, and a positive relation between the age of fathers and household

activities. Şanlı and Öncel (2014) discovered in their study that as women aged their functional status increased, helping them to feel more responsible and experienced regarding infant care and household activities, particularly in cases where the women had a higher number of children. This last finding, as it relates to its positive effect on functional status, can be attributed to the fact that the women's exposure to postpartum functions helped them to be better prepared to handle the process. The women and their partners who participated in the present study were in the fertile age group, which could serve to explain the insignificant impact of age on functional status.

The results from this study found that educational level had a positive effect on functional status. As the educational level increased, the mean scores obtained by the women in the area of household activities were higher in the postpartum sixth month. Fathers with a higher educational level were determined to have higher mean scores in the area of infant care activities in the postpartum eighth week, and in the area of child care activities in both the postpartum second and the postpartum sixth month. Contrary to the results of the present study, Tulman et al., (1990) reported that women with a higher educational level had lower functional status than that of women with a lower educational level. Posmontier (2008) stated in his study that as the educational level increased, the self-care of women increased. Furthermore, in a study by Şanlı and Öncel (2014), primary school graduate women were observed to have a higher functional status in the area of infant care responsibilities in the postpartum sixth week, while university and higher education graduate women had a higher functional status in the area of self-care activities in the postpartum third and the postpartum sixth months when compared to other groups. In examining the studies on the functional status of fathers, McVeigh (2001) found that there was an increase in the functional status of fathers who had a higher educational level. Similarly, Sevil and Özkan (2010) also determined that the functional status of fathers was positively affected by a higher educational level. Although other studies have shown different results, the present study observed that women who had a caesarean section had stable mean scores in the area of household activities in the postpartum eighth week but higher mean scores in the same area in the postpartum sixth month, compared to those who gave normal birth. However, it was observed that the delivery method had no effect on the other activities.

Tulman and Fawcet (1988), in their study, determined that functions were regained after the postpartum sixth week, and that in general, women who gave vaginal birth were able to recover faster than those who had a cesarean section. In a related study, Beji et al., (2003) found that women who gave vaginal birth took on the infant responsibility earlier. Lastly, Şanlı and Öncel (2014) determined that women who gave vaginal birth regained their functional status of infant care faster than those who had a caesarean section. In the present study, the higher scores obtained by women in the area of household activities were associated with the higher level of support received regarding housework in the postpartum period.

The present study, found that the number of children, planned or unplanned status of pregnancy, and the provision of outside support for household activities and infant care did not affect the functional status. Tulman et al., (1990) in their study, observed that the increased number of pregnancies and births resulted in increased social-community activities and self-care activities. The study by Şanlı and Öncel (2014) determined that the number of children affected the postpartum functional status of women in the postpartum sixth week and the postpartum third month. Moreover, the same study found that with the increased number of children, household activities, infant care and total functional status increased, while the functional status regarding self-care decreased, with the difference being maintained in the postpartum sixth month. To continue, in the studies by Apay and Pasinlioğlu (2009) and Özkan and Sevil (2007), the women who had planned pregnancies were shown to have higher

postpartum functional status. On the other hand, Şanlı and Öncel (2014) found that the women who had unplanned pregnancies had higher mean scores in the area of infant care in the postpartum sixth week, in the area of household activities in the postpartum third month, and in the area of self-care activities in the postpartum third and sixth months. In the present study, the variable of planned/not planned pregnancy was shown to not affect the functional status of women in the postpartum second and sixth month, a finding associated with the fact that all the women who participated in the study had planned pregnancies.

Women who received support in their household activities have been shown to have less domestic activities, lower total postpartum functional status and less infant care responsibilities (Apay et al., 2009; Özkan et al., 2007). In the study conducted by Mc Veigh (2000b), it was determined that women who received social support in their postpartum sixth week, third month, and sixth month had significantly higher household, social-community and self-care activities. Özkan and Sevil (2007), in their study, found that women who received support in infant care had higher functional status regarding infant care. In the study conducted by Apay and Pasinlioğlu (2009), it was observed that women who received support in infant care had lower functional status regarding household activities and infant care. Further, Şanlı and Öncel (2014), determined that women who received no support in infant care and household activities returned to their functional status earlier. In the same study, it was reported that the lower functional status of women who received support in infant care and household activities was associated with the fact that routine work that had at once been done by them was now undertaken by supporters. In contrast to the studies cited above, the present study found there to be no difference in the women' functional status in terms of having/not having support in household activities and infant care. Considering the results of this study and the other studies mentioned above, it was concluded that having/not having outside help in household activities and infant care produces different effects on the women' functional status.

Nurses and midwives have important functions in the transition process of parents to their changing roles and in the development of healthy children. In Turkey, by identifying the functional status of women and their partners after the birth and by determining the factors affecting the functional status, women and their partners will have the opportunity to be given full support and thereby be better adapted to the postpartum period. With that said, it is important that nurses and midwives support parents throughout the pregnancy process, including up to the completion of the transition to their new roles of mother and father, and help them to make the changes necessary to prepare them for this new period of their life.

ORCID

Sevgi Özkan https://orcid.org/0000-0001-8385-210X
Sinem Göral Türkcü https://orcid.org/0000-0003-1574-0186
Pınar Serçekuş Ak https://orcid.org/0000-00029326-3453

5. REFERENCES

Anderson, A.M. (1996). The father- infant releationship; becoming connected. *Journal of the Society of Pediatric Nurses*,1(2), 83-92.

Apay Ejder, S., & Pasinlioğlu, T. A. (2009). Sectional study: the invertigation of functional situations of the women after the labor. *Hemşirelikte Araştırma Geliştirme Dergisi*, 1(1), 20-29.

Barclay, L., Donovan, J., & Genovese, A. (1996). Men's experiences during their partner's first pregnancy: a grounded theory analysis. *Australian Journal of Advanced Nursing*, 13(3),12-24.

Barclay, L., & Lupton, D. (1999). The experiences of new fatherhood: a socio-cultural analysis. *Journal of Advanced Nursing*, 29(4), 1013-1020.

Başbakkal, Z. (1999). *Babaların bebek bakımındaki rolü* (Doktora Tezi). [Role of parents in baby care] (Doctoral Thesis). Ege Üniversitesi, Institute of Health Sciences, Izmir.

Beji, N.K., Coşkun, A., & Yıldırım, G. (2003). Effects of childbirth on the functional ability of women. *Hemşirelikte Araştırma Geliştirme Dergisi*, 1, 22-29.

Fawcett, J., & Tulman, L.(1990). Building a programme of research from the roy adaptation model of nursing, *Journal of Advenced Nursing*, 15(6), 720-725.

Fawcett, J., Tulman, L., & Myers, S. (1988). Development of the inventory of functional status after childbirth. *Journal of Nurse-Midwifery*, 33(6), 252-260.

Fichardt, A.E., Van Wyk, N.C., & Weich, M. (1994). The needs of postpartum women. *Curationis*, 17(1), 15-21.

Gjerdingen, D.K., & Chaloner, K.(1994). Mothers' experience with houshold roles and social support during the first postpartum years. *Women Health*, 21(4), 54-74.

Henderson, A.D, & Brouse, A.J. (1991). The experiences of new fathers during the first 3 weeks of life. *Journal of Advanced Nursing*,16(3), 293-298.

Herbert, P. (1994). Support of first- time mothers in three months after birth. *Nursing Times*, 90(24), 36-37.

Hodnett, E.D. (1996). Nursing support of the laboring woman. *Journal of Obstetric, Gynecologic and Neonatal Nursing*, 25(3), 257-264.

McVeigh, C.A. (1997). Functional status after childbirth: a comparison of Australian women from english and non-english speaking backraunds. *Australian College of Midwives Incorporated Journal*, 10(2), 15-21.

McVeigh, C. A. (2000a). Anxiety and functional status after childbirth. *Australian College of Midwives Journal*, 13(1), 14-18.

McVeigh, C.A. (2000b). Satisfaction with social support and functional status after childbirth. *Maternal Child Nursing Journal*, 25(1), 25-30.

McVeigh, C.A. (2001). Functional status after fatherhood, an Australian study. *Journal of Obstetric, Gynecologic and Neonatal Nursing*, 31, 32-38.

McVeigh, C.A., Chaboyer, W. (2002). Reability and validity of inventory of functional status after childbirth when used in Australian population. *Nursing & Health Sciences*, 4(3), 107-112.

Özkan, S., & Sevil, Ü. (2007). The study of validity and reliability of inventory of functional status after childbirth. *TSK Koruyucu Hekimlik Bülteni*, 6, 199-208.

Posmontier, B. (2008). Functional status outcomes in mothers with and without postpartum Depression. *Journal of Midwifery & Women's Health*, 53(4), 310-318.

Sevil, Ü., & Özkan, S. (2010). Psychometric properties of a Turkish version of the inventory of functional status-fathers (IFS-F). *Turkiye Klinikleri Journal of Nursing Science*, 2, 16-23.

St John, W., Cameron, C, & McVeigh. C. (2005). Meeting the challenge of new fatherhood during the early weeks. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 34(2), 180-189.

Şanlı,Y., & Öncel, S. (2014). Kadınların doğum sonrası fonksiyonel durumları ve etkileyen faktörlerin belirlenmesi [Evaluation of the functional status of woman after childbirth and effective factors]. *Journal* of *Turkish Obstetric* and *Gynecology Society*, 2, 105-114.

Tulman, L., & Fawcett, J. (1988). Return of functional ability after childbirth. *Nursing Research*, 37, 70-75.

Tulman, L., Fawcett, J., Groblewski, L., & Silverman, L. (1990). Changes in functional status after childbirth. *Nursing Research*, 39(2), 70-75.

Tulman, L.J., Fawcett, J., & Weiss, M. (1993). The inventory of functional status−fathers: development and psychometric testing. *Journal of Nurse-Midwifery*, 38, 276-282.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 524–543

DOI: 10.21449/ijate.444882

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

The Relationship between Attitudes towards Asylum Seekers and Compassion Levels of University Students

İsmail Yelpaze[1,*], **Deniz Güler**[2]

[1]Kahramanmaraş Sütçü İmam University, Faculty of Education, Kahramanmaraş, Turkey
[2]Anadolu University, Faculty of Education, Eskişehir, Turkey

Abstract: The aim of this study is to adapt the attitude scale towards asylum seekers and refugees to Turkish culture and to examine the relation of university students' attitudes towards asylum seekers to certain personality traits. The study was conducted with the participation of 340 university students. The attitude scales for both asylum seekers and refugees were adapted to Turkish culture and validity and reliability studies were conducted in the study. Findings have shown that they provide the necessary conditions on both scales and can be used as a reliable and valid measuring tool in studies. In addition, students with a high level of compassion, empathy and understanding have been found to have a more positive attitude toward asylum seekers / refugees at a meaningful level. Also, the students studying at the faculty of theology have more positive attitude than the students studying at other faculties. However, the attitude towards asylum seekers / refugees did not differ according to gender. Findings are discussed in the light of the literature.

ARTICLE HISTORY
Received: 03 March 2018
Accepted: 07 July 2018

KEYWORDS
Refugee,
Asylum Seeker,
Attitude Scale,
Sensitivity,
Empathy,
Understanding,
Compassion

1. INTRODUCTION

Mankind has been forced to migrate from some place to another due to various reasons since the first day of its life. Thus, migration is a question as old as the human history. The ones who migrate to create a better future for themselves are called immigrants. Immigration may take place in two different forms, i.e. willingly or forcedly. Forced migration refers to the individuals' abandonment of their countries temporarily, but forcedly, due to the threats aimed at their safety, arising from the ongoing civil wars, ethnic, religious, or political oppressions in their countries (Martin, 2002). The persons under such circumstances are given the status of refugee or asylum-seeker (Aksu-Kargın, 2016).

Some countries are ranked at the top of the list of countries preferred the most by the immigrants, thanks to their economic opportunities, state policies, and geographical locations. Turkey has become a country immigrant population of which has been increasing rapidly in the recent years (Ünal, 2014). Especially due to the civil war erupted in Syria as of 2011, millions of Syrians have taken shelter in Turkey. Various factors play a role in making such a choice. Turkey has a critical position for Syrian refugees both as a passageway to Europe, and as an

CONTACT: İsmail Yelpaze ismailyelpaze@gmail.com Kahramanmaraş Sütçü İmam University, Faculty of Education, Kahramanmaraş, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

asylum (Kolukırık, 2014). In the meantime, Turkey's adoption of the open-door policy since the first days of the war in Syria (Aras-Gökalp & Mencütek, 2015), Syrians' trust in Turkey, and religious motives are also influential in this choice (Disaster and Emergency Management Agency of T.R. Prime Ministry-AFAD, 2013).

Since the arrival of the first refugee convoy in Turkey as of 2011, it was estimated as of 2013 that, the number of the Syrians living in the camps reached 600,000, while the ones living outside the camps were more than 400,000 (Dinçer, Federici, Ferris, Karaca, Kirişçi, & Çarmıklı, 2013). While the the number of Syrians was announced as 2.749.140 as of 2016 (The Office of the United Nations High Commissioner for Refugees-UNHCR, 2016), it was estimated as 3.412.368 Syrian people as of December 2017 (Directorate-General of Immigration Administration of T.R. Ministry of Internal Affairs, 2017).

Not restricting the Syrian asylum-seekers to living within a specific region, but allowing them to take part in the daily life, Turkey has offered the Syrian asylum-seekers various opportunities for their orientation. For instance, the ones who decide not to live in the tent cities are provided with the right to reside at the downtowns of the cities. In addition, they are given the permission to open up businesses, healthcare support, most importantly, the students are given the opportunity to resume their education not only at universities, but also at the primary and secondary educations (Yıldız, 2013). In order to resume their higher-education, the Syrian students are at first provided with the opportunity to learn Turkish, and then to resume their higher-education. It appears that the Syrian students seize this opportunity very well. In fact, while there were only 455 Syrian students studying at the higher-education institutions in Turkey as of 2010, the same figure has significantly risen up to 17967 only for those studying at the state universities (out of the total of 20701) as of the academic year of 2017-2018 (YÖK (Higher-Education Institute), 2018).

Apart from the state's receptive attitude, whether the native population is evenly receptive is a critical factor for the psychological and social harmony of the refugees or asylum-seekers who have left their countries and been trying to adapt to their new settlements. Just like in elsewhere around the world, the asylum-seekers and refugees may encounter negative attitudes and behaviors, such as social exclusion, discrimination, racism, xenophobia, alienation, also in Turkey (Erdoğan, 2015; Ünal, 2014). In this respect, it is suggested that, such negative attitudes of the native population against the asylum-seekers/refugees has a great potential of negatively affecting the rehabilitation of the asylum-seekers.

It draws attention that, the first studies with regard to the attitudes towards the asylum-seekers/refugees were performed in the countries being most frequently applied by the asylum-seekers/refugees, such as Canada, Australia, USA, South Africa (Gordon, 2016; Nickerson & Louis, 2008). It is emphasized in this study that, cognitive, affective, and behavioral factors stand out predominantly among those influencing the society's attitudes towards the refugees (Croucamp, O'Connor, Pedersen, & Breen (2017). Economic reasons (Christophersen, Liu, Thorleifsson, and Tiltnes, 2013; Ceobanu & Escandell, 2010), nationalism and misbeliefs (Pederson, Attwell & Heveli, 2005), safety, social harmony, cultural threat, and social dominance tendency (Nickerson & Louis, 2008) have been singled out the most in this respect among the factors affecting the attitude towards refugees.

Personality has been suggested as a critical factor affecting the attitude towards refugees, like individual's belief system, socio-cultural and political tendencies. However, this factor seems to have not been taken into consideration seriously in the research studies. In one of the few number of studies performed in this regard, the traits of being agreeable and open to experience have been singled out as a critical precursor of the positive attitude towards refugees (Gallego & Pardos-Prado, 2014). Compassion, as being a trait, is also suggested to be among the critical factors determining the attitude towards the refugees. Compassion signifies the state

of emotion, thinking, and behavior, consisting of helping the others in overcoming their grief and problems without judging them, and being aware of tenderness and sharing (Neff, 2003). Compassion is a feature that enables us to understand the inner-world of the individual, and drives us to support him/her (Demirci-Seyrek, Ersanlı & Tunç, 2016). In view of the studies performed respectively, a meaningful positive change has been observed in the attitudes and behaviors of the school staff, having undergone a sensibility-based training program, towards the children suffering trauma (Wilson, 2013). In the meantime, sensibility trainings seem to be utilized in the arrangements intended for the refugee, hybrid, or trauma-victim children, and in bringing up qualified persons and suitable environments intended for eliminating the increasing degeneration in the society (Arnot, Pinson, & Candappa, 2009; Boyden, 2009). After all, it may be suggested that the persons with higher levels of sensibility/compassion are to bear more positive attitudes towards the persons under difficult situations, such as refugees or asylum-seekers.

It has been foreseen that the increase in the number of refugee and asylum-seeker population may have certain negative effect on both Turkish society and on the refugee and asylum-seeker population (Ünal, 2014). It is implied that Turkish society's attitudes towards Syrian asylum-seekers are not all that positive (Demir, 2016). It is emphasized that seeing the asylum-seekers as an economic burden, causing disturbance in the social order and daily life are the motives that create negative attitudes among a part of Turkish society, which is expanded day by day (Yıldız, 2013).

It is deemed significant to support these opinions emphasized by the authors with empirical research studies. However there is a limited number of studies performed on this issue, and almost all of them seem to be qualitative studies attained by interviews made up of open-ended questions (Ergin, 2016; Karakuş & Göktuna-Yaylacı, 2015; Özdemir & Öner-Özkan, 2016). Qualitative studies provide more in-depth knowledge in comparison to the knowledge obtained by quantitative assessment instruments. However, it is in need of performing more comprehensive research studies in order to bring light onto the question of how and at which level both native population and the asylum-seekers themselves are affected from the ever-increasing number of the refugees and asylum-seekers (Ünal, 2014). Thus valid and reliable assessment instruments are sought after in order to evaluate Turkish society's attitudes towards the asylum-seekers/refugees, and the relation of these attitudes with psycho-social and personal factors. As a result of the literature search within the scope of the study, no assessment instrument having undergone validity and reliability studies in Turkish culture was found available in order to assess the society's attitude towards refugees/asylum-seekers Lack of a scale to assess the society's attitude towards the asylum-seekers makes it difficult to examine the relation of the other variants with the society's attitude. It is hereby intended first to have the assessment instruments, namely the "Illegal Aliens Scale", developed by Ommundsen & Larsen (1997), and the "Attitudes Towards Asylum Seekers Scale", developed by Pederson, Attwell & Heveli (2005), adapted to the Turkish culture, and undergo comparison in terms of their psychometric features in order to assess the attitude towards asylum-seekers/refugees. It is intended in the second part of the study to examine the relations between attitude towards asylum-seekers, refugees and sensibility/compassion.

2. METHOD

This research has been made in accordance with the relational model of general screening model. In addition, this study is a scale adaptation study using the basic survey model. Firstly, two scales were adapted to Turkish culture and secondly, the relations among some personality traits and attitudes to refugee and asylum seekers.

2.1. Study Group

The study group of this study is composed of students at undergraduate level at a university. The students were selected from 7 different faculties through appropriate sampling method. The average age of participants ranging from 17 to 29 is 20.95. A total of 340 participants consisted of 101 male, 234 female and 5 ones not explained.

2.2. Instruments

2.2.1. *Attitudes towards Refugees Scale*

The original scale (Illegal Aliens Scale) was developed by Ommundsen and Larsen (1997) to measure the attitudes towards illegal immigrants. At the first stage 80 statements related to attitudes to illegal immigrants were written and these were applied to 75 university students. After the item analysis, 30 items 14 positive and 16 negative statements remaindered and their item-total correlation ranged from .59 to .76. In the second stage, a different university student sample with 100 people was used. There was a significant correlation of .89 in the scale split method. In the third stage, a similar sample of 115 university students was examined in relation to similar scales. In this study, attitude scales for African-Americans and homosexuals were used within the scope of concurrent validity. As a result of the analysis, there was a significant correlation between the current scale and the African-American attitude .44 and the homosexual attitude .60.

2.2.2. Attitudes towards Asylum-Seekers Scale:

The original scale (Attitudes Towards Asylum Seekers Scale) was developed by Pederson et al. (2005) to measure the attitude of the public towards asylum seekers. During the development phase of the scale, firstly people from different socio-economic status were interviewed. Later, a yearly publication related to asylum-seeker of a journal was reviewed and twenty-five scale items were written based on these reports. These items were sent to 20 people for clarification and to improve the writing rules. This review resulted in an 18-item form containing 9 negative and 9 positive statements

An exploratory factor analysis was performed to determine the factor structure of the scale. At the end of the analysis, a three factor structure emerged which is higher than the value of 1, which explains 66.9% variance. However, a single-factor structure has been accepted because it has a one-factor structure in the screen plot graph. As a result, a one-factor structure was acknowledged, which has Cronbach alpha reliability coefficient of .94 and item-total correlation of .30 above.

2.2.3. Compassion Scale

The Compassion Scale was developed by Demirci-Seyrek, Ersanlı and Tunç (2016) with the aim of measuring the sensitivity of adult individuals. The measuring tool was developed in the sample of academic staff. In the development process of the scale, exploratory factor analysis was performed and a four-dimensional, 15-item measurement tool was formed. Dimension names are called mercy, empathy, understanding, and humanitarianism. Item-total correlation and sub-over-group comparisons were examined within the context of item analysis. Confirmatory factor analysis also found that the scale's structure was confirmed. In the reliability analysis, Cronbach alpha values were .70 for the mercy factor, .65 for the empathy factor, .62 for the understanding factor, .60 for the humanitarianism factor and .80 for the total scale.

2.3. Data Collection and Analysis

In the first part of the works intended for the adaptation of the attitude towards asylum-seekers scales, among the authors in charge of the assessment instruments, Knud Larsen was

contacted for the Attitude towards Illegal Immigrants Scale, and Anne Pedersen was contacted for the Attitude towards Asylum-Seekers Scale, thereupon the permits necessary for the adaptation of them to the Turkish culture were obtained. Further required permits were obtained from the Ethics Board of the related university with regard to the practicability and applicability of the study.

Language validity of the assessment instruments was maintained upon the consensus of five persons having command of both languages, and four of whom were to serve in the field of psychological counseling. Two ones translated each item into Turkish independently. Later, it was decided which translations were more appropriate for Turkish language and culture. For validity and reliability of translated scale data were collected from 340 university students by the researchers. Application permit was obtained from the lecturer of the course upon showing him/her the ethics board report. Although being present in the classroom during the application, the lecturers did not interfere in the application. Seeking for the voluntariness of the students with regard to the application, they were further provided with the essential information on the study by the researchers themselves. The applications lasted for an average of 15 minutes.

Data were selected from the students studying at seven different faculties of a state university. Not only there are asylum-seekers living together with the native society, there is also an asylum-seekers camp in the city where the university is domiciled. It was therefore suggested that the students may be familiar with the asylum-seeker concept. Furthermore, the participants were sought to have been residing in the city for no shorter than 1.5 years, so that they had come by the asylum-seekers anyhow.

Before transferring the data attained by making use of assessment instruments to the computer, two forms not suitable to the directive were excluded. Before starting the analysis of the data, average values were used in place of missing data, normal distribution criteria were reviewed, and the data were thereby made ready for the analyses. SPSS and AMOS programs were used in data analysis.

Confirmatory factor analysis was performed for the construct validity of the assessment instruments. Chi-Square fit test ($\chi 2$), the ratio of the Chi-Square to the degree of freedom ($\chi 2/sd$), Comparative Fit Index (CFI), Normalized Fit Index (NFI), Incremental Fit Index (IFI), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Root Mean Square Error of Approximation (RMSEA), and the fit indices thereof were examined in the evaluation of the model's construct. CFI, NFI, IFI GFI, and AGFI values being more than .90, and RMSEA value being less than .05 are the tokens of good fit (Byrne, 2001).

3. FINDINGS

3.1. Validity Studies of Attitudes towards Refugees Scale

This subsection includes results of validity studies about Attitudes towards Refugees Scale.

3.1.1. Translation study

In the studies of adaptation from another language, translation suitability is quite critical for the proper understanding of the expressions in the language and culture to which the adaptation is to be made. In the translation process, first of all, the items of the assessment instruments were translated into Turkish by two persons independent of each other, who had a command of both languages, and were serving in the field of psychological counseling. The translations were subsequently reviewed by the researchers, and having the translations considered as being more meaningful and to be understood more easily in the culture accepted consensually, Turkish form of the assessment instruments were thereby created (see appendix 1).

3.1.2. Construct validity

Confirmatory factor analysis was performed in order to ascertain whether the factor structure of the unique form of the scale was similar in Turkish culture. Confirmatory factor analysis is utilized in testing the hypotheses about the structures previously established on a theoretical basis (Brown, 2006).

While KMO was reviewed in order to decide about the data's suitability for the factor analysis, Bartlett Test was reviewed in order to ascertain whether the correlation matrix was actually a unit matrix. As a result of the analysis, KMO value was found as .93, and the Bartlertt test was found as meaningful ($p \leq .001$). These values reveal that the sample is suitable for undergoing factor analysis (Hair, Anderson, Tatham and Black, 2003).

Fit indices of the single-factor structure of the scale's Turkish form were examined by way of DFA analysis. In the first analysis, not only the fit indices were found as being inadequate, but also the factor loads of some of the items were found as being quite low. That is why the items 7, 12, 21, and 22 with item-factor loads lower than .30 (Martin & Newel, 2004) were phased out of the analysis. Taking a look at the aforementioned items deleted, they were found either as being not directly related with the asylum-seekers (item 7: Entering our country is quite easy), or not setting an example for the conditions of the asylum-seekers living in Turkey (item 21: Asylum seekers should not encounter discrimination). Confirmatory factor analysis was repeated with the remaining 26 items, and data regarding DFA were shown in Figure 1.

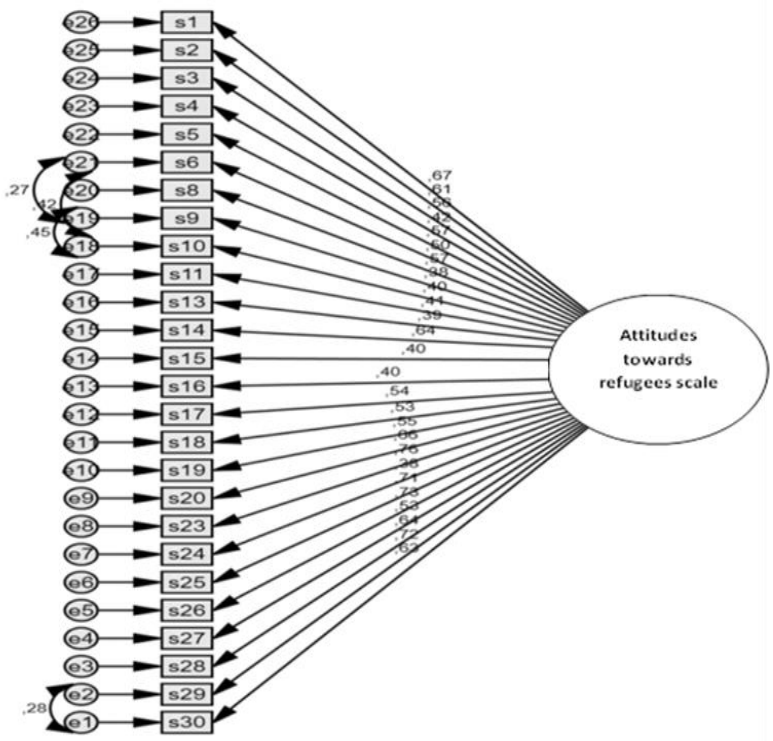


Figure 1. Results of Confirmatory Factor Analysis of Attitudes to Refugees

Standardized estimates of the items are shown in Figure 1. As being seen in the figure, standardized estimates vary between .39 and .76. These estimates are meaningful at the level of $p \leq 0.05$. The model fit ($\chi 2$=363.09, sd=271, $p \leq .005$, $\chi 2/sd$=1.34) and the fit indices (RMSEA=.03 NFI=.89 CFI=.97 IFI=.97 AGFI=.90 GFI=.92) of the single-factor scale construct with 26 items were found to be at good level.

3.1.3. Concurrent Validity

The attitude towards asylum-seekers scale having undergone adaptation in the current study was utilized in examining the similar scale validity of the attitude towards refugees scale. Pearson correlation coefficient between the two assessment instruments was sought after. Meaningful relation of .84 was found between the two scales as an outcome of the analysis ($p \leq .001$). Such a relation is deemed as being at a high level (Cohen, 1988).

3.1.4. Discriminant validity-Comparison of 27% Lower-Upper Group Averages

Attitude towards Refugees Scale was examined in terms of its adequacy in distinguishing the persons as per the trait it assessed. This analysis allows researchers to establish the scale's ability to distinguish those who have a positive attitude towards the characteristic that is being measured from those who have a negative attitude towards the same (Erkuş, 2005). In this context, averages of lower 27% (91 participants) and upper 27% groups were compared as per each item and total score of the scale. t-test analysis was performed for ascertaining whether there was a meaningful difference between these groups (Table 1). As a result of the analysis, a meaningful relation was found between the lower and upper group averages both as per each item and total score of the scale [$t(180)=-42.662$, $p<.001$]. Although there is a meaningful relation between the lower and upper group averages, it is recommended to have the effect size calculated in order to ascertain whether the same relation is significant in theory and practice (Huck, 2008). The effect size of the study was calculated by $\eta 2$ as .91. This value is considered to be a major effect (Cohen, 1988).

3.2. Reliability Studies of Attitudes towards Refugees Scale

This subsection includes results of reliability studies about Attitudes towards Refugees Scale.

3.2.1. Internal Consistency Coefficient

Internal consistency reliability analyses were examined at first in the context of the reliability of the assessment instrument. Having the internal consistency of the assessment instrument calculated by means of Cronbach Alfa coefficient, it was found as .92. This value indicates that the internal consistency of this value assessment instrument is quite high (Pallant, 2005).

Corrected item-total correlation was examined in order to determine the level of the relation between the items of the scale. Since the item being examined was excluded from the remaining of the scale in this examination, the corrected item-total correlation was preferred. The correlation values seem to vary between .36 and .71 as a result of the analysis (Table 1). It is asserted that these values are to be higher than .30 in order to be accepted (Erdoğan, Nahcivan & Nihal, 2014).

Table 1. T test results of comparison of 27% group, Cronbach alpha, item-total correlation, and test-retest values of Attitudes towards Refugees Scale

Item No	t value for 27% groups	Cronbach α	Item-total correlation	Test-retest correlation
s1	-17.268**	.92	.62	.31*
s2	-12.749**	.92	.57	.46**
s3	-13.003**	.92	.54	.57**
s4	-8.988**	.92	.42	.55**
s5	-12.873**	.92	.54	.70**
s6	-11.321**	.92	.52	.53**
s8	-12.249**	.92	.54	.49**
s9	-7.529**	.92	.40	.49**
s10	-8.474**	.92	.44	.69**
s11	-7.753**	.92	.39	.59**
s13	-7.545**	.92	.38	.41**
s14	-15.006**	.92	.62	.76**
s15	-8.379**	.92	.37	.50**
s16	-10.311**	.92	.40	.52**
s17	-12.772**	.92	.53	.65**
s18	-11.414**	.92	.50	.67**
s19	-13.432**	.92	.56	.70**
s20	-16.267**	.92	.62	.62**
s23	-22.140**	.91	.71	.53**
s24	-7.723**	.92	.36	.32*
s25	-16.908**	.91	.67	.80**
s26	-17.923**	.91	.69	.56**
s27	-11.120**	.92	.48	.47**
s28	-13.763**	.92	.60	.36*
s29	-21.698**	.91	.69	.63**
s30	-15.173**	.92	.61	.55**
Total	-42.662**	.92	-	.67**

3.2.2. Test-retest reliability

Test-retest analysis was performed in order to ascertain the temporary stability of the assessment instrument, or whether the constant scores remained constant during the transition from one condition to another (Burns & Grove, 2003). In this context, the assessment instrument was applied to 41 university students twice in three weeks of interval. The relation between the two assessments was examined by way of Pearson Correlation analysis. Meaningful relation of .67 was found between the two assessments as an outcome of the analysis ($p \leq .001$). Such a relation is deemed as being at a high level (Cohen, 1988).

3.3. Validity Studies of Attitudes towards Asylum-Seekers Scale

This subsection includes results of validity studies about Attitudes towards Asylum-Seekers Scale.

3.3.1. Translation study

In the studies of adaptation from another language, translation suitability is quite critical for the proper understanding of the expressions in the language and culture to which the adaptation is to be made. In the translation process, first of all, the items of the assessment instruments were translated into Turkish by two persons independent of each other, who had a command of both languages, and were serving in the field of psychological counseling. The translations were subsequently reviewed by the researchers, and having the translations considered as being more meaningful and to be understood more easily in the culture accepted

consensually, Turkish form of the assessment instruments were thereby created (see appendix 2).

3.3.2. Construct validity

Confirmatory factor analysis: Confirmatory factor analysis was performed in order to ascertain whether the factor structure of the unique form of the scale was confirmed in Turkish culture. Confirmatory factor analysis is utilized in testing the hypotheses about the structures previously established on a theoretical basis (Brown, 2006).

While KMO was reviewed in order to decide about the data's suitability for the factor analysis, Bartlett Test was reviewed in order to ascertain whether the correlation matrix was actually a unit matrix. As a result of the analysis, KMO value was found as 0.85, and the Bartlertt test was found as meaningful (p≤.001). These values reveal that the sample is suitable for undergoing factor analysis (Hair, Anderson, Tatham and Black, 2003).

Fit indices of the single-factor structure of the scale's Turkish form was examined by way of DFA analysis. In the first analysis of the adaptation study, not only the fit indices were found as being inadequate, but also the factor loads of some of the items were found as being quite low. That is why the items 3, 4, 7, 9, 12, 14, and 15 with item-factor loads lower than .30 (Martin & Newel, 2004) were phased out of the analysis. Taking a look at them, there are items that do not fit to the situation of the shelter-seekers in Turkey (item 7: Asylum-seekers are rightful to go on hunger strike to draw attention to their conditions). Confirmatory factor analysis was repeated with the remaining 11 items, and data regarding DFA were shown in Figure 2.

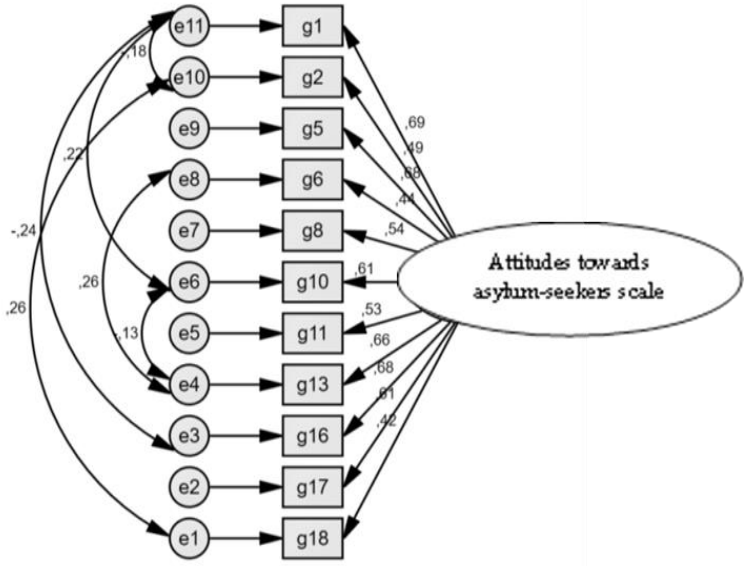


Figure 2. Results of Confirmatory Factor Analysis of Attitudes to Asylum-Seekers

Standardized estimates of the items are shown in Figure 2. As being seen in the figure, standardized estimates vary between 0.42 and 0.69. These estimates are meaningful at the level of p≤0.05. The model fit (χ2=64.68, sd=38, p≤ .005, $\chi 2/sd$=1.70) and the fit indices (RMSEA=.04 NFI=.94 CFI=.97 IFI=.97 AGFI=.94 GFI=.96) of the single-factor scale construct with 11 items were found to be at good level.

3.3.3. Concurrent validity

The attitude towards refugees scale having undergone adaptation to Turkish culture in the current study was utilized in examining the similar scale validity of the attitude towards asylum-seekers scale. Pearson correlation coefficient was sought after between the two assessment instruments. Meaningful relation of .84 was found between the two scales as an outcome of the analysis ($p \leq .001$). Such a relation is deemed as being at a high level (Cohen, 1988).

3.3.4. Discriminant validity-Comparison of 27% Lower-Upper Group Averages

The Attitude towards Asylum-Seekers Scale was examined as per its ability to distinguish those who have a positive attitude towards the trait that is being measured from those who have a negative attitude towards the same (Erkuş, 2005). In this context, averages of lower 27% and upper 27% groups were compared as per each item and total score of the scale (Table 2). Unpaired t-test analysis was performed for ascertaining whether there was a meaningful difference between these groups. As a result of the analysis, a meaningful relation was found between the lower and upper group averages both as per each item and total score of the scale [$t(180)=-39.073$, $p<.001$]. Although there is a meaningful relation between the lower and upper group averages, it is recommended to have the effect size calculated in order to ascertain whether the same relation is significant in theory and practice (Huck, 2008). By means of et square ($\eta 2$) the effect size was calculated for this study, and found as .89. This value is deemed as being a strong effect (Cohen, 1988).

3.4. Reliability Studies of Attitudes towards Asylum-Seekers Scale

This subsection includes results of reliability studies about Attitudes towards Refugees Scale.

3.4.1. Internal consistency coefficient

Internal consistency reliability analyses were examined at first in the context of the reliability of the assessment instrument. Having the internal consistency of the assessment instrument calculated by means of Cronbach Alfa coefficient, it was found as 0.85. This value indicates that, the internal consistency of this value assessment instrument is quite high (Pallant, 2005). Each and every item was further examined as per its relation with the total reliability of the scale. In this examination, it was found that, deletion of any one of the scale's items was not to change the scale's total reliability significantly (Table 2). It has been suggested that researchers should not make changes in the scale if excluding the item does not lead to major changes in Cronbach's alpha coefficient (Cortina, 1993).

Item-total correlation was examined in order to determine the level of the relation between the items of the scale. While the item being examined was excluded from the remaining of the scale in this examination, therefore the corrected item-total correlation was preferred. The correlation values seem to vary between 0.41 and 0.62 as a result of the analysis (Table 2). It is asserted that, these values are to be higher than .30 in order to be accepted (Erdoğan, Nahcivan, & Nihal, 2014).

3.4.2. Test-retest reliability

Test–retest reliability is an important analysis as it shows the extent to which the responses to the items in the scale may change over time (Burns & Grove, 2003). In this context, the assessment instrument was applied to 36 university students twice in three weeks of interval. The relation between the two assessments was examined by way of Pearson Correlation analysis. Meaningful relation of .52 was found between the two assessments as an outcome of the analysis ($p \leq 0.01$). Such a relation is deemed as being at a medium level (Cohen, 1988). Temporary stability of each item was further examined. As a result of this examination, all but one (g6) of the items was found to possess stability valid for the assessments (Table 2).

Table 2. Cronbach alpha, item-total correlation and test-retest values of Attitudes towards Asylum-Seekers

Item No	t value for 27% groups	Cronbach α	Item-total correlation	Test-retest correlation
g1	-18.166**	.83	.60	.69**
g2	-10.968**	.84	.47	.51**
g5	-18.797**	.83	.61	.42*
g6	-9.299**	.84	.42	.32
g8	-13.331**	.84	.50	.39*
g10	-13.995**	.83	.55	.45**
g11	-12.597**	.84	.49	.63**
g13	-15.849**	.83	.62	.57**
g16	-14.943**	.83	.59	.42*
g17	-14.614**	.83	.55	.83**
g18	-10.875**	.84	.41	.60**
Total	-39.073**	.85	-	.52**

3.5. The Relations of Attitudes towards Refugee and Asylum-Seekers with Some Variables

This part includes results of some demographic variables and the relations of them with attitudes to refugees and asylum-seeker. Firstly, the participants' attitudes towards asylum-seekers and refugees were examined as per their genders, and outcomes attained therefrom were shown in Table 3.

Table 3. T-test results of comparisons of attitudes towards refugees and asylum-seekers in terms of gender

	Group	n	Mean	Sd	df	t	p
Refugee	Male	100	87.75	21.62	332	-.697	.48
	Female	234	89.47	20.21			
Asylum-seeker	Male	100	36.49	9.87	332	.120	.90
	Female	234	36.36	8.82			

As being seen in Table 3, university students' attitudes towards asylum-seekers and refugees do not reveal a meaningful variation as per their genders (p>.05).

Relation of the participants' attitudes towards asylum-seekers and refugees with various variants were examined. The relation between the compassion personality scale of the attitude scales in this direction and the sub-dimensions thereof is shown in Table 4.

Table 4. The results of correlations among attitudes towards refugees, asylum-seekers and compassion

Variables	Asylum-seeker	Refugee	Compassion (total)	Compassion (subscale)	Empathy	Understanding	M	Ss
Asylum-seeker	-						6.35	9.18
Refugee	86**	-					8.87	20.78
Compassion	-17**	-17**	-				3.93	7.32
-Mercy	-12*	-13*	78**	-			2.41	2.66
-Empathy	-18**	-16**	72**	39**	-		1.55	2.61
-Understanding	-10	-13*	75**	53**	41**	-	7.08	2.34
-Humanitarian	-08	-06	63**	30**	26**	25**	2.90	2.54

As seen in Table 4, a very highly meaningful relation at the level of .86 was found between the attitude towards asylum-seekers and the attitude towards refugees (p≤.01). While the attitude towards asylum-seekers scale displayed meaningful relation of -.17 with compassion trait, of -.12 with mercy sub and -.18 with empathy, that are among the sub-dimensions of compassion (p≤.05), it displayed no meaningful relation with understanding and humanitarian features (p>.05).

While the attitude towards refugees scale displayed meaningful relation of -.17 with compassion trait, of -0.13 with mercy, -0.16 with empathy and -0.13 with understanding that are among the sub-dimensions of compassion (p≤.05), it displayed no meaningful relation with humanitarian features (p>.05).

Whether the university students' attitudes towards refugees varied as per the faculties they were studying at was analyzed via ANOVA, and the results thereof are shown in Table 5.

Table 5. Anova results of comparisons of attitudes towards refugees in terms of study departments

Source	Sum of Squares	df	Mean Square	F	p≤
Between Groups	19436.821	6	3239.470	8.499	.001
Within Groups	126545.056	332	381.160		
Total	145981.876	338			

As being seen in Table 5, the participants' attitudes towards refugees vary as per the faculties they study at (p≤.05). Scheffe among the post-hoc tests of ANOVA was utilized in order to ascertain the source of this variation. Scheffe Test is known to be the firmest test so as to avoid 1st Type error (Huck, 2000). Findings regarding Scheffe Test are shown in Table 6.

Table 6. Post-hoc (scheffe) results of comparisons of attitudes towards refugees in terms of study departments

	1	2	3	4	5	6
1.Education						
2.Economics& administrative sciences	.32					
3.Theology	20,77*	20.45*				
4.Science&Letters	4.08	3.76	-16.68*			
5.Health sciences	-11.29	-11.61	-32.06*	-15.38		
6.Engineering	-4.71	-5.03	-25.48*	-8.79	6.58	
7.Agriculture	4.86	4.55	-15.90*	.78	16.16	9.58

As being seen in Table 6, a meaningful difference was found (p≤.05) between the theology faculty students' attitudes towards immigrants and those of the students from other faculties. This difference is to the favor of the theology faculty students. In other words, the theology faculty students' attitudes towards refugees are meaningfully more positive than those of the students from other faculties. No significant difference was found among the students from other faculties.

Whether the university students' attitudes towards asylum-seekers varied as per the faculties they were studying at was analyzed via ANOVA, and the results thereof are shown in Table 7.

Table 7. Anova results of comparisons of attitudes towards asylum-seekers in terms of study departments

Source	Sum of Squares	df	Mean Square	F	p≤
Between Groups	3898.925	6	649.821	8.763	.001
Within Groups	24619.477	332	74.155		
Total	28518.402	338			

As being seen in Table 7, the participants' attitudes towards asylum seekers vary as per the faculties they study at (p≤.05). Scheffe among the post-hoc tests of ANOVA was utilized in order to ascertain the source of this variation. Findings regarding Scheffe Test are shown in Table 8.

Table 8. Post-hoc (scheffe) results of comparisons of attitudes towards asylum-seekers in terms of study departments

	1	**2**	**3**	**4**	**5**	**6**
1. Education						
2. Economics& administrative sciences	2.05					
3. Theology	10.31*	8,264*				
4. Science&Letters	3.33	1,284	-6,980*			
5. Health sciences	-2.17	-4,230	-12,494*	-5,514		
6. Engineering	-.97	-3,032	-11,296*	-4,315	1,198	
7. Agriculture	2.86"	,806	-7,458*	-,478	5,036	3,838

As being seen in Table 8, a meaningful difference was found (p≤.05) between the theology faculty students' attitudes towards asylum-seekers and those of the students from other faculties. This difference is to the favor of the theology faculty students. In other words, the theology faculty students' attitudes towards asylum-seekers are meaningfully more positive than those of the students from other faculties. No significant difference was found among the students from other faculties.

4. DISCUSSION AND CONCLUSION

Reviewing the findings of the research from a holistic point of view, it may be said that the Turkish form of both the attitude towards asylum-seekers scale and the attitude towards refugees scale is valid, and possesses a reliable structure.

Structural validity of the attitude towards refugees scale has not been verified just as in the original scale. Since some of the items thereof do not suit to their situation, these items may therefore have not made sense to the refugees in our country. Furthermore, it seems difficult to discern whether some of the items are related directly with the refugees. That is why the analyses were resumed after deleting four items from the original scale, and the structural validity was thereby maintained. The data attained with regard to the reliability of the scale show that, the assessment instrument has performed a reliable assessment. In fact both the internal consistency coefficient and the data attained from the test-retest method show that, the conditions required for reliable assessment have been maintained. After all, a one-dimensional, 26-item attitude towards refugees scale with proven validity and reliability, and rendering 34% of variance, has been brought in Turkish culture.

Structural validity of the attitude towards asylum-seekers scale has also not been verified as in the original scale. Since seven items were found to be not setting an example for the conditions of the asylum-seekers living in Turkey, and not being sufficiently related with the whole of the scale, they were deleted from the assessment instrument. After that structural

validity of the scale could have been verified in its new form. The scale has been found to be not only maintaining distinctive validity, but also possessing sufficient internal consistency coefficient, and remaining at a sufficient level in terms of test-retest score. After all, a one-dimensional, 11-item attitude towards asylum-seekers scale with proven validity and reliability, and rendering 40% of variance, which may be utilized for the assessment of the attitude towards asylum-seekers, has been brought in Turkish culture.

Considering the fact that, the difference between refugee and asylum-seeker has not been well-discerned in Turkish society, both two assessment instruments have been adapted to Turkish culture. In fact, reviewing the studies performed in Turkey, it may be seen that, the concepts of Syrian asylum-seeker (Kardeş, Banko and Akman, 2017; Topkaya & Akdağ, 2016), and Syrian refugee (Doğanay & Çoban-Keneş, 2016; Kağnıcı, 2017) have been used to identify the same group. Meanwhile, the high correlation (.86) between the instruments assessing the attitude towards refugees and those assessing the attitude towards the asylum-seekers in the current study may also be interpreted as a sign of the students' unawareness of the difference between the two concepts. After all, as long as they are valid and with proven reliability, the attitude towards refugees/asylum-seekers scales may be utilized interchangeably. However, the attitude towards asylum-seekers scale may be deemed as being more useful thanks to the lesser number of its items, and for having rendered greater variance.

As the second objective of the study, the relation of the students' attitudes towards the asylum-seekers with certain variants has been examined. It has been found that the students' attitudes towards refugees and asylum-seekers do not change as per their genders. Similarly, no significant factor was found between the gender and the attitude towards asylum-seekers in the study performed by Kanbur (2017) among teachers, and in the study performed by Anderson (2017) among university students as well. On the other hand, there is an empirical research (Karaoğlu, 2015), and a meta-analysis study (Anderson & Ferguson, 2017), both having shown the fact that males possess significantly more negative attitude than that of the females. Reviewing the research findings from a holistic point of view, the relation of gender with the attitude towards asylum-seekers varies. It may be deemed necessary to perform more comprehensive studies, by way of including different variants in the research studies to be performed respectively in order to ascertain the source of this variation.

It has been found that the students' attitudes towards the refugees and asylum-seekers are related with sensibility, traits, and the elements of compassion, empathy, and understanding that compose such traits. Despite the fact that the higher are the traits they possess, the more positive the attitudes they display towards refugees and asylum-seekers become, no significant relation has been found between human traits and the attitudes towards asylum-seekers.

Sensibility signifies the state of emotion, thinking, and behavior, consisting of helping the others in overcoming their grief and problems without judging them, and being aware of tenderness and shares (Neff, 2003). Sensibility is a feature that enables us to understand the inner-world of the individual, and drives us to support him/her (Demirci-Seyrek, Ersanlı, and Tunç, 2016). In view of the studies having been performed respectively, a meaningful positive change has been observed in the attitudes and behaviors of the school staff, having undergone a sensibility-based training program, towards the children suffering trauma (Wilson, 2013).In another study it was found the lower the individuals' empathy skills, the higher their threat perceptions and prejudices regarding the refugees (Karaoğlu, 2015). Contemplating on the findings of the current study in conjunction with the outcomes attained from the previous studies, it may be said that the ones possessing higher sensibility trait are the ones also possessing higher empathic skills, and those who approach people in difficult situations affectionately. In this respect, it is a highly anticipated outcome that, the persons with higher sensibility traits are to bear positive attitudes towards the asylum-seekers trying to live in

Turkey under difficult situations and severe traumas. After all, it may be suggested that the persons with higher levels of sensibility are to bear more positive attitudes towards the persons being under difficult situations.

It has been examined whether the university students' attitudes towards refugees and asylum-seekers vary as per the faculties they study at. Students from seven faculties were compared with this intent, and students from the faculty of theology were found as having adopted the sought attitude being meaningfully more positive than the students from other faculties. However, no significant difference was found respectively among other faculties. This outcome suggests the existence of a relation between religious attitude or devotion and the attitude towards asylum-seekers. In fact, it has been reported from several studies that, the ones with higher religious devotion bear more negative attitude towards asylum-seekers (Anderon & Ferguson 2017; Perry, Paradies & Pedersen, 2014). This situation is referred to the religion's feature of being a determining factor in numerous prejudices (Shariff, Willard, Andersen, & Norenzayan, 2016). However, it is seen that the study group in the previous studies were composed of Christians persons being. In view of the fact that the refugees are mostly devouts of other religions, a cultural conflict may be in question. In the current study, for both the participants and the refugees are the devout of the same religion, a cultural sensitivity may come to the fore from the point of view of the theology faculty students.

Despite the significant contributions having been brought in the field by this research, it also has certain limitations. First of all, after the deletion of the items from the assessment instrument, the findings attained from the same sample were dealt with. Having not performed a reconfirmation process in a different sample is a limitation of the study. On the other hand, having not approached a different sample in order to examine the relation of the demographic features and levels of sensibility of the participants with their attitudes towards asylum-seekers is also considered as a limitation.

In conclusion, taking the findings attained from the study into consideration, a number of suggestions may be made for both the researchers and the practitioners. First of all, two assessment instruments with proven validity and reliability may be utilized for the purpose of assessing the attitudes towards the asylum-seekers and refugees. On the other hand, these assessment instruments may be retested in samples with different demographic features. Furthermore, different variants that may be related with the individuals' attitudes towards asylum-seekers may also be uncovered. Practitioners, too, may organize trainings on the development of the individuals' sensibility characteristics which are critical for both the asylum-seekers' adaptation to our country, and for the society's recognition of their presence.

ORCID

İsmail Yelpaze https://orcid.org/0000-0003-4428-0502
Deniz Güler https://orcid.org/0000-0001-6006-5795

5. REFERENCES

AFAD. (2013). *Türkiye'deki Suriyeli sığınmacılar 2013: Saha araştırması sonuçları* [Syrian refugees in Turkey 2013: Field survey result]. Retrieved March, 2015 from https://www.afad.gov.tr/Dokuman/TR/60-20131230154 91-syrian-refugees-in-turkey-2013_baski_30.12.2013_tr.pdf

Aksu-Kargın, İ. (2016). *The Syrian refugees in Turkey: Their living conditions and the ways in which their presence has affected Turkish society*. Doctoral dissertation. Indiana University.

Anderson, J. R. (2017). The moderating role of socially desirable responding in implicit–explicit attitudes toward asylum seekers. *International Journal of Psychology*. doi:10.1002/ijop.12439

Anderson, J. & Ferguson, R. (2017). Demographic and ideological correlates of negative attitudes towards asylum seekers: A meta- analytic review. *Australian Journal of Psychology*. doi:10.1111/ajpy.12162

Aras-Gökalp, E. N. & Mencütek, Z. Ş. (2015). The international migration and foreign policy nexus: The case of Syrian refugee crisis and Turkey. *Migration Letters, 12*(3), 193-208.

Arnot, M., Pinson, H. & Candappa, M. (2009). Compassion, caring and justice: Teachers' strategies to maintain moral integrity in the face of national hostility to the 'non-citizen'. *Educational Review, 61*(3), ss.249-264, doi:10.1080/00131910903045906

Bello, V. (2016). Inclusiveness as Construction of Open Identity: How Social Relationships Affect Attitudes Towards Immigrants in European Societies. *Social Indicators Research, 126*(1), 199-223.

Boyden, J. (2009). What place the politics of compassion in education surrounding non-citizen children? *Educational Review, 61*(3), ss.265-276, doi:10.1080/00131910903045914

Burns, N. & Grove, S. K. (2003). *Understanding nursing research* (3th ed., pp. 265–307).

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Ceobanu, A. M. & Escandell, X. (2010). Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: A review of theories and research. *Annual Review of Sociology, 36*, 309-328. doi:10.1146/annurev.soc.012809.102651

Christophersen, M., Liu, J., Thorleifsson, C. M. & Tiltnes, A. A. (2013). Lebanese Attitudes towards Syrian Refugees and the Syrian Crisis. *Results from a National Opinion Poll. FAFO Paper*, (13).

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Croucamp, C. J., O'connor, M., Pedersen, A. & Breen, L. J. (2017). Predicting community attitudes towards asylum seekers: A multi- component model. *Australian Journal of Psychology, 69*(4), 237-246.

Çetinsaya, G. (2014). *Büyüme, kalite, uluslararasılaşma: Türkiye yükseköğretimi için bir yol haritası* [Growth, quality, internationalization: A road map for higher education in Turkey]. Eskişehir: Anadolu University Press.

Çiçekli, B. (2014). *Yabancılar ve Mülteci Hukuku* [Foreigners and Refugee Law] (5th ed.). Ankara: Seçkin Punlication.

Demir, O. Ö. (2016). Crisis or opportunity? Public debate and perception on Syrian Refugees in Turkey. In (D. Eroglu, J. H. Cohen and I. Sirkeci). *Turkish Migration 2016 Selected Papers*. London: Transational press.

Demirci-Seyrek, Ö., Ersanlı, K. & Tunç, T. (2016). Duyarlık Ölçeği [Validation of the compassion scale]. *Electronic Journal of Social Sciences, 15*(56), 284-294.

Dinçer, O. B., Federici, V., Ferris, E., Karaca, S., Kirişci, K. & Çarmıklı, E. Ö. (2013). *Suriyeli Mülteciler ve Türkiye: Sonu Gelmeyen Misafirlik* [Turkey and Syrian refugees: The limits of hospitality]. International Strategic Research Organization (USAK).

Doğanay, Ü. & Çoban Keneş, H. (2016). Yazılı basında Suriyeli 'Mülteciler': Ayrımcı söylemlerin rasyonel ve duygusal gerekçelerinin inşası [Syrian 'Refugees' in printed

press: Construction of rational and emotinal justifications of discriminatory discourses]. *Mülkiye Dergisi, 40*(1), 143-184.

Erdogan, M. (2015). *Türkiye'deki Suriyeliler: Toplumsal Kabul ve Uyum* [Syrians in Turkey: Social acceptance and integration research]. İstanbul: Bilgi Üniversitesi Yayınları

Erdoğan, S., Nahcivan, N. & Nihal M. N. (2014). *Hemşirelikte Araştırma Süreç, Uygulama ve Kritik* [Nursing research. Process, implementation, and critics]. Ankara: Nobel publication.

Ergin, H. (2016). Turkish University Students' Perceptions Towards Their Syrian Classmates. *Education and Science, 41*(184). doi: http://dx.doi.org/10.15390/EB.2016.6214

Ergün, E. (2012). *Avrupa insan hakları sözleşmesi ve Türk hukukunda sınır dışı etme geri gönderme ve geri verme* [European human rights treaty and deportation in Turkish law]. Ankara: Yargı Publication.

Gallego, A. & Pardos-Prado, S. (2014). The big five personality traits and attitudes towards immigrants. *Journal of Ethnic and Migration Studies, 40*(1), 79-99.

Gordon, S. L. (2016). Welcoming refugees in the rainbow nation: contemporary attitudes towards refugees in South Africa. *African Geographical Review, 35*(1), 1-17.

Gökçearslan-Çifci, E., Özdemir, B., Arslan, E. & Akgül-Gök, F. (2016). Attitudes of shopping mall workers towards Somali asylum seekers. In (D. Eroglu, J. H. Cohen and I. Sirkeci). *Turkish Migration 2016 Selected Papers*. London: Transational press.

Hair, J. F. Jr., Anderson, E.R., Tatham, L.R. & Black, W.C. (2003). Multivariate Data Analysis, 5th edition. New Delhi: Pearson Education.

Huck, S. W. (2000). *Reading statistics and research*. New York: Addision Wesley Longman

Kağnıcı, D. Y. (2017). Suriyeli mülteci çocukların kültürel uyum sürecinde okul psikolojik danışmanlarına düşen rol ve sorumluluklar [School counselors' roles and responsibilities in cultural adaptation process of Syrian refugee children]. *İlköğretim Online, 16*(4). doi:10.17051/ilkonline.2017.342990

Kanbur, N. İ. (2017). *İlkokulda görev yapan öğretmenlerin mülteci öğrencilere yönelik tutumlarının incelenmesi* [The examination of the primary school teachers' opinions about the problems of refugee students]. Master Thesis. Sakarya University, Educational Science Institute. Sakarya.

Karakuş, M. & Göktuna-Yaylacı, F. (May, 2015). Student perceptions on refugees: an evaluation based on multiculturalism. *International Conference on Contemporary Issues in Education*. Dubai, United Arab Emirates.

Karaoğlu, E. (2015). *The role of social dominance orientation, empathy and perceived threat in predicting prejudice of Turkish citizens toward Syrian immigrants*. Master Thesis. Midde East Technical University, Graduate school of social sciences, Ankara

Kardeş, S., Banko, Ç. & Akman, B. (2017). Sosyal medyada Suriye'li sığınmacılara yönelik algı: bir sözlük değerlendirmesi [Perceptions about Syrian refugees on social media: an evaluation of a social media platform]. *Göç Dergisi, 4*(2), 185-200.

Martin, P. L. (2002). Germany: Managing migration in the 21st century. *Institute of European Studies*.

Neff, K. D. (2003). Self- compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85-101.

Nickerson, A. M. ve Louis, W. R. (2008). Nationality Versus Humanity? Personality, Identity, and Norms in Relation to Attitudes Toward Asylum Seekers. *Journal of Applied Social Psychology, 38*(3), 796-817.

Özdemir, F. & Öner-Özkan, B. (2016). Türkiye'de sosyal medya kullanıcılarının Suriyeli mültecilere ilişkin sosyal temsilleri [Social representations of social media users toward Syrian refugees in Turkey]. *Nesne, 4*(8), 227-244. doi: 10.7816/nesne-04-08-04

Pedersen, A., Attwell, J. & Heveli, D. (2005). Prediction of negative attitudes toward Australian asylum seekers: False beliefs, nationalism, and self-esteem. *Australian Journal of Psychology, 57(*3), 148-160. DOI: 10.1080/00049530500125157

Perry, R., Paradies, Y. & Pedersen, A. (2014). Religious ambivalence: Suppression of pro-social attitudes toward asylum seekers by right-wing authoritarianism. *The International Journal for the Psychology of Religion*, 25, 1-33. https://doi.org/10.1080/10508619.2014.921473

Sansonetti, S. (2016). *Female refugees and asylum seekers: The issue of integration*. Policy department C: Citizens' Rights and Constitutional affairs.

Directorate-General of Immigration Administration of T.R. Ministry of Internal Affairs (2017). Göç istatistikleri: Geçici koruma [Migration statistics: Temporary protection]. Retrieved December 2017 from http://www.goc.gov.tr/icerik6/temporary-protection_915_1024_4748_icerik

Topkaya, Y. & Akdağ, H. (2016). Sosyal Bilgiler Öğretmen Adaylarının Suriyeli Sığınmacılar Hakkındaki Görüşleri -Kilis 7 Aralık Üniversitesi Örneği- [Prospective Social Studies Teachers' Views About Syrian Defector-Kilis 7 Aralık University Sample-]. *Çankırı Karatekin University Journal of Institute of Social Sciences, 7*(1), 767-786.

UNHCR (2016). *Syria regional refugee response inter-agency information sharing portal*. Retrieved April 2016 from http://data.unhcr.org/syrianrefugees/country.php?id=224

Ünal, S. (2014). Türkiye'nin beklenmedik konukları:"Öteki" bağlamında yabancı göçmen ve mülteci Deneyimi [Turkey's unexpected guests: Foreign immigrants and refugees experience in the context of "other"]. *Zeitschrift für die Welt der Türken/Journal of World of Turks*, *6*(3), 65-89.

Van Der Veer, K., Ommundsen, R., Hak, T. & Larsen, K. S. (2003). Meaning shift of items in different language versions. A cross-national validation study of the Illegal Aliens Scale. Quality & Quantity, *37*(2), 193-206.

Ward, C. & Masgoret, A. M. (2008). Attitudes toward Immigrants, Immigration, and Multiculturalism in New Zealand: A Social Psychological Analysis1. *International Migration Review, 42*(1), 227-248.

Yıldız, Ö. (2013). Syrian refugees in the Turkey's camps: Problems, Expectations, Turkey and Future Perception. *Journal of Sociological Research, 16*(1), 140-169.

YÖK-Higher Education Institute (2018). *Yükseköğretim bilgi yönetim sistemi: Öğrenci istatistikleri* [Higher education information management system: Student statistics]. Retrieved June, 2018 from https://istatistik.yok.gov.tr/

Appendix-1: Attitudes towards Refugees Scale

1= Hiç katılmıyorum, 2= Çok az katılıyorum , 3= Kısmen katılıyorum, 4= Çoğunlukla katılıyorum 5= Tamamen katlıyorum

Mültecilere Yönelik Tutum Ölçeği

1	Sığınmacılar benim verdiğim vergiden faydalanamamalıdır.	1	2	3	4	5
2	Bizim vergilerimiz ülkemizde bulunan sığınmacılar için kullanılabilmelidir.	1	2	3	4	5
3	Ülkemizde herkese yetecek yeterince yer var.	1	2	3	4	5
4	Sığınmacılar ülkemizin kaynaklarını kötüye kullanmıyorlar.	1	2	3	4	5
5	Sığınmacılar toplumumuza sıkıntı vermektedir.	1	2	3	4	5
6	Ülkemizin sınırları uluslararası tüm sığınmacılara açık olmalıdır.	1	2	3	4	5
7	Sığınmacılar gıda yardımı alabilmelidir.	1	2	3	4	5
8	Ülkemizde doğum yapan sığınmacılar ülke vatandaşlığına alınmalıdır.	1	2	3	4	5
9	Ülkemiz tüm siyasi sığınmacıları kabul etmelidir.	1	2	3	4	5
10	Sığınmacıların ülkemize her yıl milyonlarca dolar maliyeti olmaktadır.	1	2	3	4	5
11	Sığınmacılar ülkemize önemli ölçüde insan kaynağı sağlamaktadırlar.	1	2	3	4	5
12	Devletimizi sığınmacılara ücretsiz olarak eğitim, bakım, sağlık hizmeti sunmalıdır.	1	2	3	4	5
13	Sığınmacılar yüzünden ülkemizin nüfusu aşırı derecede artmaktadır.	1	2	3	4	5
14	Sığınmacılar bu ülkenin vatandaşlarıyla aynı haklara sahip olmamalıdır.	1	2	3	4	5
15	Sığınmacıların da hakları var.	1	2	3	4	5
16	Başka ülkelerin vatandaşlarına yardım etmek bizim ülkemizin sorumluluğunda değildir.	1	2	3	4	5
17	Sığınmacıların bu ülkenin vatandaşlarıyla aynı haklara sahip olmaya hakları vardır.	1	2	3	4	5
18	Mülteciler kendi ülkelerine dönmeye zorlanmalıdır.	1	2	3	4	5
19	Benim vergim mültecileri desteklemek için kullanılmamalıdır.	1	2	3	4	5
20	Ülkemizde yaşayan herkese eşit fırsatlar sunulmalıdır.	1	2	3	4	5
21	Sığınmacılar büyüyen bir soruna dönüşmektedir.	1	2	3	4	5
22	Sığınmacıları ülkemizin dışında tutmak için, sınırlarımızdaki güvenlik artırılmalıdır.	1	2	3	4	5
23	Kendi evimizi nasıl koruyorsak, ülkemizi de sığınmacılardan korumalıyız.	1	2	3	4	5
24	Sığınmacıların ülkemizi kendi sorunlarına bir sığınak olarak görmelerini istemiyorum.	1	2	3	4	5
25	Sığınmacılar toplumun yozlaşmasına neden olmaktadır.	1	2	3	4	5
26	Sığınmacılar kanunları ihlal ediyorlar.	1	2	3	4	5

Ters kodlanan maddeler: 2,3,4,6,7,8,9,11,12,15,17,20

Puanın yüksekliği olumsuz tutuma işaret etmektedir.

Appendix-2: Attitudes towards Asylum-Seekers Scale

	Sığınmacılara Yönelik Tutum Ölçeği					
1	Sığınmacılar, çeşitli eylemlerde bulunarak devletimize istediklerini yaptırıyorlar.	1	2	3	4	5
2	Sığınmacılar ülkemizde uzun süre kalmak istiyorlarsa bu hak onlara verilmelidir.	1	2	3	4	5
3	Sığınmacılar, ülkemizde çeşitli eylemler yaparak nankör davranıyorlar.	1	2	3	4	5
4	Sığınmacıların içinde bulundukları duruma üzülüyorum.	1	2	3	4	5
5	Devletin sığınmacılara yönelik politikasını doğru buluyorum.	1	2	3	4	5
6	Sığınmacılar, kendine zarar verme gibi davranışlara başvurarak insanları ndırmaya çalışıyorlar.	1	2	3	4	5
7	Sığınmacılar, toplumumuza uyum sağlamak için çaba harcamıyorlar	1	2	3	4	5
8	Sığınmacılar meşru mültecidir ve onlar hoşgörü ile karşılanmaları gerekir.	1	2	3	4	5
9	Sığınmacılar toplumda nefret uyandırıyor.	1	2	3	4	5
10	Eğer sığınmacılar ülkemizde mutlu değilse, kendi ülkelerine gönderilsin.	1	2	3	4	5
11	Sığınmacıların, kamplar yerine toplumun içinde olmaları daha iyi olur.	1	2	3	4	5

Ters kodlanan maddeler: 2,4,5,8,11

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 544–566

DOI: 10.21449/ijate.425190

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

Vocational High School Teachers' and Students' Opinions about Performance Tasks

Ayşegül Altun [1*], **Hülya Kelecioğlu** [2]

[1] Ondokuz Mayıs University, Faculty of Education, Samsun, Turkey.
[2] Hacettepe University, Faculty of Education, Ankara, Turkey

Abstract: The aim of the study is to determine the opinions of teachers and students in a Vocational High School about performance tasks. This study was carried out with 17 vocational high school students and 13 teachers. This study was conducted through qualitative research design by analyzing the records of the individuals who shared their own interviews, observations and documents. The data were analyzed through content analysis. The first codes were generated for each question and the collected data were arranged in an inducible approach by determining appropriate themes for these codes. All the codes and themes were checked by second researchers. Moreover, the consistency in coding was determined. In addition to the interviews, observations and document analysis were performed. The results of this study demonstrate that both the students and the teachers do not have sufficient knowledge about what the performance task is and how it is carried out. Therefore, the performance tasks carried out do not reach their goals. Performance tasks should measure higher order thinking, and they should be related to real life situations. The results of the study show that performance tasks do not demonstrate these features. In addition, students should have knowledge and understanding of the subjects in order to complete their performance tasks. On the other hand, they are supposed to perform a performance task about a subject which they mostly do not understand or they do not know. This may be the reason why performance task does not reach its goal.

ARTICLE HISTORY

Received: 04 January 2018
Revised: 23 April 2018
Accepted: 03 june 2018

KEYWORDS

Performance task,
Content analysis,
Vocational high school,
Teachers' opinions,
Students' opinions

1. INTRODUCTION

One of the aims of classroom assessment is to improve the quality of teaching activities. The change in primary education programs enabled the addition of new and different assessment approaches such as performance task, project, portfolio, etc. to the traditional assessment and evaluation techniques. In 2013-2014 academic year, the Ministry of National Education renewed the programs in secondary education, and the competent authority also introduced the performance task as a compulsory activity in secondary education. When teachers assess the students in order to determine their achievement level, there is a decision-making orientation based on more comprehensive and multi-source knowledge within a student-centered approach to education. In this orientation method students should not be

CONTACT: Ayşegül ALTUN ✉ aysegul.altun@omu.edu.tr Ondokuz Mayıs University, Educational Faculty, Department of Educational Sciences, Ankara, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

evaluated through a couple of exams in terms of their success status and teachers should not make a decision about students' success by means of one-type of evaluation (Berberoğlu, 2006).

The researchers suggest the use of alternative measurement and evaluation approaches which enable students to demonstrate their differences and skills, critical thinking, creativity, and higher order thinking skills, to put their information into practice in real life, to construct their new knowledge by using the old ones and to establish a connection among them (Kan, 2007). Ignoring the alternative assessment procedures when measuring the student's achievement also leads students to acquire one-way work habits, and predominantly being one-sided of their learning (Berberoğlu, 2006). For this reason, the researchers have started to focus on new approaches enabling to find out the characteristics of the students such as knowledge, skill, motivation, interest, attitudes, and to recognize these characteristics. One of these approaches is the performance-based assessment.

Performance is the vigorous effort of the students in the process of producing a product by using the knowledge and skills they acquire (Kutlu, Doğan & Karakaya, 2014). The performance-based assessment measures the knowledge and the skills students have. The Ministry of National Education (MoNE) has obliged the use of performance tasks at high schools pursuant to a circular issued in 2013. In the same circular, the ministry defined the performance tasks as the activities of groups or individual students who shared their experiences based on the research by using the skills of critical thinking, creativity, problem-solving and reading comprehension. In addition, these tasks should be conducted under the supervision of teachers.

Students are expected to think critically and to solve complex problems during the performance-based assessment process (Wright, 2008). At the same time, students should use their knowledge and skills in a situation that is linked to daily life (Başol, 2010). For an evaluation process which is based on performance, the observable performance of the product should include higher order thinking process, social skills and study process, interdisciplinary transition and exchange of information (Berberoğlu, 2007). Performance-based assessments are linked to real life situations (Neukrung & Fawcett, 2010), and they give students the opportunity to demonsrate their own knowledge in real life situations (Russel & Aurasian, 2012). Performance-based assessments should have a clear purpose, identify observable performances or product, provide appropriate setting for assessing the performance and provide scores for performance (Russel & Aurasian, 2012).

Performance tasks are a type of assessment that is critical for the determination of students' real life skills. However, it requires both students and teachers to get out of their usual learning and assessment patterns. For this reason, many studies have been carried out in order to determine the views of students and teachers on performance task. Some of the studies on this subject were summarized below.

Teachers stated that projects and performance tasks were useful but evaluating them was time comsuming (Güvey, 2009). Similar to these results, Dağhan and Akkoyunlu (2005) found out that if performance-based assessment was used effectively, it contributed to the lessons. However, these tasks are not preferred because of time constraints and difficulty in using these methods. Teachers say that they give performance tasks to their students since the tasks make them bring together the topics taught in everyday life, direct them to individual work and research, and help them to make self-assessments. Most of the teachers taking part in the study also agree that performance tasks are useful (Doğan-Temur, Bayındır & İnan, 2010). In the research conducted by Duruhan & Çavuş (2013), teachers say that performance tasks reinforce the students' learning, encourage the collaboration and socialization among the students, direct the students to research, and develop creative thinking and sense of responsibility. But they also say most of the tasks do no reach their specific goals. Another research which was conducted

with the teachers reveals that performance tasks make students think critically and ensure more permanent learning for them. Teachers also give feedback and make guidance to students while preparing performance tasks. In addition, they consider their students' needs, interests and requests while determining the subjects (Yılmaz & Benli, 2011). Another research conducted by Birgin and Baki (2012) shows that teachers have difficulties in implementing alternative assessment methods in classrooms, and that they have lack of knowledge about them. On the other hand, other researchers state that teachers prefer performance tasks more commonly than other alternative methods and that they think they are more proficient at perfomance tasks (Duran, Mıhladız & Ballıel, 2013). In addition, these teachers say that they use performance tasks to support students' learning, and that they give the tasks to their students due to compelling reasons (Kaya, Karaçam, Eş & Tuncel, 2013). According to the results of the study performed by Çiftçi (2010), the teachers state that performance tasks have positive effects such as developing creative thinking and sense of responsibility and that it contributes to classes when it is used effectively.

In addition to these statements about performance task, teachers express the problems encountered in the system such as negative attitudes of parents and their lack of interest (Çiftçi, 2010; Yılmaz & Benli 2011), time constraint (Dağhan & Akkoyunlu, 2005; Güvey, 2009; Yılmaz & Benli 2011), lack of equipment and crowded classes (Çiftçi 2010) and high costs (Yılmaz & Benli, 2011). Another point that teachers complain about is that the students copy the tasks from the internet (Çiftçi, 2010; Yılmaz & Benli, 2011) or that the tasks are performed by the parents (Yılmaz & Benli, 2011). Insufficient physical facilities of schools and classrooms are also considered as a problem while using alternative assessement methods in classes (Yiğit & Kırımlı, 2014).

In the studies conducted on students about performance tasks, it has been found out that the tasks contribute to the development of students' numerous skills such as establishing connection between the knowledge and real life and communicating and doing research, that the students can associate the mathematics with the daily practices, that they improve their communication skills with their friends and environment, that the performance task enables students to do self-criticism, that the tasks reveal the real achievements of the students more efficiently, and that the performances allow the students to be aware of their deficiencies by making positive contribution to their learning (Bal, 2012; Çetin & Çakan, 2012). In those studies, students have talked about some negative aspects of the performance tasks such as time-consuming, exhausting and stressful nature of the preparation process, the difficulties encountered while drawing a picture in comparison with the task of writing a document, and limited access to the resources to be used for the tasks. (Bal, 2012; Karaçam, 2013). The students who watch TV less frequently, and who benefit from the books as resources instead of internet have higher scores in performance tasks (Hastürk & Özkan, 2015). Akay and Küçükkaragöz (2016) stated that there was a positive, high and significant correlation between performance task scores and mathematical achievement scores and that there was also a positive and significant and moderate correlation between performance tasks and attitudes towards matmetamics. Therefore, they conclude that performance tasks increase the students' achievements.

Sayhan (2017) conducted a study in order to determine the opinions of high school students about performance tasks, and she found out that the students preferred doing performance tasks because they could do them out of school, and performance tasks increased their grades. Half of the students say that the performance tasks enhance their knowledge, make them active and help them in making self-assessments, revealing their real achievements more efficiently, and contributing to the discovery of their interest and talents. In contrast to these positive statements, students also remark that doing performance tasks in all courses is time

consuming, and that they cannot study for the exams effectively because of these tasks. Preparing tasks is not easy and they have some difficulties in affording the materials.

1.1. Purpose of the research

MoNE changed the curriculum in 2005-2006 academic year, and it became compulsory for the students to prepare performance tasks for each lesson in primary education. In secondary education, from 2013-2014 academic year on, students have to prepare performance tasks for each lesson in every semester. Secondary school teachers have encountered these regulations for the first time although the students who are in secondary education are familiar with these concepts. For this reason, it is important to determine the problems that teachers have especially in secondary education. Since the students will also suffer from the problems faced by the teachers, the opinions of the students should also be evaluated.

Evaluation of the skills in vocational courses at the vocational high school through the performance evaluation will ensure that students determine their ability to acquire skills in these areas more efficiently, and it will also provide effective feedback to teachers and students. In addition, the students at this vocational school have low achievement grades, and performance task requires higher order thinking skills and connecting their knowledge with real life situations. In this sense, it is important to identify what the teachers and students at the vocational high school understand about the performance-based assessment and the problems they experience in the performance-based assessment. For this reason, this study was carried out on the students and teachers of a vocational high school.

This study has aimed to determine what the students and teachers at the vocational high school understand about the concept of performance task, to evaluate the performance tasks at a vocational high school and the process of preparing and evaluating performance tasks, and to identify the problems encountered in this process.

1.2. The Importance of Research

There are many studies about performance tasks in the literature. However, these studies are mostly based on the views of primary school students, teachers and/or parents because the performance tasks were compulsory in primary education. (Ayva, 2010; Bal, 2012; Çetin & Çakan, 2010; Güneş & Soran, 2013; Karaçam, 2013) (Birgin & Baki, 2012; Doğan-Temur, Bayındır & İnan, 2010; Duruhan & Çavuş, 2013; Güvey, 2009; Yılmaz & Benli, 2011). There are not many studies on the views of secondary school teachers and students in the literature. From this point of view, it is thought that this study will contribute to the field.

Since the students at vocational schools in Turkey receive professional education, the students have to adapt the information they have learned in vocational courses to their daily life. Hence, performance evaluation is more important for vocational high schools than other high schools. We consider that the results of this study focusing on the implementation of performance tasks and the problems faced by both the teachers and the students and the suggestions which will be made in the light of these results for the solution of above mentioned problems will be useful for more effective implementation of the performance task at the vocational high school.

2. METHOD

This study aims to to get more detailed and comprehensive information about the opinions of the teachers and students at a vocational high school on performance tasks. Therefore, a qualitative method was chosen.

2.1. Participants

In this study, we chose the appropriate sampling design among purposeful sampling types. Purposeful sampling assumes that the researcher has to choose a sample to learn, develop insights into any subject, and learn in great detail what s/he wants to learn (Merriam, 2013). We carried out this study with the participation of 17 students and 13 teachers who were thought to represent the views of the teachers and students at the vocational school on performance tasks. Before asking the people to participate in the study, we explained the purpose of working with teachers and students, and we stated that we would usethe results of the interviews. Volunteer teachers and students participated in the study, and we provided detailed feedbackon the participants' performance tasks. All of thirteen teachers who were willing to cooperate were included in the study. Table 1 shows the distributions of the teachers according to their gender, the faculties they graduate from and the years of service.

Table 1. Demographics of participants teachers

Features	f	%
Gender		
Female	5	38,46
Male	8	61,54
Faculties they graduated		
Science and Literature	5	38,46
Education	8	61,54
Service year		
5 years and less	6	46,15
6-10 years	2	15,38
11-15 years	2	15,38
15 years and more	3	23,08

In Table 1, it is seen that the majority of the teachers are male, that they graduated from the Faculty of Education and that they have been working for 5 years or less.

The school where we conducted the study is located in a district of Ankara, and mostly the students who were not able to succeed in the high school entrance exams receive education at this high school. We selected one successful (the students whose average scores are above 3 have been considered as successful), one unsuccessful (the students whose average scores are below 2 have been considered as unsuccessful), and two moderately successful (whose average scores are about 2) students at each grade level. We interviewed 17 students in total – 4 of them were from the 10^{th} grade, 4 from the 11^{th} grade, 4 from the 12^{th} grade and 5 from the 9^{th} grade.

2.2. Data Collection Tools

We prepared semi-structured interview forms in order to determine the opinions of the teachers and students about the performance tasks in the research. We drew up these forms by analyzing the relevant studies in the literature, and we put the forms into their final form by taking the opinions of two teachers who did not take part in the study and two experts of measurement and evaluation in terms of the reasonableness of the statements and the clarity of the questions.

We ensured the methodological diversity by analyzing the documents (criteria or scales that teachers use when evaluating the performance tasks) of the teachers who shared their interviews, observations and documents during the data collection phase. We made the observations at the same time during the data collection process in the school but we did not use any data collection tool the observation. During the observation, we examined the

performance tasks prepared by the students within the scope of the performance evaluation and presented on the walls and panels of the classrooms. In addition, we took into consideration the attitudes and discourses of the teachers in the teachers' room, and we took notes for these aspects out of the school.

2.3. Process

We collected the data of the study from the teachers and students from a vocational school in Ankara. At the beginning of the interview, we informed the teachers and students about the purpose of the research and the process of data collection, and we stated that the data obtained in the research could be used for a scientific study. We arranged the appointments with students and teachers, and then we held the interviews with the teachers and students at appropriate times. We made negotiations when the teachers were free or in the afternoon. We also held the interviews during lunch breaks or after lessons. In this way, we tried to remove the obstacles to the collection of research data.

We asked the questions in the forms to the students and teachersr, and we recorded the interviews with the tape recorder. Two teachers who did not want their interviews to be recorded answered the questions by writing. Then we deciphered and wrote down the data obtained by voice recording. The interviews with the teachers of cultural courses took 125 minutes 36 seconds, while the interviews with the vocational teachers lasted for 63 minutes 51 seconds. The interviews with the teachers took 189 minutes and 27 seconds in total. We prepared a 50-page document as the written copy of these interviews. The interviews with the students took 121 minutes and 7 seconds in total. When we wrote down these records, we obtained a 23-page document. The interviews with the teachers and students lasted for 5 hours 10 minutes 35 seconds, and when we put these negotiations in writing, we drew up a 73-page document in written form. There was no restriction on the answers of the participants in the interviews. The teachers and students made important explanations by expressing their own ideas. In general, we asked the same questions to the participants in the same order. However, we also asked the questions such as “Do you open up a little more, explain what you mean, what exactly do you mean?” in order to get more comprehensive answers from the participants who gave general answers. Thus, we sought detailed and comprehensive information on the performance task from the point of view of teachers and students. We examined the documents of the teachers who shared the evaluation criteria they used in performance evaluation so as to determine how teachers prepared and evaluated the performance tasks. The students did not give any document because they told they submitted their assignments to their teachers.

2.4. Validity and Reliability of Study

Reliability in qualitative research is based on whether the results are consistent with the collected data, rather than whether they give consistent results when applied at different times. Internal validity is related to the fact that the findings reflect the actual situation (Merriam, 2013). Throughout the process, the researcher must act with a critical approach and check to what extent the results obtained reflect the reality (Yıldırım & Şimşek, 2005). Internal validity of the qualitative research can be controlled by means of member control or member inquiry, adequate and appropriate participation in data collection processes, examination of a specialist (Merriam, 2013). While preparing interview forms, first of all, we took expert opinion from the specialist of measurement-evaluation and two teachers not involved in the study for the evaluation of expressiveness of the expressions, the clarity of the questions and the provability of the questions. Thus, we tried to ensure the validity and reliability of the study.

In this study, we have tried to determine the ways of completing the performance task and the opinions of the teachers and students at vocational high schools about the tasks by examining the interviews of the participants, the observations made in the school, the records

of the teachers sharing the documents and the results obtained from the collected data. Then the second researcher checked all the codes, and we decided the ones thought to be faulty together. Since a second person did not do any re-coding of the data, we did not calculate the consensus percentage. In addition, the researcher encoded the randomly-selected interview forms of two teachers and of two students for the second time in two weeks so as to determine the consistency in coding. Consistency in coding for student interviews is 0, 89 while this rate was calculated as 0, 91 for teacher interviews.

2.5. Analysis of the Data

Before starting the analysis, we made an encoding of SxÖy for the students to express the following components: S: class, x: class level, Ö: student and y: the order in which the interview was done. For example, S9O1 means that we firstly interviewed the student in the 9th grade. Similarly, the teachers were coded as Öx: Ö: teacher and x: the order in which the interview was made. For instance, Ö9 means the teacher who was interviewed was in the 9th rank.

We analyzed the data by content analysis. Because of the lack of a conceptual structure according to the purpose of this research, we attempted to reveal a structure as a result of the analysis of the data with an induction approach, and in this way, we made the content analysis. We generated the codes for each question and we arranged the collected data in an inducible approach by determining appropriate themes for these codes. We read the raw data obtained from the students and the teachers and in some cases we generated the according to the meaning of the data. Then, we put together the related codes in order to create themes. In the following phase, the second researcher checked the codes, and we made the final decision on what was believed to be faulty.

3. FINDINGS

We asked 15 questions to the students in order to determine their perceptions on the performance task, and we asked 17 questions to the teachers for identifying their views. We have arranged the answers to these as follows. We have determined the views of the students and teachers on the performance task.

The codes and the themes including the opinions of the students regarding the performance task are presented in the table below.

Table 2. Thoughts of students towards the content

Theme	Code	f
Requests topic / content of the task	Easy for the student to do	15
	The tasks with which students can show their performance	2
	Different	3
Determining the subject	Teachers choose the subject	14
	Sometimes we learners sometimes teachers choose it	3
Who will determine the subject	I can determine the subject	16
Who will determine the subject	The teacher can state the subject	1

In Table 2, 15 students expressed that they wanted the performance tasks to be easy, and understandable, and to focus on more different subjects. Only two students wanted to prepare performance tasks on the topics that they can prove themselves. 14 of the students stated that the performance tasks were determined by the teachers while 16 students remarked that they wanted to determine the subjects by themselves.

Students expressed their demands in the following sentences.

S9Ö3: "Our teachers. I want. For example, everyone is doing the same kind of homework. It is better to do different kinds of assignments."

S10Ö3: "Teachers give us 4-5 topics to choose from. Is that what it usually is? In most of them. Did you want to determine why? Yeah. I choose easier. I choose the subject that i know. I don't have any difficulty in doing research on the subject."

S9Ö4: "Our teachers. ... I was willing to do it if I was better. I wanted to show my own thoughts and abilities."

Codes and themes about the teachers' opinions on content of performance tasks are given in the table below.

Table 3. Points to consider while teachers determine the performance task

Theme	Code	f
The way of detirming content of task	Purpose of the course	1
	The curriculum	5
	The students' level	5
	Students' Interest	1

As seen in Table 3, six of the teachers stated that the performance task's content was based on the curriculum, while eight teachers selected the content by considering the students. Teachers expressed their opinions as follows:

Ö9: "I gave the most easily understood subjects to this year's ninth grade and tenth grade."

Ö11: "According to the level of the class, I gave the subject (9th grade students' subjects are different from 10th grade students' topics) ... For example, I am asking what they can do before they have seen and for example, which topic do you want to study, the students answer and I say okay for the topic. I say to the students: tell me the subject and present it to me. Then, for instance, I told them to solve 20 questions for the performance: question-answer, question-answer."

While the students preferred to do the performance tasks that they chose, it is seen that the teachers took into consideration the needs of the students and the curriculum when determining the subjects.

The table below includes the themes and codes of the student expectations regarding the performance task.

Table 4. The expectations of the students from the performance tasks

Theme	Code	f
Expectation	Mark	11
	Understand the subject	7
	Test what they can do	1
	Admiration of teachers	1

Table 4 shows that the majority of the students expected good scores from the performance tasks, that 7 students expected to learn in a better way, one student aimed to test himself about what a student can do and what he can not do, and one student's aim was to win the approval of his teacher.

The students expressed their expectations from the performance task as follows:

S10Ö1:"Taking a high score and raising marks. Learn the topic"

S11Ö4: "Of course, it is our only goal to be able to pass high grades and pass only that course. Otherwise, nobody has a desire to do performance work."

The codes and themes related to the expectations of the teachers regarding the performance tasks of the students are given in Table 5.

Table 5. Expectations of the teachers from students regarding the performance task

Theme	Code	f
Study	Study	3
	Something to do individually	3
	Learning of the subject	2
	Inquiry	1
Make the students love the subject	Make the students love the subject	1
Responsibility	Be serious	1
	Create a question bank	1

Three of the teachers stated that there was no expectation from the students regarding the performance task. While 3 of the teachers were waiting for the students to do something individually, 2 of them wanted the students to learn the subject. There are teachers who asked the students to make inquiry, made the students enjoy the subject, took the tasks seriously and prepared a question bank. Teachers explained their views on this issue as follows:

Ö1: "I did not receive any results from my homework. While assigning the performance task, my expectation was that student had to learn the meaning of the homework, the study, the research, the research. My expectation was not met because we think of our students as well; they have no infrastructure, no resources in their hands, not a lot of computers""

Ö6: "Actually, I did not have any anticipation while giving the performance task. I just waited for my students to deliver the assignment."

MÖ1: "My expectation is that learner should take it seriously. I would have wanted that the students cared product more than the exam."

When we compare the expectations from the performance task, it seems that both groups are expected to learn the topic. It seems that the priority of the students is the score while the priority of the teachers is to teach the subjects.

The code and themes of the points that the students took into account while doing the performance task were given in Table 6 below.

Table 6. Points about which students are careful when they doing the performance task

Theme	Code	f
Points to note	Physical appearance	17
	Content	5
	Study	2
	Wishes of the teachers	3
	My behaviors	1

As it can be seen from Table 6, the students paid more attention to physical appearances while doing the performance tasks. On the other hand, there are 5 students who expressed that they are attentive to the content. The students expressed their expectations from the performance task as follows:

S10Ö2: "To the layout, cleanliness, content. What do you mean with content? Does it explain the subject better? Are you impressed when you read it or is it a normal homework? That's it."

S12Ö1: "The front cover, the beautifulness of the writing, the bibliography and something else? Nothing else."

S12Ö4: "First of all, I pay attention to the order. The page should be nice, the writing should be nice, the cover should be beautiful. What else? The content should be nice, too, my teacher. Now I can give full information; also, I can give semi-informative information. It is necessary to give the right information in the right place."

The code and the themes of the student opinions on how the teachers evaluate the performance task are given in Table 7.

Table 7. Students' opinions on how teachers evaluated the performance task

Theme	Code	f
Evaluation criteria	Writing beauty and rules	4
	Appearance	14
	Scale	2
	Content	5
	Personality and behaviors against teachers	4
	Exam mark	2
Fair scoring	Taking the rightful mark	8
	Usually taking the rightful mark	6
	Marks are given according to the person	4
	Not taking the rightful mark	1
Scoring results	No surprise if criteria are known	5
	Marks which students took are not surprise	5
	Marks which students took become surprise	3
	Don't know the criteria	5
	Sometimes it becomes surprise	5

As seen in Table 7, the students thought that the teachers evaluated the performance task according to the physical appearance. While 5 of the students said that the performance tasks were evaluated according to the contents of the tasks, 4 of the students said that they were evaluated according to their personality and behaviors against the teacher. Only 2 students said that their tasks were evaluated according to scales. Likewise, two students said that their performance tasks were graded according to the exam grade. One of the students stated that he did not get the score that he deserved in the performance task. There are 6 students who remarked that they received noteworthy scores, and that they have received the scores they deserved. There are 5 students stating that the scores from the performance task were not a surprise, and 5 students stated that they were not surprised if the criteria were known in advance. There are 3 students who said that it was a surprise, 5 students remarked that they did not know the criteria, and 5 students stated that some of the results were surprising and that they knew the scores in certain tasks.

The opinions of the students on the evaluation of the performance tasks are as follows:

S11Ö2: "Sometimes. Now, it is up to teachers. How is it happening? So if the teacher loves you, he/she gives you a lot of mark. If there is a teacher we do not meet, he just looks at the homework; if we have a teacher we meet, he/she who says this student is good, and honest gives high marks. That's a bit of respect for this point."

S12Ö3: "First they are looking at the writing, and if the writing is nice then they are starting to read it. Then they look at the contents, they think we write beautifully in the content and we have good information, and if the content and information are good, we get a nice oral exam mark."

S11Ö3: "It's a bit of a surprise ... I'm getting higher."

S10Ö2: "The teachers don't give points. They give the mark according to the person. I did homework. Our teachers say your marks are low and don't give you a good mark for your performance task. They give the points according to the exam marks. But some of them do not do this and they give the score we deserve.*"*

The following table includes the codes and themes related to the criteria teachers use for evaluating the performance task.

Table 8. Criteria used by teachers to evaluate performance task

Theme	Code	f
Information	Understanding the topic	7
	Content	6
Appearance	Bibliography	4
	Materials	2
	Plan	2
	Presentation	4
Responsibility	Submitting on time	6
According to teachers 'views	In line with the goodwill of the teachers	5
	In line with the active participation in the lessons	1

According to the table given above, 7 of the teachers said that their criterion in the evaluation of the performance task was the students' comprehension ofthe meaning in the task, and they said that they evaluated the tasks according to the content. The main criterion was the bibliography for 4 teachers, material for 2 teachers, plan for 2 teachers, and presentation for 4 teachers. 6 teachers made evaluation according to timely submission and 5 teachers evaluated the performance tasks in line with their own initiatives. Only one teacher said that he/she evaluated the students according to their participation in the lessons.

Teachers expressed their views on this subject as follows:

Ö4: "…preparation process, writing phase and presentation phase. I pay attention to how they create and finalize them."

Ö6: "I used. Bibliography, content, I wanted presentation from some classes; If the material is something visual I paid attention to the timely delivery of the material. I gave them topics. Are there topics in the assignment? I look at them."

MÖ2: "My criteria are that do the students catch the essence of a subject first? Do they realize what they're doing? Do they reflect this?... Then I have time criteria. How soon did they bring it? Then, how is the layout, layout and presentation? What did they gain from this research? I'm asking if I think that the students have worked. I have the information criteria. I have criteria. Of course I do not ask if I notice they do not work."

While evaluating the performance task, the teachers said they attached more importance to the subject, then to the content and then to the appearance. However, since the students thought that the teachers paid more attention to the appearance, it was determined that the students first attached importance to the the physical appearance.

Table 9 contains the codes and themes on how students received help or guidance from their teachers when performing performance tasks.

As it can be seen in Table 9, 11 students stated that the teachers guided them efficiently, 2 students expressed that they weren't guided; 4 students indicated that they did not need help. In addition, 9 students stated that both written and oral explanation were done, 7 students said oral explanation was done, and one student claimed that written explanation was done.

Table 9. Students' views on teachers' guidance during performance task

Theme	Code	f
Whether or not guidance is given	Guidance at enough level	11
	Not enough guidance	2
	Not want	4
Guiding style	Oral	7
	Written	1
	Written or oral	9

The students explained their views on this subject as follows:

S9Ö2: "They give it, help is good. Some of our teachers give the help on paper and some give help orally. "

S11Ö1: "They said it orally. The teachers chose the topic from the book, so we did it by taking the material and searching it by reading it beautifully. So the teachers said and us write it: do the performance task from here and we didn't do anything."

S12Ö4: "No. Now, the teacher says us your homework is this, do it and bring it to me…He/she doesn't give any detailed information… Each teacher doesn't do this…I mean generally oral. … For example; we do performance tasks, everyone comes to the class, the topics are determined according to the class list and the deadline for the homework's are stated. That's all."

Table 10 provides the codes and themes for the guidance of the teachers for their students during the performance task.

Table 10. Guidance of teachers during the performance task

Theme	Code	f
Guidance	Done	8
	Sometimes	2
	Never done/never asked	3

As seen from the table above, 9 teachers said they guided students, 2 of them sometimes guided them, 1 teacher said that he did not intervene in the process because he did not think his guidance would work, and 2 teachers did not offer their assistance because the students did not demand it. Teachers explained their views on this subject as follows:

Ö5: "If you show the subject as practical at the first week of the school, then they do not need guidance."

Ö11: "During the task. I always say to the students that if you do not understand, if there is a place you have difficulty while doing your performance assignment, come and ask me again ... Because the important thing is that students should learn. I mean, I lead my students and this is very good. Teacher must have a leading role anyway. …"

MÖ4: "…Now we already have a problem: I have been in this profession for two years, but the level of students is a bit low, in general the numerical and comprehension skills of the students are miserable, so we have students who do the task wrongly although we tell them how to fulfill the tasks very much, but they definitely need guidance, hence we have to do guidance. We can not get feedback when we are not guiding."

During the performance task, the students generally stated that they were guided efficiently, and the teachers expressed that they guided the students.

The codes and themes of the resources that the students use while doing the performance task are given in Table 11.

Table 11. Resources that the students use while doing the performance task

Theme	Code	f
Resources	Internet	17
	Books	12
	Encyclopedia	5
	Notebook	2
	My sister	3

When asked about the resources they were using while performing the performance task, all of the students answered the question as internet (one student said that s/he used internet only if s/he couldn't find the data in other sources). 12 students stated that they used books, 5 students said that they referred to encyclopedias, 2 students stated that they benefitted from notebooks, and 2 students remarked that they received supports from their siblings.

As it can be understood from these results, students are most likely to benefit from the internet and books while doing their performance tasks. Students describe the resources they use in the performance tasks as follows:

S9Ö1: ''Internet, books, books related to the lesson, my sister.''

S9Ö5: ''Internet, books, notebook.''

The codes and themes related to the problems students encounter when performing performance tasks are as follows:

Table 12. Problems students encounter when performing performance tasks

Theme	Code	f
No problems	No problems	2
Problems encountered	Layout and appearance	13
	Preparing product	5
	Time	3
	Research	5
	Presentation	1

It is seen in Table 12 that the students had the greatest difficulty in preparing the external appearance which was considered to be the most important point for the students in the performance task. There are 5 students who had problems in preparing the materials with which they would present their performance tasks. The students explained the time-related problems by stating that they were supposed to prepare performance tasks for all courses and that the submission period of the tasks coincided with the exam dates. In addition, 5 students said that they had difficulty in finding a subject and reaching the sources while doing research. Students expressed the problems of the performance tasks as fellows:

S9Ö1: "I have encountered difficulties with external appearance. In writing, the cover is intended to correct the picture when I make the picture, to do it properly."

S10Ö1: "…Time is becoming a problem. There are exams, so it's been trouble when there are exams."

The themes and codes of the teacher's views on the nature of the problems encountered during the preparation phase of the performance task are given in Table 13.

Table13. Problems of performance task according to teachers

Theme	Code	f
Student-oriented	Do not study by trusting the performance task	5
	High mark expectation	1
	Irresponsibility	3
	Doing carelessly	2
	Do not submit on time	5
	Bringing printout from the internet	1
	Do not understand- do not have the ability to do so	3
	Make someone else do (not themselves)	4
System-oriented	Giving performance task to each student from every lesson	5
	Giving performance task to one who doesn't want to do it	1
Teacher-oriented	Work load	4
	Different shape	2
	Evaluation	1
	Time	1
Nothing	Nothing	3

As observed in Table 13, 5 of the teachers thought that giving performance tasks to every student was a problem. One of the most common problems during performance tasks was that students did not study by relying on the scores to be obtained from the performance tasks and that they did not submit the tasks in a timely manner. The teachers stated that the students did not do their tasks and that the performance tasks were a burden on teachers. Teachers expressed the problems of the performance task as follows:

Ö1: "The fact that every student is given a performance duty and that their deadline is also the same date as the exams, so the student is doing the homework carelessly and I see this as a negative side. At the same time, when he gets his performance duties from all the courses, he makes his sister, brother or a friend do it, or prints it out on the internet."

MÖ3: "I do not think there is anything negative about it. I think the negative aspect of the performance task is misunderstanding. It is understood by both students and teachers as if it is a term paper. In other words, it is understood that the student has to bring out a document or business with the province and bring out the product, which is the negative side."

Ö9: "It is workload on a teacher if I think of it from my own perspective. The paper consumption was incredible and students say, for example, I have practiced almost every performance task, so I didn't do my homework, I did not study the writing, I did not have my homework, because I had my homework, but they do not do the homework properly. They do not study properly in the classroom, so when the child makes four or five pages of homework and spends it on the complex internet, they are expecting a high grade from you. From that point of view, I think that the oral note scale of the teacher is better."

While preparing the performance task, the students stated that they often had problems with the appearance of their tasks. On the other hand, the teachers stated that the students mostly did not bring their tasks on time, that they did not understand the tasks, and that they did not study for the lessons since they relied on the performance tasks, etc.

The positive and negative aspects of the performance task according to the students' thoughts are given in the table below.

Table 14. Positive and negative points of the performance task in terms of students

Theme	Code	f
No benefits	No benefits	3
Benefits of the performance task	Mark	10
	Studying and learning	14
	Research	1
	Improving the skills of writing and drawing	2
	Self-recognition	1
Negative sides	Doing homework from each lesson	2

There are 3 students who said that the performance task had no benefit. 10 students stated that the benefit of the performance task was the scores. When the code and the themes related to the performance task are examined, 14 students indicated that the performance task was beneficial in terms of studying and learning the subject. Some of the students said that they benefited from research, self-recognition and writing skills development (1 student benefited from the research and self-recognition, and two students found the opportunity to improve the writing skills). Two of the students stated that they had difficulties in doing tasks and getting prepared for the exams because they were supposed to do performance tasks in each lesson, and they remarked that there was no negative aspect of the performance task if there was not any time constraint.

The statements of the students on the benefits of the performance task are as follows:

S12Ö3: ''Of course, there are also benefits outside the note, and if we mention about them, we learn different things. Last year I got my performance assignment from mathematics: logarithma, now we study it in the same classroom, but most of my friends have forgotten it but I did not forget.''

S9Ö2: ''There was no benefit. Nothing usually happened, it was a point upgrade. It would not have been very beneficial in terms of learning the subject. Even if ı didn't take a performance task, it wouldn't make a difference.''

S10Ö2: "There is no positive side. I think so. You just said we get the information, the students' marks rise. Another? On the other hand, even if you write directly from the internet, it actually teaches you something... For example, I do not want to get a performance task as ı have high points from it. But we are obliged to have it. At the end, I have high mark and what happens whether I do my homework or not, then I am doing badly and my average decreases."

The code and the themes generated from the teachers' views on the contribution of the performance task to the students are expressed in the following table.

Table 15. Contribution of students to the performance task in terms of teachers

Theme	Code	f
No positive side	No positive side	3
Learning	Completing my shortcomings	5
	Learning individual study	1
	Making research	2
Communication	Improving communication skills	2
	Discovering students	1
Student-centered	Making students active	1
	Measuring high-level skills	1
Document	Document for oral exam	1

As shown in Table 15, three teachers thought that there was no positive effect of the performance task. While 5 of the teachers indicated that the students understood the topic better with the help of the performance task, 2 teachers expressed that they started to do research and to learn how to do it. In addition, a teacher said that the students learned to work individually and to produce an individual product. 2 teachers indicated that the performance tasks allowed for student-centered assessment. One teacher remarked that the performance tasks enabled them to obtain written documents for giving the scores of oral examination. The contribution of the performance task to the students is expressed by the teachers as follows:

MÖ2: "The positive side is that as I said, if they come to such curious students who are willing, they are really trying, learning. They come during this research, asking what they do not understand, what else they can do ..."

Ö2: "Positive aspects, as I said, it shows us silent students, their different features, creativity that we do not fully see and the different thoughts of them during teamwork. Also, students can prove their existence with these homework."

Ö6: "The student realizes his / her shortcomings when he / she makes a performance assignment. He/she realizes the missing aspects and tries to complete them... At least they do research and realize where they can use the information gained from the homework in everyday life."

Getting high scores and learning the topics were the positive aspects of the performance tasks from the students' points of views while the teachers stated that the main positive aspects of the tasks were the acceleration of the students' learning process and the development of their communication skills.

The code and the themes of the student opinions on whether the performance task is necessary are given in Table 16.

Table 16. Necessity of performance task according to students

Theme	Code	f
Whether the performance task should be or not	It shouldn't be	3
	I should take it from the lessons that I want or I am bad at	6
	Better if it is	8

As seen in the table above, 3 students said that there should be no performance task. 6 students expressed that they should do the tasks for the lessons that they wanted or in which they failed. 8 students stated that the performance tasks should be given. Five of them said that the performance tasks were beneficial because these tasks increased their scores. A student stated that the tasks contributed to the learning process since they enabled him/her to revise the subjects. Another student said the performance tasks were necessary since they gave responsibilities to the students.

S10Ö4: ''I think it should be like the project… I think that if you are not good at a lesson, you can take homework from it and prepare it better and pay more attention to upgrade that course mark. It may be from the lessons we want. So free. It isn't similar to the project, I mean you can take homework as much as you want…''

S12Ö2: ''Actually, we should go to 12 this year, so I do not need performance work from my point of view. However; in my opinion, the performance task should be for class 9, because when the students prepare for university, they will not forget the information and it will help them.''

The codes and themes on the necessity of performance tasks according to the teachers are given in Table 17.

Table 17. The necessity of performance task according to teachers

Theme	code	f
Necessity	Unnecessary	6
	It should not be given to every student	2
	It should not be at the same time with the exams	2
	Necessary	1

Table 17 shows that six of the teachers found the performance tasks unnecessary, only one teacher thought that the tasks were necessary. Two teachers remarked that the tasks should not be given to all students, and two teachers stated that the submission period should not coincide with exam dates. Teachers' expressions about the performance task are as follows:

Ö8: "I think it is unnecessary because I follow them in the course of time for example, I am giving to + according to the attendance to the course and if the student answers a question. I check their notebooks and books."

Ö11: "I think it is unnecessary. Then why is it given? Actually, I do not understand why it is given. Because you are giving to your students something, he /she has not already done it properly… For example, you ask one of the questions which the student finds but he/she cannot answer any of the questions s/he finds. Because he/she just wrote something to take notes and brought something only for show not to learn."

MÖ2: "I cannot say it is necessary but I can say that it is useful. So it is not necessary to say, but it would be useful if this job is done voluntarily."

The majority of the teachers participating in the study thought that the performance tasks should not be used, while the majority of the students thought that they should be given. The codes and the themes on how the teachers defined the performance tasks are given below.

Table 18. Definition and the aim of the performance task according to teachers

Theme	Code	f
Learning the given topic	Study	2
	Understanding	2
	Struggling	1
Evaluation	Broad evaluation	3
	Reflect His/ her learning	3
	Process	3
	Instead of an oral note	3
Creativity	Developing creativity and adapting to the real life	1
Responsibility	Developing responsibility	2

The teachers were asked to explain what the performance task was and what the purpose of these tasks might be. Five of the teachers stated that their students studied and learned the topics. Twelve of the teachers said that they used performance assessment to evaluate students. Teachers indicated that the performance based assessment under the evaluation theme allowed for a wider assessment and reflected students' learning. 3 teachers stated that the performance task provided a written document for oral examination and it ensured the participation of the students in the class.

Ö1: "I think that the task of performance should be the mission that brings the creativity of a studend to the forefront. At the same time, it is useful in terms of understanding the parts that the student does not understand about the course."

Ö3: "The performance task is the reflection as the performance of what students learn in their own way, beyond what they have learned in class, or actually what they have learned

in class. So when we look at the meaning of performance, it means that students have to make an effort to do about that lesson."

Ö5: "Activities that cover a period. It should not be just homework."

The following table shows the code and themes on how the teachers understand whether the students do their performance tasks by themselves or not.

Table 19. How teachers determine whether students do their performance task themselves or not

Theme	Code	f
I do not understand	I do not understand	2
The method of understanding whether the student has done the homework himself/herself	Asking	7
	I made students have presentation	1
	I know the student	5
	Performance tasks are the same or similar	2
	I assess the process	1
	I listen to the students one by one	1
	I give tasks which somebody else cannot do.	1

The most common way for the teachers to understand whether the students did their performance tasks by themselves or not was to ask the students the afore-said question directly. Five of the teachers stated that they could understand whether the student did the task because they knew the students, one teacher stated that he/she made them present their tasks, two of them said that they checked the similarity among performance tasks, one teacher said that he assessed the process so the students could not have his/her performance task done by someone else.

Ö3:"…For the homework I gave, somebody else cannot do that…Because we have already told about it in the classroom. He/She will comment on what we talked about. Students can make anybody commented on what I'm talking about…."

Ö7: "For example, homework was delivered in a perfectly fine, uniform, well understood manner… I said you did not do this homework. Student already said "yes, my teacher I did not do that" He/She confesses. He/She confessed because his/her writing is not so smooth, I know he/she did not understand the topic. You have already know the student."

MÖ1: "It is already emerging from the student's success. What he can or cannot do…"

The codes and themes related to the responsibilities and the performance task according to the teachers' point of view are given in the table below.

Table 20. Responsibilities and tasks related to performance tasks according to teachers

Theme	Code	f
Teacher	Being a guide	4
	Checking	3
	Determining appropriate content and standards according to my students level	4
	Take it seriously	1
Student	He /She should be responsible	3
	He /She should cooperate with the teacher	2
	Take it seriously	2
	Study	1
Parents	They will follow	6
	They will support	4
	They will provide resources	2
	Not doing homework	1

Four of the teachers indicated that the teachers should guide the students about the performance tasks; four of them said that teachers should determine appropriate subjects and standards for students. In addition to this, three teachers said that performance tasks should be checked by the teachers, one teacher said that the tasks should be taken seriously. Teachers stated that the students should be responsible for their performance tasks, that they should cooperate with their teacher, and that they should carry out study on the topic. The teachers stated that the the parents should be responsible for the follow-up of the performance tasks of their children, that they should support them, provide resources and not intervene in their homework.

Teachers expressed the responsibilities and duties related to the performance task as follows:

Ö7: "The task of the students, the student is fully aware of the task and he /she will begin immediately to investigate. They will led their teachers constantly check their homework… He/She will always keep in touch with the teacher. Parents at least, our parents are a little uninterested in this issue… If the parents knew, followed their child, warned the child…

Ö10: "…Does the performance tasks provide student with a benefit or not? Does the student enjoy or not? We need to talk about these aspects…We cannot give it exactly because we do not know each student very well. We have lots of students."

MÖ1: "…the parents have to stop to entrust their student to school. Teacher, school and parents have to act together *…"*

The codes and themes regarding the teachers' proposals to make the performance task more effective are shown in the following table.

Table 21. The proposals of the teachers about the performance task

Theme	Code	f
Topics	Related to the Daily life	1
	Developing creativity	1
	Suitable for the level of the learner	4
	The one which the students will like	2
	Student determine the topic	1
	Teacher teaching the same lesson determine the topic	1
	It should add something to student	1
Time and Place	It should been spread broader time	1
	There must be appropriate environment and facilities	1
Responsibility	There must be family support	1
	It should be taken seriously	1
	Good guidance should be done	1
Education System	Education policy must be changed	1
	Exams should been removed	1
Obligation	It should not been given the students who do not want	2
	It should be removed	3
Method	Process can be evaluated in the course	2

In Table 21, most of the teachers indicated that the selection of the topics for the performance tasks should be taken into consideration in order to improve the quality of the performance task and to make it more effective. In addition to this, the teachers thought that the performance tasks could become more efficient with the help of the improvements in the conditions and physical facilities of the school, the allocation of broader time periods for the performance tasks, and the fulfillment of the responsibilities by the families and teachers. Apart from that, there are teachers who said that the educational policy should be changed in order to

increase the functionality of the performance task. Moreover, it was indicated that the performance task should not be mandatory and that process assessment in the course was also a performance evaluation. Three teachers said the performance task should be removed from the evaluation process.

Teachers' suggestions for performance tasks are expressed as follows:

Ö10: " ……So you increase the effectiveness of the performance task in this way, all students are included in, there should be suitable environment to enjoy. .. I have no factory. I have no place to do this in my lab…"

Ö9: " …Group teachers who teach the same lesson must come together and set a standard, draw a sketch. Secondly, maybe… he studies can be done about what your students want to do, what they are expecting… As teachers we have been pushing performance assignments for two weeks to three weeks. Maybe it is more likely that we will give it at the beginning of the year and spread it over a wider period of time, students can be expected to do something more efficient…"

Ö7: "If I think of it in the name of our school, I do not think that it will be implemented properly. For example; in our school, performance task should completely be removed... The children assume that they will get 100 when they take a colorful piece of paper and paste two pages on it. In order that this work properly, the learner needs to change, the student must change his mind…"

In addition to these findings, 10 teachers who participated in the study stated that the performance task did not reach its goal, only 3 teachers' stated that it reached its goal.

4. DISCUSSION, CONCLUSION, SUGGESTIONS

This stuy examines the opinions of the students and teachers at a vocational high school about performance tasks. According to the result of this study, the expectations of the students from performance task are to have high grades and to understand the subject. Although there have been teachers who do not have any expectations from the students regarding the performance task, it has been determined that the teachers generally expect the students to study, to enjoy the course, and to take responsibility.

The teachers have stated that they assess the performance task firstly according to the content and secondly according to the physical appearance. Most of the students, on the other hand, have remarked that the teachers assess the performance task according to the physical appearance, and that they do not give grades fairly. Thus we have concluded that the students pay attention to the physical appearance rather than the content (Çetin & Çakan, 2010) and it is the most difficult part for them.

Similar to the results of the study carried out by Yılmaz and Benli (2010), we have concluded that the teachers guided the students sufficiently, and that they helped the students while doing the performance tasks.

In addition to these, the performance tasks have many advantages from the perspective of the students. The most positive aspect of these tasks is to encourage the students to study the subject and to learn it. The grades that the students get are other positive aspects of the performance tasks. We have found out that the positive aspects of the performance tasks from the perspective of the teachers are the acceleration of students' learning process and the development of their communication skills.

We have also concluded that the students generally prefer much easier subjects they can do for the performance tasks. On the other hand, the teachers mostly take into consideration the needs and levels of the students and the content of the lessons while determining the subject of the performance tasks. These findings are similar to the findings of the study carried out by

Yılmaz and Benli (2011) in which the interest, motivation and the needs of the students are taken into consideration while determining the subjects of the performance tasks.

We have determined that the students have problems in making the physical appearance of the task proper, while the teachers state that they encounter student-based problems such as irresponsible behaviors, not understanding the task and the topic, and having the tasks done by someone else. The teachers also perceive the tasks as an extra workload for them. The problem of having the performance tasks done by someone else is also seen in the studies carried out Yılmaz and Benli (2011). This study also has similar results with the studies carried out by Duruhan and Çavuş (2013) in terms of doing the performance task for the sake of getting a high grade and downloading and getting it printed from the internet.

Generally, the students have a positive attitude towards the performance task. This finding is parallel to the findings of the study with the primary school students by Ayva (2010). The teachers express that the performance task have not reached its goal. These results also show similarities to the results of the study carried out by Duruhan and Çavuş (2013).

The results of this study put forward the fact that both the students and the teachers do not have enough information about what the performance task is and how it should be done; therefore, the performance tasks have not reached their goals. Performance tasks should measure higher order thinking and they should be related to real life situations. The results of the study show that performance tasks do not possess these features. Moreover, the students should have knowledge and understanding of the subjects in order to do performance tasks. Students try to do a performance task about a subject which mostly they do not understand or they do not know. This may be a reason why performance tasks do not reach its goal. Like other alternative assessment methods, performance tasks make the students active in the process and enable them to take more responsibilities. But in the current study, teachers complain about irresponsibility of students, and most of them say that performance task should not be given to all students. They even say that the performance tasks should not be used for the evaluation of the students anymore. Teachers' lack of knowledge about performance task may be the reason of these opinions. In general, the habits of the students which they bring from their primary and secondary school, the insufficient experience and knowledge of the teachers about the performance task can be the most important reasons of these problems mentioned in the study. The applied education can be given to both the students and the teachers in order to eliminate these problems. Similar studies may be carried out at different high schools.

Acknowledgments

This study was presented at Vth Measurement and Evaluation Congress as an Oral presentation

ORCID

Ayşegül Altun https://orcid.org/0000-0002-8046-2242
Hülya Kelecioğlu https://orcid.org/0000-0002-0741-9934

5. REFERENCES

Akay, Y. & Küçükkaragöz, H. (2016). Öğrencilerin Performans Puanları ile Başarı Puanları ve Tutum Puanları Arasındaki İlişkinin İncelenmesi [To determine of the correlations between the scores of students' performance tasks with their mathematics achievements and attitudes]. *Uluslararası Hakemli Sosyal Bilimler E-Dergisi*, *58*, 539-548.

Ayva, Ö. (2010). *Sosyal Bilgiler Dersi Öğrenme Öğretme Süreci İle İlgili Öğrenci Görüşleri* [Students' views on social study lesson's teaching- learning process]. International Conference on New Trends in Education and Their Implications, Antalya-Turkey.

Bal, A. (2012). İlköğretim Öğrencilerinin Matematik Dersi Performans Görevi Hazırlama Sürecine İlişkin Görüşleri ve Yaşadıkları Sorunlar [Primary School Students' Views and

Challenges on Performance Task Preparation Process in Mathematics Course].*Pegem Eğitim ve Öğretim Dergisi, 2*(1), 11-23.

Başol, G. (2015). *Eğitimde Ölçme ve Değerlendirme* [Measurment and evaluation in education]. (3. Baskı) Ankara: Pegem Akademi Yayınları.

Berberoğlu, G. (2007). *Sınıf içi ölçme ve Değerlendirme Teknikleri* [Classroom measurement and evaluation techniques], Ankara: Morpa Yayınları

Birgin, O., & Baki, A. (2012). Sınıf Öğretmenlerinin Ölçme-Değerlendirme Uygulama Amaçlarının Yeni Matematik Öğretimi Programı Kapsamında İncelenmesi [An investigation of the purposes of measurement and assessment practice of primary school teachers within the context of the new mathematics curriculum]. *Eğitim ve Bilim, 37*(165), 152-167.

Çetin, M.O., & Çakan, M. (2010). Fen ve Teknoloji Dersi Başarılarının Farklı Yaklaşımlarla Ölçülmesi ve Bu Yaklaşımlara İlişkin Öğrenci Görüşleri [Measurement of science and technology course achievement by different approaches and student opinions about these approaches]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(2), 93-99.

Çiftçi, S. (2010). İlköğretim birinci kademe 4. ve 5. Sınıf Öğretmenlerinin Performans Görevlerine İlişkin Görüşleri [The opinions of the teachers in upper primary classes concerning the student performance tasks]. *İlköğretim Online, 9*(3), 934-951.

Dağhan, G., & Akkoyunlu, B. (2005). Bilişim Teknolojileri Dersinde Kullanılan Performansa Dayalı Değerlendirme Yöntemlerine İlişkin Nitel Bir Çalışma [A qualitative study about performance based assessment methods used in information technologies lesson]. *Kuram ve Uygulamada Eğitim Bilimleri, 14*(1), 1-17.

Doğan-Temur, Ö., Bayındır, N., & İnan, H.Z. (2010). Öğretmenlerin Performans Görevlerinin Verilme Gerekçelerine İlişkin Algı Düzeyleri [Teachers' perspectives' on the reasons for giving students performance tasks]. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 7*(14), 247-258.

Duran, M., Mıhladız, G., & Ballıel, B. (2013). İlköğretim Öğretmenlerinin Alternatif Değerlendirme Yöntemlerine Yönelik Yeterlik Düzeyleri [The competency level of elementary school teachers' towards the alternative assessment methods]. *Mehmet Akif Ersoy Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi, 2*(2), 26-37.

Duruhan, K., & Çavuş, A.(2013). İlköğretim 7.Sınıf Sosyal Bilgiler Dersi Performans Görevlerinin Etkililiğine İlişkin Öğretmen Görüşleri [Views of teachers on the effectiveness of primary education 7th grade social science lesson performance tasks]. *İlköğretim Online, 12*(4), 949-968.

Güneş, P., & Soran, H. (2013). İlköğretim İkinci Kademe Öğrencilerinin Dereceli Puanlama Anahtarına İlişkin Görüşleri [Secondary school students' opinions on rubrics]. *Kastamonu Eğitim Dergisi, 21*(4), 1327-1344.

Güvey, E. (2009). *İlköğretim 1-5 Öğretim Programlarında Yer Alan Proje ve Performans Görevlerine İlişkin Öğretmen ve Veli Görüşleri* [Teacher and parent opinions on project and performance tasks in primary education 1-5 teaching programs]. Yayımlanmamış Yüksek Lisans Tezi. Eskişehir Osman Gazi Üniversitesi, Sosyal Bilimler Enstitüsü, Eskişehir.

Hastürk, H. G., & Özkan, B. (2015). 7.Sınıf Öğrencilerinin Fen Bilimleri Dersi Performans Görevlerine İlişkin Tutumlarını Etkileyen Değişkenler [The investigation of seventh graders' attitudes and opinions towards the performance tasks used in science course]. *International Journal of Humanities and Education, 1*(1), 122-140

Kan, A. (2007). Performans Değerlendirme Sürecine Katkıları Açısından Yeni Program Anlayışı İçerisinde Kullanılabilecek Bir Değerlendirme Yaklaşımı: Rubrik Puanlama Yönergeleri [An alternative method in the new educational program from the point of performance based assessment: rubric scoring scales]. *Kuram ve Uygulamada Eğitim Bilimleri, 7*(1), 129-157.

Karaçam, S. (2013). İlköğretim Öğrencilerinin Farklı Formattaki Performans Görevlerine İlişkin Görüşleri [Views of elementary school students related performance tasks in different formats]. *Bartın Üniversitesi Eğitim Fakültesi Dergisi, 2*(2), 235-266

Kaya, S., Karaçam, S., Eş, H., & Tuncel, M. (2013). 4. ve 5. Sınıf Öğretmenlerinin Fen ve Teknoloji Dersinde Proje ve Performans Görevlerine İlişkin Görüşleri [The 4th and 5th grade primary teachers' opinions related to project and performance tasks at science and technology course]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 33*(1), 187-201.

Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2014). *Ölçme ve Değerlendirme: Performans ve Portfolyaya Dayalı Durum Belirleme* [Measurement and evaluation: performance and portfolio based assessment] (4. Baskı) Ankara: Pegem Akademi Yayınları.

MEB (2013), Milli Eğitim Bakanlığı Orta Öğretim Kurumlar Yönetmeliği [Ministry of national education regulation on secondary education institutions]. Retrieved from http://www.resmigazete.gov.tr/eskiler/2013/09/20130907-4.htm

Merriam, S. B. (2013). *Nitel Araştırma: Desen ve Uygulama için Bir Rehber* [Qualitative research: a guide to design and implementation] (S. Turan, Çev. Ed.). Ankara: Nobel Yayınları.

Neukrung, E.S., & Fawcett, R. C.(2010). *Essentials of Testing and Assessment: A pratical guide for counselors, social workers and psychologists*. Brooks/Cole Cengage Learning, USA.

Russel, M.K., & Airasian, P. W. (2001). *Classroom assessment: Concepts and Aplications(7th Ed.)*. Mc Graw Hill Companies. New York.

Sayhan, A. (2017). Ortaöğretim Öğrencilerinin Performans Görevlerine İlişkin Görüşleri [View of the secondary school student related to performance task]. *Elektronik Sosyal Bilimler Dergisi*, *16*(60), 60-75.

Wright, R. J. (2008). *Educational Assesment. Test and Measurments in the Age of Accountability*. United States of America: Sage.

Yıldırım, A., & Şimşek, H. (2005). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri* [Qualitative research in social sciences]. (5. Baskı), Ankara: Seçkin Yayıncılık

Yılmaz, M., & Benli, N. (2011). İlköğretim I. Kademede Verilen Performans Görevlerinin Veli Görüşleri Açısından Değerlendirilmesi [Evaluation performance task given in the first level of elementary school in the aspects of parents' views]. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 8*(16), 287-302.

Yiğit, F., & Kırımlı, B. (2014). Türkçe Öğretmenlerinin Alternetif Ölçme ve Değerlendirme Yöntemlerini Uygulama Biçimleri ve Uygulamada Karşılaştıkları Problemler [Turkish course teachers' forms of implementing alternative assessment-evaluation methods and problems encountered while implementing these methods]. *Electronic Turkish Studies*, *9*(3), 1621-1639.

International Journal of Assessment Tools in Education
2018, Vol. 5, No. 3, 567–592

DOI: 10.21449/ijate.445919

Published at http://www.ijate.net http://dergipark.gov.tr/ijate *Research Article*

A Systematic Review Research: *'Mathematics Anxiety'* in Turkey

Vesile Alkan [1,*]

[1] Pamukkale University, Faculty of Education, Kınıklı Campus, Denizli, Turkey

Abstract: The aim of this study was to conduct a systematic review research concentrating on studies regarding *'Mathematics anxiety'*. 59 papers were reanalyzed in order to answer the questions derived from the main aim in terms of the rules of systematic review method. These studies were reviewed and analyzed by taking account of their aims, designs, sampling and results. While doing this, the similarities and differences of the reviewed studies were also found. It was seen that most of these studies conducted with middle school students. It was also found that the majority of reviewed studies were designed according to quantitative approaches. The review shows that mathematics anxiety is resulted from students' achievement, self-efficacy and fear along with parents' and teachers' lack of supports in mathematics. The results gathered from reviewed studies suggest that studies designed with approaches like qualitative and mixed-method and studies focused on various topics related with *'Mathematics anxiety'* with different sampling are needed. In conclusion, this systematic review study provides some fruitful information for the area and so for the further studies.

ARTICLE HISTORY

Received: 02 April 2018
Revised: 12 July 2018
Accepted: 18 July 2018

KEYWORDS

Mathematics Anxiety,
Systematic Review,
Causes of Mathematic Anxiety,
Reducing Mathematics Anxiety

1. INTRODUCTION

> *"I fear in mathematics. Mathematics is really difficult to me. I can't ask anything to my teacher when I can't solve a problem. I am not sure he can help me, yes maybe he can. […] But firstly, I don't want my friends to laugh at me at my mistakes and secondly, I don't want my teacher to embarrass me in front of my friends. […] if I ask him, he may think that I am lazy, but I only can't understand mathematics."*
>
> *Alkan (2009)*

As in quotation above, some of the students feel fear and so anxiety in mathematics. This apprehension diminishes on the one hand students' success in mathematics on the other hand their interests in doing mathematical operations. Anxious students feel that they are incapable of solving problems or finding solutions in mathematics.These feelings might lead students to avoid physically and mentally attending mathematics which then create mathematical handicapped students.

CONTACT: Vesile Alkan ✉ vesile@pau.edu.tr Pamukkale University, Faculty of Education, Kınıklı Campus, Office A0205, Denizli, Turkey

ISSN-e: 2148-7456 /© IJATE 2018

Since 1950's, anxiety in mathematics has been considered as a problematic situation in educational settings. In those years both Gough (1954) and Dreger & Aiken (1957) defined anxiety as a negative emotional reactions to mathematics. Also, they describe how number anxiety was differed from general anxiety. Thereafter, there have been a wide body of research focusing on anxiety in mathematics. This is not just because of mathematics' being crucial element of the school curriculum but also its being an essential element of life.

Mathematics anxiety described as a negative feeling response to mathematics (Ashcraft, 2000; Gresham, 2010; Maloney & Beilock, 2012 and Richardson & Suinn, 1972). These negative feelings are caused by poor performance and lack of comprehending notions in mathematics. Mathematics anxiety created as a result of these feelings can have an impact on all age groups from primary school students to adults (Alkan, 2013a & 2013b; Ashcraft, 1995; Cemen, 1987; Wu, Willcutt, Escovar & Menon, 2014).

In spite of an increase in technology and new techniques in order to improve learning and teaching in mathematics, it is noticed that students are still anxious in mathematics. Due to this, studies are investigating *'mathematics anxiety'* in Turkey as in other countries. It is seemed that most of the studies conducted in Turkey identified the level of mathematics anxiety (Aydın, 2011; Bekdemir, 2009; Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Dede & Dursun, 2008; Kaçar & Sarıçam, 2015; Karadeniz & Karadağ, 2014; Peker & Şentük, 2012; Taşdemir, 2013; Uysal & Selışık, 2016 and Ünlü, Ertekin & Dilmaç, 2017). Additionally, some of the studies focused on finding out the relation between mathematics anxiety and attitudes and self-efficacy (Adal & Yavuz, 2017; Akın & Kurbanoğlu, 2011; Doruk, Öztürk & Kaplan, 2016 and Kurbanoğlu & Takunyacı, 2012). Apart from these studies, some of the other studies adapted mathematics anxiety scales into Turkish (Akçakın, Cebesoy & İnel, 2015; Akın, Kurbanoğlu & Takunyacı, 2011; Baloğlu, 2010 and Baloğlu & Balgalmış).

Considering Turkish context, even though several studies were done about *'mathematics anxiety'*, there is a lack of studies concerning systematic review of studies about *'mathematics anxiety'*. It is expected that a systematic review of research about *'mathematics anxiety'* can be helpful for the researchers to see the scope of research done about this topic. It is also expected that the results of this study can provide fruitful information to the area. In addition to this, another expectation is that the results of this study can enable researchers to design their studies with different approaches. Furthermore, it is expected that the results presenting the aim of the previous research can provide new perspectives to the future researchers.

In this study, it was aimed to review previous studies focused on *'mathematics anxiety'* conducted in Turkey systematically. To this aim, following questions were answered:

1) What are the main aims of the studies?
2) What type of methodology is used in the studies about mathematics anxiety?
3) Who are the main participants of the studies reviewed?
4) What are the main outcomes?
5) What are the main similarities and differences of the studies reviewed?
6) What needs to be done in the area reviewed?

The questions formed according to the main aim of this study were answered according to the rules of systematic review research. By this review, the main similarities and differences of the studies about *'mathematics anxiety'* were identified. In addition to this, the research which differed from the broad of previous research related with *'mathematics anxiety'* was determined.

2. METHOD

This study aimed to investigate the previous studies related with *'mathematics anxiety'* considering their aims, designs and results. Therefore, a systematic review research method was used to review previous studies focusing on *'mathematics anxiety'* done in Turkish context. A systematic review is a research method comprising of selection of previous studies related with the decided topic, critically evaluation of these studies and analysis of all relevant studies in terms of systematic rules (Millar, 2004; Littell, Corcoran & Pillai, 2008 and Torgerson, 2003). In this study, it is aimed to select and evaluate research about the topic *'mathematics anxiety'* conducted in Turkey. Consequently, studies investigated topics related with *'mathematics anxiety'* are re-examined by using systematic review process.

According to Millar (2004), while using systematic review research, a researcher should follow such phases. These phases are; a) determination of the aim, b) selection of the studies which are consistent with the aim founded on particular criteria, and c) drawing inferences from the gathered information. In this study, all these phases are paid attention and the process is carried out based on these.

In terms of systematic review rules and based on research questions, primarily 117 studies published between the years 2007 and 2017 were downloaded. These studies were downloaded through such databates as ERIC, EBSCOHost, ULAKBIM and Google Scholar. In order to reach papers through the given databates, the keyword *'matematik kaygısı'* in Turkish and *'mathematics anxiety'* in English is searched.

It should be noted that theses, books, projects and conference papers are not included to the searching process thereby, to the review process. After the completion of downloading papers published in journals, those having only abstracts were excluded. In addition to this, even though a lot of papers were listed after the search, it was seen that some of them were not related with the main topic. Therefore, those not having the words *'mathematics anxiety'* and *'matematik kaygısı'* in the keyword part and not particularly related with the main topic were also excluded.

The papers' being directly related with *'mathematics anxiety'* was the significant criterion in the selection of papers for this study. In other words, papers focusing on such topics as *'mathematics teaching anxiety'*, *'test anxiety in mathematics'* and specifically *'attitudes in mathematics'* were excluded. In addition to this, although some papers both titled with the term *'mathematics anxiety'* and included the term *'mathematics anxiety'* in the keyword part, they were excluded from the review process as they were not focusing mainly on mathematics anxiety and not suitable for the aim of this study. Finally, 59 papers were included in the review process.

The aim and the questions derived were decisive in this study. In this context, this study designed in terms of systematic review method followed the criteria below: *These papers*

- published between 2007 and 2017
- focused mainly on mathematics anxiety
- concerned with mathematics anxiety in educational settings
- published in journals

3. FINDINGS

The selected articles included in review process were analysed in accordance with the questions outlined above. Firstly, the aims, designs and participants of studies were given. Then, the results of studies were presented.

3.1. Main Aims of the Studies

Considering the main aims of the previous studies it can be seen that most of the studies focused on identifying the level of anxiety in mathematics (e.g. Aydın, 2011; Bekdemir, 2009; Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Dede & Dursun, 2008; Kaçar & Sarıçam, 2015; Karadeniz & Karadağ, 2014; Peker & Şentük, 2012; Taşdemir, 2013; Uysal & Selışık, 2016 and Ünlü, Ertekin & Dilmaç, 2017). It can be also seen that some studies about identification of anxiety in mathematics targeted determining various issues related with mathematics anxiety. For instance, such issues as *relationship between attitudes and anxiety in mathematics* (e.g. Hacıömeroğlu, 2017; Karadeniz & Karadağ, 2014, Şimşek, Şahinkaya & Aytekin, 2017 and Yenilmez, Girginer & Uzun, 2007), *relationship between some variables and mathematics anxiety* (e.g. Çatlıoğlu, Gürbüz & Birgin, 2014; Doruk & Kaplan, 2013; Taşdemir, 2013; Oksal, Durmaz & Akın, 2013; Peker & Şentürk, 2012 and Uysal & Selışık, 2016), *relationship between anxiety, attitudes and self-efficacy in mathematics* (e.g. Adal & Yavuz, 2017; Akın & Kurbanoğlu, 2011; Doruk, Öztürk & Kaplan, 2016 and Kurbanoğlu & Takunyacı, 2012) *relationship between anxiety and achievement in mathematics* (e.g. İlhan & Öner Sünkür, 2013, Şad, Kış, Demir & Özer, 2016 and Yılmaz & Bindak, 2016), *relationship between mathematics anxiety and classroom assessment* (e. g. İlhan, 2015), *relationship between metacognitive awareness and mathematics anxiety* (e. g. Gökbulut & Akdağ, 2016 and Kaçar & Sarıçam, 2015) and *relationship between mathematics anxiety and learning strategies* (e.g. Arslan, Güler & Gürbüz, 2017) were investigated.

Table 1 indicated the whole picture of the main aims of the previous studies. As seen in Table 1, while some studies focused on finding out the level of mathematics anxiety and its relation with such variables as mentioned above, some of them showed variety of topics. It was seen that there were studies on *adaptation of mathematics anxiety scales to Turkish* (e.g. Akçakın, Cebesoy & İnel, 2015; Akın, Kurbanoğlu & Takunyacı, 2011; Baloğlu, 2010; Baloğlu & Balgalmış, 2010). There was one study focusing on *reducing mathematics anxiety* at primary school level (e.g. Alkan, 2013a) and one study focusing on the *relation between mathematics anxiety and mothers* (e.g. Alkan, 2013b). On the other side there were two reseach investigated the *relationship between mathematics anxiety and mathematics teaching anxiety* (e.g. Peker & Ertekin, 2011 and Ünlü, Ertekin & Dilmaç, 2017).

Table 1. Main Aims of the Studies

Aims of the Studies	Examples
The level of anxiety in mathematics	Baloğlu, 2008; Dede & Dursun, 2008; Aydın, Delice, Dilmaç & Ertekin, 2009; Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Pamuk & Karakaş, 2011; Aydın, 2011; Taşdemir, 2015
Mathematics anxiety and its causes	Bekdemir, 2007; Alkan, 2010; Alkan, 2011; Özdemir & Sezginsoy Şeker, 2017
Relationship between mathematics anxiety and such demographic variables	Peker & Şentürk, 2012; Doruk & Kaplan, 2013; Taşdemir, 2013; Çatlıoğlu, Gürbüz & Birgin, 2014; Uysal & Selışık, 2016
Relationship between mathematics anxiety and attitudes towards mathematics	Yenilmez, Girginer & Uzun, 2007; Karadeniz & Karadağ, 2014; Hacıömeroğlu, 2017; Şimşek, Şahinkaya & Aytekin, 2017
Relationship between mathematics anxiety, attitudes towards mathematics and self-efficacy	Akın & Kurbanoğlu, 2011; Kurbanoğlu & Takunyacı, 2012; Doruk, Öztürk & Kaplan, 2016; Adal & Yavuz, 2017.
Relationship between mathematics anxiety and metacognitive awareness	Kaçar & Sarıçam, 2015; Sarıçam & Ogurlu, 2015; Gökbulut & Akdağ, 2016

Table 1. Main Aims of the Studies (Cont.)

Aims of the Studies	Examples
Relationship between mathematics anxiety and self-regulation	İşleyen, 2015; Yurt & Kurnaz, 2015
Relationship between mathematics anxiety and achievement in mathematics	Şad, Kış, Demir & Özer, 2016
The impacts of mathematics anxiety on mathematical achievement	Bekdemir, 2009; İlhan & Öner Sünkür, 2013
The effects of achievement and social comparison on mathematics anxiety	Kesici & Erdoğan, 2010; Erdoğan, Kesici & Şahin, 2011
The predictive power of students' perceptions of classroom assessment environment for their mathematics anxiety.	İlhan, 2015
Relationship between the achievement in mathematics and mathematics anxiety and test anxiety	Yılmaz & Bindak, 2016
The predictive power of mathematics anxiety and perceived social support from teacher	Erden & Akgül, 2010
The relation between anxiety and achievement in mathematics and achievement in geography	Bekdemir & Başıbüyük, 2011
The predictive power of mathematics anxiety, positive and negative perfectionism to the mathematics achievement	İlhan & Öner Sünkür, 2012
Relationship between Mathematics Teaching Anxiety and Mathematics Anxiety	Peker & Ertekin, 2011
The impacts of teachers' mathematics anxiety on students mathematics achievement	Aslan, Gürgah Oğul & Taş, 2013
Investigation test and mathematics anxiety according to such variables	Oksal, Durmaz & Akın, 2013
Mothers and their relation with mathematics anxiety	Alkan, 2013b
Relationship between students mathematics anxiety and their learning strategies	Arslan, Güler & Gürbüz, 2017

Table 1. Main Aims of the Studies (Cont.)

Aims of the Studies	Examples
Relationships between Mathematics Anxiety, Mathematics Teaching Anxiety, Self-efficacy Beliefs towards Mathematics and Mathematics Teaching	Ünlü, Ertekin & Dilmaç, 2017
The effects of students' basic psychological needs on motivational regulations towards mathematics and mathematics anxiety	Durmaz & Akkuş, 2016
The ways implemented by teachers to reduce mathematics anxiety	Alkan, 2013a
Relationship between self-concept and mathematics anxiety	Işıksal, Curran, Koç & Askun, 2009
Relationship between mathematics anxiety and mathematical beliefs	Hacıömeroğlu, 2013
Relationship between mathematics anxiety and epistemological beliefs	Delice, Ertekin, Aydın & Dilmaç, 2009
The effect of geogebra software on mathematics anxiety and mathematics teaching anxiety	Zengin, 2017
Relationships between students' perception of self-efficacy, mistake-handling learning awareness, and mathematical anxieties	Aksu, Özkaya, Gedik & Konyalıoğlu, 2016
The effect of a "Geometry Garden" on mathematics anxiety	Kurt & Özel, 2013
Classification of students' mathematics anxiety according to the PISA 2012 results.	Erten Tatlı, Atalan Ergin & Demir, 2016
Adaptation of MARS-SV into Turkish	Baloğlu, 2010
Adaptation of MARS-E into Turkish	Baloğlu & Balgalmış, 2010
Adaptation of R-MARS into Turkish	Akın, Kurbanoğlu & Takunyacı, 2011
Adaptation of MAS-R into Turkish	Akçakın, Cebesoy & İnel, 2015
Validity and Reliability Study of MASS	Özdemir & Gür, 2011

When the reviewed papers are considered, it can be said that most of the studies tried to determine whether mathematics anxiety is related to gender, type of the school, attitudes toward mathematics and self-efficacy. In addition to this, there are few studies investigated the reasons of mathematics anxiety. Another point gathered from the reviewed papers is that there is not any study about developing a scale between 2007 and 2017. On the other hand, there are four studies focused on adaptation of such scales used in other countries.

3.2. Types of Methodology Used in Studies

The studies reviewed in this study indicated that a majority of these studies used quantitative approach. These quantitative studies designed according to the survey method as illustrated in Figure 1 below. Beside this, it can be seen that very few studies used qualitative and mixed method approaches compare to those quantitatives.

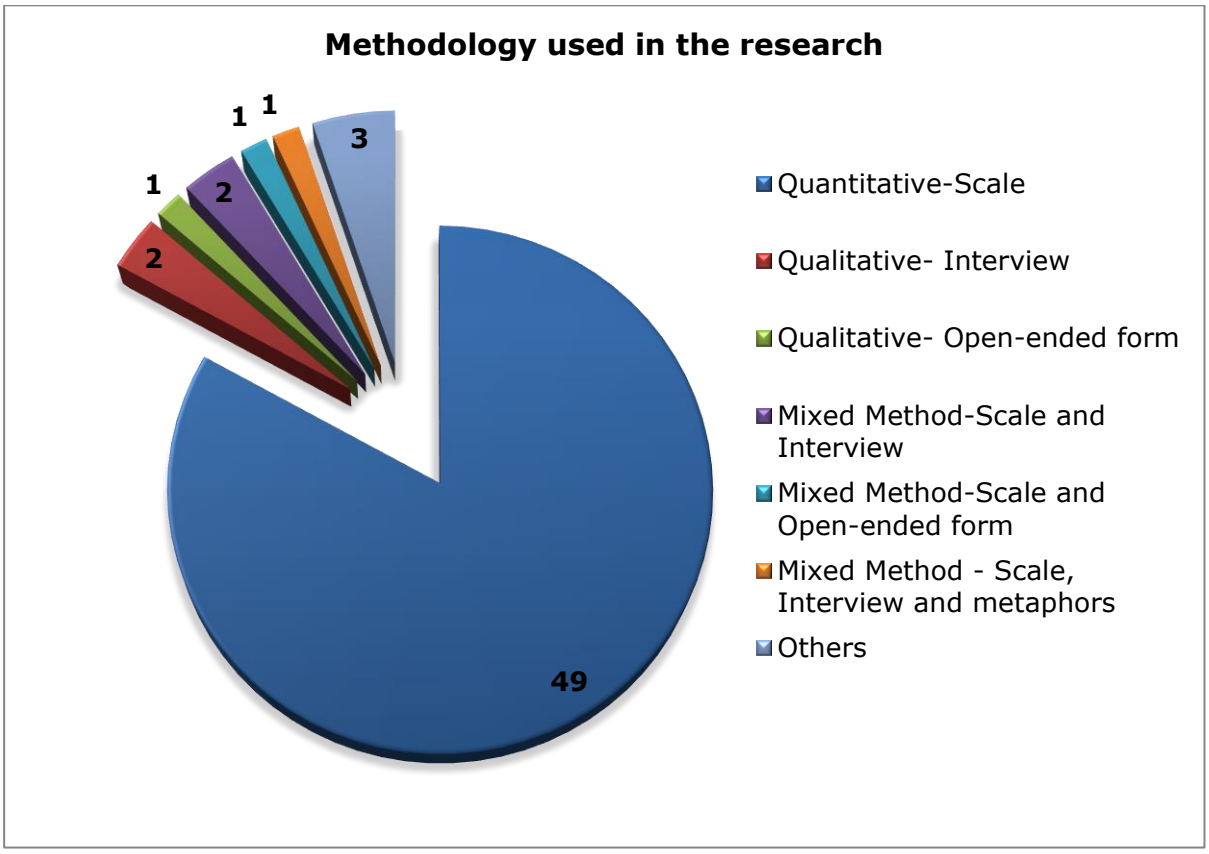


Figure 1. The distribution of the type of methodology in studies

As said earlier and as given in Table 2, it can not be said that studies were used variety of different approaches. In other words, papers were mostly designed according to quantitative methodology. Considering the data collection tools used in these quantitive studies, it is seen that scales are used by most of them. The majority of researchers used mathematics anxiety scale whereas some of them collected the data with attitude or self-efficacy scale beside mathematics anxiety scale.

The result of this review indicates that multiple data collection tools were used in recent studies (e.g. Aslan, Gürgah Oğul & Taş, 2013; Hacıömeroğlu, 2017; Ünlü, Ertekin & Dilmaç, 2017). This means that some researchers tried to find out the relationship between mathematics anxiety and other factors with different perpectives apart from previous studies. It can be said that investigating mathematics anxiety with various scales could be helpful to the area in order to understand the situation.

Although multiple data collection used by quantitative studies in recent years, it could be said that there is a need to use other research methodologies. Since this can enable researchers to discover the phenomenon in detail. As seen in Table 2, interview and open-ended form were the data collection tools in research designed with qualitative approach. In addititon to this, it was emerged that data collection tools in those designed with mixed-method approach were interviews and open-ended form with scales. There are very few studies done according to such methods like meta-analysis, literature review and document analysis. On the other hand, for instance, it is seen that none of these studies used focus group technique in data collection.

Table 2. Type of Methodology used in Research

Research Approach	Data Collection Tools	Examples
Quantitative *(survey)*	* Mathematics Anxiety Scale	* Bekdemir, 2007; Yenilmez, Girginer & Uzun, 2007; Baloğlu, 2008; Dede & Dursun, 2008; Işıksal, Curran, Koç & Askun, 2009; Aydın, Delice, Dilmaç & Ertekin, 2009; Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Baloğlu, 2010; Baloğlu ve Balgalmış, 2010; Pamuk & Karakaş, 2011; Akın, Kurbanoğlu & Takunyacı, 2011; Özdemir & Gür, 2011; Doruk & Kaplan, 2013; Taşdemir, 2013; Akçakın, Cebesoy & İnel, 2015; Taşdemir, 2015
	* Mathematics Anxiety Scale and Epistemological beliefs Scale	* Delice, Ertekin, Aydın & Dilmaç, 2009
	* Mathematics Anxiety Scale, Achievement Motivation Scale and Social Comparison Scale	* Kesici, & Erdoğan, 2010; Erdoğan, Kesici & Şahin, 2011
	* Mathematics Anxiety Scale and Perceived Social Support Scale-R	* Erden & Akgül, 2010
	* Mathematics Anxiety Scale and Mathematics Teaching Anxiety	* Peker & Ertekin, 2011
	* Mathematics Anxiety Scale, Mathematics Teaching Anxiety, Self-efficacy beliefs towards Mathematics Scale and Self-efficacy beliefs towards Mathematics Teaching Scale	* Ünlü, Ertekin & Dilmaç, 2017
	* Mathematics Anxiety Scale and Attitude Scale	* Peker & Şentürk, 2012; Karadeniz & Karadağ, 2014; Çatlıoğlu, Gürbüz & Birgin, 2014; Şimşek, Şahinkaya & Aytekin, 2017

Table 2. Type of Methodology used in Research (Cont.)

Research Approach	Data Collection Tools	Examples
Quantitative *(survey)*	∗ Mathematics Anxiety Scale, Attitude Scale and Self-efficacy Scale	∗ Akın & Kurbanoğlu, 2011; Doruk, Öztürk & Kaplan, 2016
	∗ Mathematics Anxiety Scale, Attitude Scale and Motivation Scale	∗ Kurbanoğlu & Takunyacı, 2012
	∗ Mathematics Anxiety Scale, Attitude Inventory and Students' grades	∗ Hacıömeroğlu, 2017
	∗ Mathematics Anxiety Scale and Positive and Negative Perfectionism Scale	∗ İlhan & Öner Sünkür, 2012
	∗ Mathematics Anxiety Scale and students' grades	∗ İlhan & Sünkür, 2013
	∗ Mathematics Anxiety Scale and Mathematical Beliefs Scale	∗ Hacıömeroğlu, 2013
	∗ Mathematics Anxiety Scale and Test Anxiety Scale	∗ Oksal, Durmaz & Akın, 2013
	∗ Mathematics Anxiety Scale, Test Anxiety Scale and students' grades	∗ Yılmaz & Bindak, 2016
	∗ Mathematics Anxiety Scale, Beliefs survey, Number and Operation Task and Geometric Shapes Sorting Task	∗ Aslan, Gürgah Oğul & Taş, 2013
	∗ Mathematics Anxiety Scale and Self-Regulation Skill Scale	∗ İşleyen, 2015
	∗ Mathematics Anxiety Scale, Basic Psychological Needs Scale and Academic Self-Regulation Scale	∗ Durmaz & Akkuş, 2016
	∗ Mathematics Anxiety Scale and Classroom Assesment Environment Scale	∗ İlhan, 2015
	∗ Mathematics Anxiety Scale and Meta-Cognition Questionnaire	∗ Kacar & Sarıçam, 2015; Gökbulut & Akdağ, 2016
	∗ Mathematics Anxiety Scale and Metacognitive Awareness Inventory	∗ Sarıçam & Ogurlu, 2015

Table 2. Type of Methodology used in Research (Cont.)

Research Approach	Data Collection Tools	Examples
Quantitative *(survey)*	∗ Mathematics Anxiety Scale and Self-efficacy Scale	∗ Yurt & Kurnaz, 2015
	∗ Mathematics Anxiety Scale, Mistaken Handling Learning Scale and Self- Efficacy Scale	∗ Aksu, Özkaya, Gedik & Konyalıoğlu, 2016; Adal & Yavuz, 2017
	∗ Mathematics Anxiety Scale and Learning Strategies Scale	∗ Arslan, Güler & Gürbüz, 2017
	∗ Questionnaire	∗ Aydın, 2011; Uysal & Selışık, 2016
	∗ Students' grades	∗ Kurt & Özel, 2013
Qualitative	∗ Interview	∗ Alkan, 2011; Alkan, 2013b
	∗ Open-ended form	∗ Alkan, 2013a
Mixed Method	∗ Mathematics Anxiety Scale, Achievement Test and Interviews	∗ Bekdemir, 2009; Bekdemir and Başıbüyük, 2011
	∗ Mathematics Anxiety and Mathematics Teaching Anxiety Scales and Open-ended form	∗ Zengin, 2017
	∗ Mathematics Anxiety Scale, metaphors and interviews	∗ Özdemir & Sezginsoy Şeker, 2017
Others	∗ Literature Review	∗ Alkan, 2010
	∗ Meta-analysis	∗ Şad, Demir & Özer, 2016
	∗ Document analysis	∗ Erten Tatlı, Atalan Ergin & Demir, 2016

3.3. Main Participants of Research

The main participant of the research related with *'mathematics anxiety'* is students as they are facing with this situation. Additionally, students' teachers are other significant elements of studies related with this topic. Therefore, it is seen that the majority of the research conducted with students. Only two of the studies are done with teachers. The list of participants and related research are given in Table 3.

Considering the participants, it can be said that most of the studies conducted with middle school students among those done with the group of students. This points that there could be an idea accepted in those studies that middle school students were more anxious than students at other levels. Beside this, the results show that pre-service teachers and undergraduates are the second big group of participants in the sampling. Studies on pre-service teachers' and undergraduates' anxieties in mathematics that were reviewed indicate that anxiety in these levels could be problematic for their education.

It is seen in Table 3, there are also research selected high school students as participants. Even though these are less than other groups of participants, the results show that there is a need to study with these students in order to find out the causes of mathematics anxiety. On the other hand, there is a few research conducted with primary school students (e.g. Alkan, 2011;

2013b; Hacıömeroğlu, 2017 and Peker & Şentürk, 2012) Moreover, there is only one research including kindergarten students in the sampling (e. g. Aslan et al., 2013). Apart from all these groups of participants, it is seen that there are three studies done based on the documents (e.g. Alkan, 2010; Erten Tatlı et al., 2016 and Şad et al., 2016)

The reviews about the participants of the research suggest that studying with middle-schools students, pre-service teachers and undergraduates could add extra information to the area. However, it is seen that there is a need to study with primary school students in that students face with mathematics firstly at this level. In addition to this, it could be good to understand *'mathematic anxiety'* from the lenses of different combinations of participants, for instance students and teachers or students and parents, etc.

Table 3. The main participants of research

Main Participants	**Examples**
Pre-service Teachers	Bekdemir, 2007; Işıksal, Curran, Koç & Askun, 2009; Delice, Ertekin, Aydın & Dilmaç, 2009; Peker & Ertekin, 2011; Akın, Kurbanoğlu & Takunyacı, 2011; Doruk & Kaplan, 2013; Hacıömeroğlu, 2013; Çatlıoğlu, Gürbüz & Birgin, 2014; Akçakın, Cebesoy & İnel, 2015; Kaçar, Sarıçam; 2015; Zengin, 2017
Undergraduates	Yenilmez, Girginer & Uzun, 2007; Baloğlu, 2008; Bekdemir, 2009; Aydın, Delice, Dilmaç & Ertekin, 2009; Baloğlu, 2010; Pamuk & Karakaş, 2011; Bekdemir & Başıbüyük, 2011; Akın & Kurbanoğlu, 2011; Taşdemir, 2013; Gökbulut & Akdağ, 2016; Ünlü, Ertekin & Dilmaç, 2017; Özdemir & Sezginsoy Şeker, 2017
High School Students	Erdoğan, Kesici & Şahin, 2011; Kurbanoğlu & Takunyacı, 2012; İlhan, 2015; Durmaz & Akkuş, 2016; Uysal & Selışık, 2016;
Middle School Students	Dede & Dursun, 2008; Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Erden & Akgül, 2010; Kesici & Erdoğan, 2010; Aydın, 2011; Özdemir & Gür, 2011; İlhan & Öner Sünkür, 2012; İlhan & Öner Sünkür, 2013; Oksal, Durmaz & Akın, 2013; Kurt & Özel, 2013; Karadeniz & Karadağ, 2014; İşleyen, 2015; Sarıçam & Oğurlu, 2015; Taşdemir, 2015; Yurt & Kurnaz, 2015; Aksu, Özkaya, Gedik & Konyalıoğlu, 2016; Doruk, Öztürk, Kaplan, 2016; Yılmaz & Bindak, 2016; Arslan, Güler & Gürbüz, 2017; Adal & Yavuz, 2017
Primary School Students	Alkan, 2011; Peker & Şentürk, 2012; Hacıömeroğlu, 2017
Primary and Middle School Students	Baloğlu & Balgalmış, 2010; Şimşek, Şahinkaya & Aytekin, 2017
Primary School Teachers	Alkan, 2013a
Primary School Students and their mothers	Alkan, 2013b
Kindergarten students and their teachers	Aslan, Gürgah Oğul & Taş, 2013
Documents	Alkan, 2010; Şad, Kış, Demir & Özer, 2016; Erten Tatlı, Atalan Ergin & Demir, 2016

3.4. Main outcomes of research

A general overview of the reviewed studies' outcomes is given in Table 4. The main outcomes of the studies are classified according to the participants selected in the reviewed studies as seen in the Table 4. While presenting the data, examples of the studies are provided for each result. In addition to this, these outcomes of reviewed studies about mathematics anxiety are categorised under three major themes as *'reasons of mathematics anxiety'*, *'adaptation of scales into Turkish'* and *'reduction in mathematics anxiety'*. In Figure 2 the distribution of these studies based on themes are presented. The numbers of studies presented under three themes in Figure 2 are correspondingly given in Table 4. In other words, each outcome is linked with these themes in Table 4.

As mentioned earlier, the most of the research under the review were designed in terms of quantitative research approach and mostly conducted with students at different levels. Additionally, studies used qualitative and mixed method approaches mostly collected the data based on students' views. Therefore, the results of these studies founded on students' perceptions or perspectives on mathematics anxiety and its relationship with such factors.

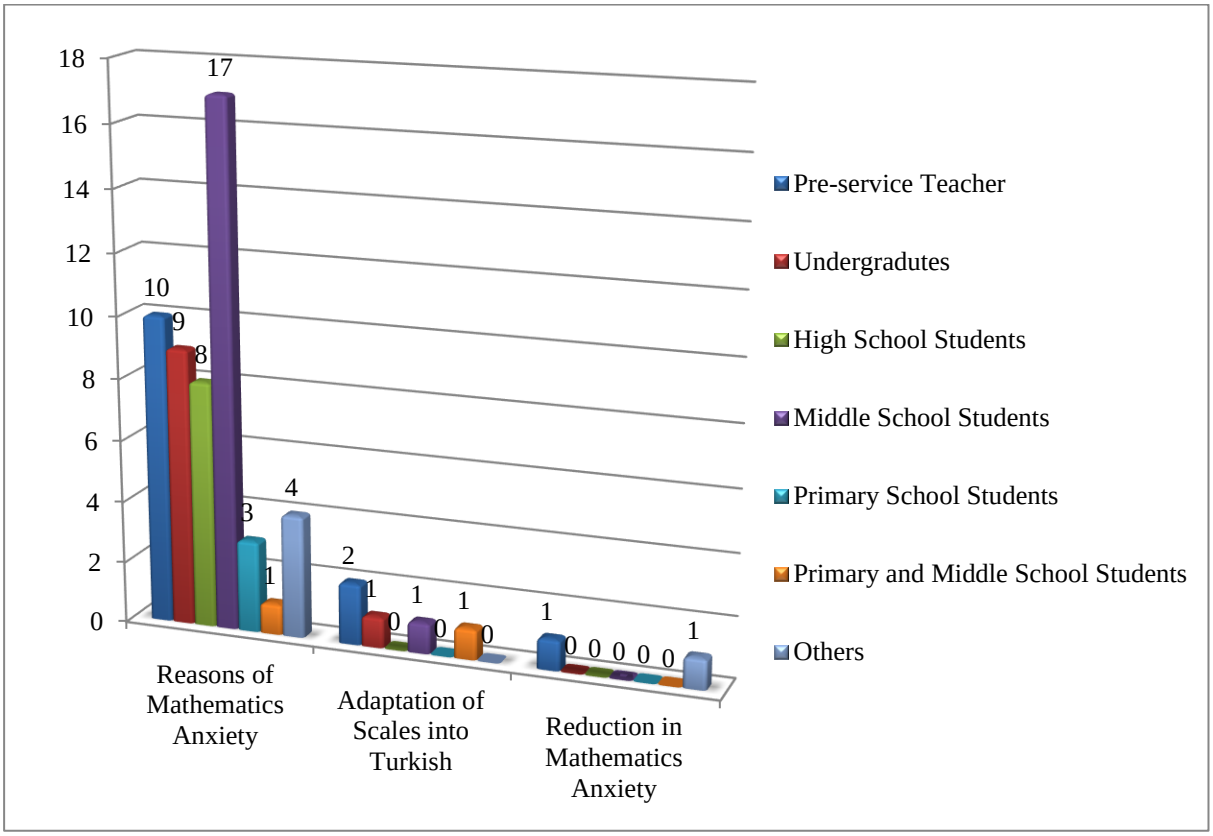


Figure 2. Themes emerged from reviewed research

It is seen in Figure 2 that almost all studies' results are related with the theme *'reasons of mathematics anxiety'*. The highest number of studies within this theme conducted with middle school students. The review results indicate that such factors as the level of achivement in mathematics, teachers' and parents' lack of support, students' fear, self-efficacy and beliefs and metacognitive awareness in mathematics can cause mathematics anxiety (see Table 4).

In some studies reviewed under the theme *'reasons of mathematics anxiety'*, it can be seen that results about the effect of gender on mathematics anxiety show contradictions. For instance, İlhan & Öner Sünkür (2013) stated that girls are more anxious in mathematics compare to boys. On the other hand, the results of studies conducted by Birgin, Baloğlu,

Çatlıoğlu & Gürbüz, (2010), Çatlıoğlu, Gürbüz & Birgin (2014) & Taşdemir (2015) emphasized that students' anxiety is not related with their being a girl or boy (see also Table 4).

There are few studies under the themes *'adaptation of scales into Turkish'* and *'reduction in mathematics anxiety'* as given in Figure 2. The adaptation of scales related with mathematics anxiety into Turkish done for pre-service teachers, undergraduates, middle school students and both for primary and middle school students. According to the results of reviewed studies on this topic, it can be said that R-MARS, BAI, MARS-SV, MAAS and MARS-E are valid and reliable for selected samples (see Table 4).

The results about the theme *'reduction in mathematics anxiety'* indicate that there are only two studies under this theme. One of these studies is done with pre-service teachers whereas the other one is done with primary school teachers. As seen in Table 4, the study conducted by Zengin (2017) suggested that the learning and teaching process designed according to GeoGebra software can reduce the level of pre-service teachers' anxiety in mathematics. In addition to this, in Alkan's (2013a) study it is found that supporting students' motivation and comprehension in mathematics, using examples and exercises and games could be helpful to reduce the level of primary school students' anxiety in mathematics.

Table 4. Main outcomes emerged from the reseach

Participants	Themes	Main Outcomes	Examples
Pre-service teacher	*Reasons of Mathematics Anxiety*	*- The causes of mathematics anxiety* o *Teachers' negative attitudes and practices;* o *Students' fear of making a mistake and asking a question to their teachers*	∗ Bekdemir, 2007 ;
	Reasons of mathematics anxiety	*- Negative relationship between mathematics anxiety and mathematical self-concept*	∗ Işıksal, Curran, Koç & Askun, 2009;
	Reasons of mathematics anxiety	*- The level of anxiety in mathematics changes according to gender, the level of class and the university*	∗ Aydın, Delice, Dilmaç & Ertekin, 2009;
	Reasons of mathematics anxiety	*- Positive relationship between epistemological beliefs and mathematic anxiety*	∗ Delice, Ertekin, Aydın & Dilmaç, 2009;
	Reasons of mathematics anxiety	*- Positive relationship between mathematics anxiety and mathematics teaching anxiety*	∗ Peker & Ertekin, 2011
	Adaptation of scales into Turkish	*- Adapted form of R-MARS is valid and reliable to be used with pre-service teachers*	∗ Akın, Kurbanoğlu & Takunyacı, 2011
	Reasons of mathematics anxiety	*- Mathematics anxiety is moderately related with self-efficacy and this relation is affected by gender.*	∗ Doruk & Kaplan, 2013

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
	Reasons of mathematics anxiety	- *The lower mathematics anxious one has the higher mathematical beliefs.*	∗ Hacıömeroğlu, 2013
	Reasons of mathematics anxiety	- *Gender has no effect on mathematics anxiety. Achievement and anxiety in mathematics has negative relationship.*	∗ Çatlıoğlu, Gürbüz & Birgin, 2014
	Adaptation of scales into Turkish	- *Adapted version of BAI is reliable and valid for measuring pre-service teachers' anxiety level.*	∗ Akçakın, Cebesoy & İnel, 2015
	Reasons of mathematics anxiety	- *Metacognitive awareness has relation with mathematics anxiety. The increase in negative metacognition level causes anxiety in mathematics.*	∗ Kaçar & Sarıçam, 2015
	Reasons of mathematics anxiety	- *There is a positive relationship between metacognitive awareness and mathematics anxiety.*	∗ Gökbulut & Akdağ, 2016
	Reduction in mathematics anxiety	- *The learning and teaching process designed with GeoGebra software enable pre-service teachers to overcome mathematics anxiety and mathematics teaching anxiety.*	∗ Zengin, 2017
Undergraduates	Reasons of mathematics anxiety	-*Relationship between mathematics anxiety and attitudes* ○ *Positive attitude increases anxiety decreases*	∗ Yenilmez, Girginer & Uzun, 2007
		○ *The increase in the level of anxiety in mathematics causes negative attitude towards mathematics so that lowers self-efficacy*	∗ Akın & Kurbanoğlu, 2011

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
Undergraduates	Reasons of mathematics anxiety	- *Positive correlations between the subjective and objective measures of mathematics anxiety level*	∗ Baloğlu, 2008
	Reasons of mathematics anxiety	-*The causes of mathematics anxiety* ○ *Teachers' negative attitudes and practices;* ○ *Students' fear of making a mistake and asking a question to their teachers*	∗ Bekdemir, 2009
	Adaptation of scales into Turkish	- *Adapted form of MARS-SV can be used for undergraduates*	∗ Baloğlu, 2010;
	Reasons of mathematics anxiety	- *Online students are more anxious than on-campus students*	∗ Pamuk & Karakaş, 2011
	Reasons of mathematics anxiety	- *Mathematics anxiety and achievement is related with geography achievement positively*	∗ Bekdemir & Başıbüyük, 2011
	Reasons of mathematics anxiety	- *The level of anxiety in mathematics does not change according to gender. Students from vocational school are more anxious than those from other.*	∗ Taşdemir, 2013
	Reasons of mathematics anxiety	- *There is a positive relationship between mathematics teaching anxiety and mathematics anxiety. Self-efficacy beliefs have negative relation both with mathematics anxiety and mathematics teaching anxiety.*	∗ Ünlü, Ertekin & Dilmaç, 2017
	Reasons of mathematics anxiety	- *Students' personality, their interaction with teachers and school facilities cause mathematics anxiety.*	∗ Özdemir & Sezginsoy Şeker, 2017

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
High School Students	Reasons of mathematics anxiety	- *The level of anxiety in Mathematics* ○ *There isn't a high relation between mathematics anxiety and variables like gender and grade*	* Dede and Dursun, 2008
	Reasons of mathematics anxiety	*- The level of students' achievement motivation and social comparison has an impact on mathematics anxiety*	* Erdoğan, Kesici & Şahin, 2011
	Reasons of mathematics anxiety	*- Students' anxiety in matematics can be changed according to the type of schools but students' gender does not affect anxiety*	* Kurbanoğlu & Takunyacı, 2012
	Reasons of mathematics anxiety	*- Performance oriented assessment environment has positive whereas learning oriented assessment environment has negative relation with mathematicsanxiety.*	* İlhan, 2015
	Reasons of mathematics anxiety	*- The basic psychological needs had an impact on mathematics anxiety.*	* Durmaz & Akkuş, 2016
	Reasons of mathematics anxiety	*- The level of anxiety in mathematics changes according to the type of school. High school students have moderate level anxiety in mathematics.*	* Uysal & Selışık, 2016
	Reasons of mathematics anxiety	*- Mathematics anxiety is related with achievement in mathematics and the socio-economic level of parents.*	* Erten Tatlı, Atalan Ergin & Demir, 2016
	Reasons of mathematics anxiety	*- There is a slight relationship between mathematics anxiety and organizing strategies and rehearsal strategies. There is a moderate relationship between mathematics anxiety and the other learning strategies (elaboration, comprehension monitoring and affective strategies).*	* Arslan, Güler & Gürbüz, 2017

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
Middle School Students	Reasons of mathematics anxiety	- *The level of students' achievement motivation and social comparison has an impact on mathematics anxiety*	∗ Kesici & Erdoğan, 2010
	Reasons of mathematics anxiety	- *There is not a relation between gender and mathematics anxiety but is a relation with grade level and mathematics anxiety. Perceived enjoyment of mathematics has an effect on mathematics anxiety*	∗ Birgin, Baloğlu, Çatlıoğlu & Gürbüz, 2010; Taşdemir, 2015
	Reasons of mathematics anxiety	- *Negative relation between mathematics anxiety and students' achievement. Teacher support is significant predictor of success and anxiety in mathematics.*	∗ Erden & Akgül, 2010
	Adaptation of scales into Turkish	- *MAAS is valid and reliable to be used with middle school students*	∗ Özdemir & Gür, 2011
	Reasons of mathematics anxiety	- *There is not a relation between gender and is a slight relation between grade and mathematics anxiety*	∗ Aydın, 2011
	Reasons of mathematics anxiety	- *Mathematics anxiety has an impact on learning, positive perfectionism has relation with achievement whereas negative perfectionism has relation with anxiety*	∗ İlhan & Öner Sünkür, 2012
	Reasons of mathematics anxiety	- *Girls are more anxious than boys. There is a difference in the level of anxiety according to the grades*	∗ İlhan & Öner Sünkür, 2013
	Reasons of mathematics anxiety	- *Test and mathematics anxiety is moderatetly telated with each other*	∗ Oksal, Durmaz & Akın, 2013
	Reasons of mathematics anxiety	- *Geometry Garden has an impact on students' learning so that their being anxious in mathematics*	∗ Kurt & Özel, 2013
	Reasons of mathematics anxiety	- *There is a negative relation between mathematics anxiety and attitudes towards mathematics.*	∗ Karadeniz & Karadağ, 2014

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
Middle School Students	Reasons of mathematics anxiety	*-Self-regulation and mathematics anxiety has negative relationship.*	∗ İşleyen, 2015
	Reasons of mathematics anxiety	*- There is a negative relationship between mathematics anxiety and mathematics anxiety of gifted student.*	∗ Sarıçam & Ogurlu, 2015
	Reasons of mathematics anxiety	*- The sources of self-efficacy have an impact on gifted students' anxiety level in mathematics.*	∗ Yurt & Kurnaz, 2015
	Reasons of mathematics anxiety	*- There is a significant relationship between self-efficacy, mistake-handling learning and mathematics anxiety.*	∗ Aksu, Özkaya, Gedik & Konyalıoğlu, 2016
	Reasons of mathematics anxiety	*- Middle schools students' level of mathematics anxiety is moderate. There is a negative relationship between anxiety and self-efficacy and attitudes in mathematics.*	∗ Doruk, Öztürk & Kaplan, 2016; Adal & Yavuz, 2017
	Reasons of mathematics anxiety	*- Achievement in mathematics has a negative impact on mathematics anxiety. The relation between test and mathematics anxiety is weak.*	∗ Yılmaz & Bindak, 2016
Primary School Students	Reasons of mathematics anxiety	*- Parents' lack of knowledge in mathematics, teachers support, students' self-efficacy and interaction with peers can cause anxiety in mathematics*	∗ Alkan, 2011
	Reasons of mathematics anxiety	*- Gender, satisfaction with the lesson and the teacher, and attitudes can be predictos of mathematics anxiety.*	∗ Peker & Şentürk, 2012
	Reasons of mathematics anxiety	*- There is a slightly negative relation between attitudes and anxiety in mathematics. The higher success the lower anxiety in mathematics*	∗ Hacıömeroğlu,2017

Table 4. Main outcomes emerged from the reseach (Cont.)

Participants	Themes	Main Outcomes	Examples
Primary and Middle School Students	Adaptation of scales into Turkish	*- Adapted version of MARS-E is reliable and valid to measure anxiety in mathematics*	∗ Baloğlu & Balgalmış, 2010
	Reasons of mathematics anxiety	*- Achievement and anxiety in mathematics has negative correlation. Fourth graders are more anxious than seventh graders. Gender has no impact on the level of attitudes and anxiety in mathematics.*	∗ Şimşek, Şahinkaya & Aytekin, 2017
Others (Primary School Teachers, Mothers, Documents)	Reasons of mathematics anxiety	*- Mathematics anxiety is the reason of students' personality, peers, parents and interaction with teachers*	∗ Alkan, 2010
	Reduction in mathematics anxiety	*- In order to reduce mathematics anxiety, primary school teachers support students' motivation and comprehension in mathematics. Use examples and exercises beside games to support students' effective learning*	∗ Alkan, 2013a ∗ Alkan, 2013b
	Reasons of mathematics anxiety	*- Mothers' ignoring anxiety, lack of knowledge, lack of support causes students' mathematics anxiety at primary school*	
	Reasons of mathematics anxiety	*- Pre-school teachers being anxious in mathematics has no effect on students' anxiety.*	∗ Aslan, Gürgah Oğul & Taş, 2013
	Reasons of mathematics anxiety	*- There is a negative relations between achievement and anxiety in mathematics*	∗ Şad, Kış, Demir & Özer, 2016,

4. CONCLUSION AND IMPLICATIONS

This systematic review research was focused on studies about *'mathematics anxiety'*. By this study, it is believed that crucial points in regard to *'mathematics anxiety'* are determined for future studies. Since a whole descriptive picture of mathematics anxiety in educational settings and its relation with such factors is provided with this review.

The reviewed studies indicated that nearly all studies done with students. Additionally, most of these students were selected from middle school levels. These findings, on the one hand suggest that students have significant role in defining anxiety and in determining its possible reasons in mathematics. Therefore, it can be said that the data gathered from this group provides a valuable insight to the area. On the other hand, it is seen that there is a need to do studies with

students from different school levels in order to delineate the anxiety in mathematics at various levels. Some of the reviewed studies done with middle school students stressed divergent results. For instance, in the study conducted by İlhan and Öner Sünkür (2013), it is claimed that girls are more anxious compare to boys whereas other studies done by Birgin, Baloğlu, Çatlıoğlu & Gürbüz (2010), Çatlıoğlu, Gürbüz & Birgin (2014) and Taşdemir (2015) emphasized that there is no significant relation between gender and mathematics anxiety. These results show that there is a need to conduct study with big samples at middle school levels and also at other levels to establish the relationship between these. Alternatively, a small sample study understanding the phenomenon deeply from different genders from all type of school levels could provide new insight into the area.

The result emerged from the reviewed studies emphasizes the role of teachers. Considering mathematics anxiety in educational settings, teachers' role in either creating or reducing anxiety in mathematics is also crucial. This means, further studies can focus on investigating mathematics anxiety according to teachers' views. In studies conducted by Alkan (2011, 2013), some points like, parents' and teachers' lack of support, interaction between teachers and students, friends and mothers are highlighted as factors of causing mathematics anxiety. These results propose that in Turkish context, further studies need to investigate mathematics anxiety from different perspectives like teachers, mothers or friends. Moreover, it could be good to conduct study with different combinations of participants such as teacher-student or student-teacher-parents.

Most of the reviewed studies aimed to measure the level of anxiety in mathematics and to find out causes of mathematics anxiety. In addition to this, in some studies while identifying anxiety, its relationship with genders, school types, students' capabilities and attitudes in mathematics are explored. However, other issues like the impacts of friends, teachers and parents or students' personalities on students' mathematics anxiety are need to be studied in detail. Another contribution of this review is that it helps to discover the lack of studies about such topics as reducing or overcoming mathematics anxiety and developing Turkish mathematics anxiety scales for various school levels.

Concerning all reviewed studies it can be said that most of them designed according to quantitative research methodology. The results emerged from these studies show that scales are commonly used data collection tools. It can be accepted that these quantitative studies provide productive information to the literature. Additionally, the scales used in these quantitative studies are useful to reach many participants at a specific period and also to maintain conclusive results. Nevertheless, it can be also seen that there is a lack of studies designed in terms of mixed method and qualitative research methodologies. For this reason, it can be said that mixed method and qualitative research studies could bring fruitful knowledge to the area. Moreover, such data collection tools as interviews, observations, focus groups are needed to be used to comprehend students' feelings about mathematics anxiety as well as to define and to exemplify anxiety in mathematics. On the other hand, it is seen that using metaphors (see e.g. Özdemir & Sezginsoy Şeker, 2017) to draw students' anxiety in mathematics could provide interesting insight into the area.

In conclusion, it is seen that most of the studies under the review designed according to the quantitative research methodology as well as obtained the data based on students' perceptions by scales. Students' anxiety in mathematics can be measured or clarified by their perceptions or perspectives, but there should be studies considering other groups of participants like teachers, parents or friends. According to the results of reviewed studies, it can be said that most of the studies conducted to find out nearly similar aims. This suggests that there should be studies focusing on special topics relating with *'mathematics anxiety'*. It might be concluded that future studies related to *'mathematics anxiety'* need to widen the scope of the topic, to use

multiple data collection tools and to gather data from various groups of participants in order to discover the whole picture of mathematics anxiety in educational settings in Turkey. This will also support triangulation of findings and enhance the trustworthiness of the results.

ORCID

Vesile ALKAN https://orcid.org/0000-0002-8630-3357

5. REFERENCES

Adal, A. A., & Yavuz, İ. (2017). Ortaokul öğrencilerinin matematik öz yeterlik algıları ile matematik kaygı düzeyleri arasındaki ilişki [The relationship between mathematics self efficacy and mathematics anxiety levels of middle school students]. *International Journal of Field Education, 3*(1), 20-41.

Akçakın, V., Cebesoy, Ü. B., & İnel, Y. (2015). İki boyutlu matematik kaygısı ölçeğinin Türkçe formunun geçerlik ve güvenirlik çalışması [Validity and reliability study of Turkish version of bidimensional mathematics anxiety scale]. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi [Gazi University Journal of Gazi Educational Faculty], 35*(2), 283-301.

Akın, A., & Kurbanoğlu, İ. N. (2011). The relationships between math anxiety, math attitudes, and self-efficacy: A structural equation model. *Studia Psychologica, 53*(3), 263-273.

Akın, A., Kurbanoğlu, İ. N., & Takunyacı, M. (2011). Revize edilmiş matematik kaygısı değerlendirme ölçeği: doğrulayıcı faktör analizi çalışması [Revised mathematics anxiety rating scale: A confirmatory factor analysis]. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED) [Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education], 5*(1), 163-180.

Aksu, Z., Özkaya, M., Gedik, S. D., & Konyalıoğlu, A. C. (2016). Mathematics self-efficacy and mistake-handling learning as predictors of mathematics anxiety. *Journal of Education and Training Studies, 4*(8), 65-71.

Alkan, V. (2009). *The Relationship between teaching strategies and styles and pupils' anxiety in mathematics at primary schools in Turkey.* Unpublished PhD Thesis. The University of Nottingham.

Alkan, V. (2010). Matematikten nefret ediyorum! [I hate Mathematics!]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi [Pamukkale University Journal of Education], 28* (II), 189-199.

Alkan, V. (2011). Etkili matematik öğretiminin gerçekleştirilmesindeki engellerden biri: kaygı ve nedenleri [One of the barriers to providing effective mathematics teaching: anxiety and its causes]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi [**Pamukkale University Journal of Education**], 29*(I), 89-107.

Alkan, V. (2013a). Reducing mathematics anxiety: The ways implemented by teachers at primary schools in Turkey. *International J. Soc. Sci. & Education, 3* (3), 795-807.

Alkan, V. (2013b). Mothers and their relation with pupils' mathematics anxiety. *International Global Research Analysis*, 2(4): 83-86.

Ashcraft, M.H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, 1, 3–34.

Ashcraft, M.H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. http://dx.doi.org/10.1111/1467-8721.00196

Aslan, D., Gürgah Oğul, İ., & Taş, I. (2013). The Impacts of preschool teachers' mathematics anxiety and beliefs on children's mathematics achievement. *International Journal of Humanities and Social Science Invention, 2*(7), 45-49.

Arslan, Ç., Güler, H. K., & Gürbüz, M. Ç. (2017). Ortaokul öğrencilerinin matematik kaygı düzeyleri ile öğrenme stratejileri arasındaki ilişkinin incelenmesi [Investigation of the relation between middle school students' mathematics anxiety and learning strategies]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi [Mehmet Akif Ersoy University Journal of Education Faculty]*, *42*, 123-142.

Aydın, E., Delice, A., Dilmaç, B., & Ertekin, E. (2009). İlköğretim matematik öğretmen adayların matematik kaygı düzeylerine cinsiyet, sınıf ve kurum değişkenlerinin etkileri [The influence of gender, grade and institution on primary school mathematics student teachers' anxiety levels]. *İlköğretim Online [Elementary Education Online],* 8 (1), 231-242.

Aydın, B. (2011). İlköğretim ikinci kademe düzeyinde matematik kaygısının cinsiyete göre farklılıkları üzerine bir çalışma [A study on secondary school students' mathematıcs anxıety in terms of gender factor]. *Kastamonu Eğitim Dergisi [Kastamonu Education Journal]*, *19*(3), 1029-1036.

Baloğlu, M. (2008). Üniversite öğrencilerinin öznel ve nesnel matematik kaygısı düzeylerinin ölçümlenmesi [Subjective and objective assessment of mathematics anxiety levels among college students]. *Boğaziçi Üniversitesi Eğitim Dergisi [Boğaziçi University Journal of Education]*, *25*(1), 1-7.

Baloğlu, M. (2010). An investigation of the validity and reliability of the adapted mathematics anxiety rating scale-short version (MARS-SV) among Turkish students. *Eur J Psychol Educ, 25*, 507–518 DOI 10.1007/s10212-010-0029-2

Baloğlu, M., & Balgalmış, E. (2010). Matematik kaygısını derecelendirme ölçeği ilköğretim formu'nun Türkçe'ye uyarlanması, dil geçerliği ve psikometrik incelemesi [The adaptation of the mathematics anxiety rating scale-elementary form into Turkish, language validity, and preliminary psychometric investigation]. *Kuram ve Uygulamada Eğitim Bilimleri [Educational Sciences: Theory & Practice]*, *10*(1), 77-110.

Bekdemir, M. (2007). İlköğretim matematik öğretmen adaylarındaki matematik kaygısının nedenleri ve azaltılması için öneriler (Erzincan Eğitim Fakültesi Örneği) [The causes of mathematics anxiety in elementary preservice teachers and proposals for decreasing mathematics anxiety (The example of Faculty of Erzincan Education]. *Erzincan Eğitim Fakültesi Dergisi [Erzincan University Journal of Education Faculty]*, *9*(2), 131-144.

Bekdemir, M. (2009). Meslek yüksekokulu öğrencilerinin matematik kaygı düzeylerinin ve başarılarının değerlendirilmesi [Evaluation of vocational school student's mathematics anxiety levels and achievement]. *EÜFBED-Fen Bilimleri Enstitüsü Dergisi [Erzincan University Journal Science and Technology]*, *2*(2), 169-189.

Bekdemir, M., & Başıbüyük, A. (2011). Sosyal bilgiler ve sınıf öğretmenliği programı öğrencilerinin matematik başarı ve kaygı düzeylerinin coğrafya başarısını yordaması [The prediction of the levels of mathematics achievement and anxiety of the social sciences and primary education students to their geography achievement]. *GÜ, Gazi Eğitim Fakültesi Dergisi [Gazi University Journal of Gazi Educational Faculty]*, *31*(2), 459-477.

Birgin, O., Baloğlu, M., Çatlıoğlu, H., & Gürbüz, R. (2010). An investigation of mathematics anxiety among sixth through eighth grade students in Turkey. *Learning and Individual Differences, 20*, 654–658.

Cemen, P. B. (1987). The nature of mathematics anxiety. (Report No. SE 048 689). Stillwater: Oklahoma State University. (ERIC Document Reproduction Service No. ED 287 729).

Çatlıoğlu, H., Gürbüz, R., & Birgin, O. (2014). Do Pre-service Elementary School Teachers Still Have Mathematics Anxiety? Some Factors and Correlates. *Bolema, Rio Claro (SP), 28* (48), 110-127.

Dede, Y., & Dursun, Ş. (2008). İlköğretim II. kademe öğrencilerinin matematik kaygı düzeylerinin incelenmesi [An investigation of primary school students' mathematics anxiety levels]. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi [Journal of Uludağ University Faculty of Education]*, *XXI* (2), 295-312.

Delice, A., Ertekin, E., Aydın, E., & Dilmaç, B. (2009). Öğretmen adaylarının matematik kaygısı ile bilgibilimsel inançları arasındaki ilişkinin incelenmesi [An investigation of the relationship between epistemological beliefs and mathematics anxiety of student teachers]. *Uluslararası İnsan Bilimleri Dergisi*, *6*(1), 361-375.

Doruk, M., & Kaplan, A. (2013). Sınıf ve ilköğretim matematik öğretmeni adaylarının matematik kaygılarının incelenmesi [Examining mathematics anxiety of prospective primary school and prospective primary mathematics teachers]. *Kastamonu Eğitim Dergisi [Kastamonu Education Journal]*, *21*(4), (Özel Sayı-Special Issue) 1505-1522.

Doruk, M., Öztürk, M., & Kaplan, A. (2016). Ortaokul öğrencilerinin matematiğe yönelik öz-yeterlik algılarının belirlenmesi: kaygı ve tutum faktörleri [Investigation of the self-efficacy perceptions of middle school students towards mathematics: anxiety and attitude factors]. *Adıyaman Üniversitesi Eğitim Bilimleri Dergisi [Adıyaman University Journal of Educational Sciences]*, *6*(2), 283-302.

Dreger, R., & Aiken, L. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology, 48*(6), 344-351.

Durmaz, M., & Akkuş, R. (2016). Öz belirleme kuramı perspektifinden matematik kaygısı, motivasyon ve temel psikolojik ihtiyaçlar [Mathematics anxiety, motivation and the basic psychological needs from the perspective of self-determination theory]. *Eğitim ve Bilim [Education and Science]*, *41*(183), 111-127.

Erden, M., & Akgül, S. (2010). İlköğretim öğrencilerinin matematik kaygısının ve öğretmen sosyal desteğinin matematik başarılarını yordama gücü [Predictive Power of math anxiety and perceived social support from teacher for primary students' mathematics achievement]. *Eğitimde Kuram ve Uygulama [Journal of Theory and Practice in Education]*, *6*(1), 3-16

Erdoğan, A., Kesici, Ş., & Şahin, İ. (2011). Lise Öğrencilerinin Başarı Güdülerinin ve Sosyal Kıyaslama Düzeylerinin Matematik Kaygılarını Yordaması. [Prediction of high school students' mathematics anxiety by their achievement motivation and social comparison]. *İlköğretim Online [Elementary Education Online]*, *10*(2), 646-652.

Erten Tatlı, C., Atalan Ergin, D., & Demir, E. (2016). PISA 2012 Türkiye verilerine göre öğrencilerin matematik kaygısının sınıflandırıcıları [Classifiers of students' mathematical anxiety according to PISA 2012 turkey data]. *İlköğretim Online [Elementary Education Online]*, *15*(2), 696-70. Doi: http://dx.doi.org/10.17051/ io.2016.20173

Gough, M. F. (1954). Mathemaphobia: Causes and treatments. *Clearing House, 28,* 290-294. Retrieved from http://www.jstor.org/stable/30176259

Gökbulut, Y., & Akdağ, M. (2016). Sınıf öğretmeni adaylarının üstbilişsel farkındalık ve matematik kaygı düzeyleri arasındaki ilişkisi [The relationship between preservice elementary teachers' math anxiety levels and their metacognitive awareness]. *Turkish Studies, 11*(9), 461-474.

Gresham, G. (2010). A study exploring exceptional education pre-service teachers' mathematics anxiety. *Issues in the Undergraduate Mathematics Preparation of School Teachers, 4,* 1-14. Retrieved from http://www.k-12prep.math.ttu.edu/

Hacıömeroğlu, G. (2013). Mathematics anxiety and mathematical beliefs: what ıs the relationship in elementary pre-service teachers? *IUMPST: The Journal*. *5* (Teacher Attributes), 1-9.

Hacıömeroğlu, G. (2017). Reciprocal Relationships between Mathematics Anxiety and Attitude towards Mathematics in Elementary Students. *Acta Didactica Napocensia*, *10*(3), 59-68.

Işıksal, M. Curran, J. M., Koç, Y., & Askun, C. S. (2009) Mathematics anxiety and mathematical self-concept: considerations in preparing elementary-school teachers. *Social Behavior and Personality*, *37*(5), 631-644.

İlhan, M., & Öner Sünkür, (2012). Matematik kaygısı ile olumlu ve olumsuz mükemmeliyetçiliğin matematik başarısını yordama gücü [Predictive power of math anxiety, positive and negative perfectionism for 8th grade students' mathematıcs Achievement]. *Mersin Üniversitesi Eğitim Fakültesi Dergisi [Mersin University Journal of the Faculty of Education]*, *8*(1), 178-188.

İlhan, M., & Öner Sünkür, M. (2013). Matematik kaygısının matematik başarısını yordama gücünün cinsiyet ve sınıf değişkenleri açısından incelenmesi [Investigation of predictive power of mathematics anxiety on mathematics achievement in terms of gender and class variables]. *Gaziantep University Journal of Social Sciences [Gaziantep University Journal of Social Sciences]*, *12*(3), 427- 441.

İlhan, M. (2015). The Predictive power of students' perceptions of classroom assessment environment for their mathematics anxiety [Öğrencilerin sınıf değerlendirme atmosferine yönelik algılarının matematik kaygılarını yordama gücü]. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi [Ondokuz Mayis University Journal of Faculty of Education], 34*(2), 1-21.

İşleyen, T. (2015). The Relationship between Secondary School Students' Mathematics Anxiety and Self-regulation. *Educational Research and Reviews*, *10*(5), 684-690.

Kaçar, M., & Sarıçam, H. (2015). Sınıf öğretmen adaylarının üstbiliş farkındalıkları ile matematik kaygı düzeyleri üzerine bir çalışma [The examination of the relationship between metacognitive awareness and math anxiety levels in pre-service primary school teachers]. *Trakya Üniversitesi Eğitim Fakültesi Dergisi [Trakya University Journal of Education]*, *5*(2), 137-152.

Karadeniz, İ., & Karadağ, E. (2014). Kırsal bölgelerdeki ortaokul öğrencilerinin matematik kaygı ve tutumları: korelasyonel bir araştırma [Mathematics anxiety and attitudes of secondary school students in rural area: A correlational research]. *Türk Bilgisayar ve Matematik Eğitimi Dergisi [Turkish Journal of Computer and Mathematics Education]*, *5*(3), 259-273.

Kesici, Ş., & Erdoğan, E. (2010). Mathematics anxiety according to middle school students' achievement motivation and social comparison. *Education*, *131*(1), 54-63.

Kurbanoğlu, N. İ., & Takunyacı, M. (2012). Lise öğrencilerinin matematik dersine yönelik kaygı, tutum ve öz-yeterlik inançlarının cinsiyet, okul türü ve sınıf düzeyi açısından incelenmesi [An investigation of the attitudes, anxieties and self-efficacy beliefs towards mathematics lessons high school students' in terms of gender, types of school, and students' grades]. *Uluslararası İnsan Bilimleri Dergisi [International Journal of Human Sciences]*, *9*(1), 110-130.

Kurt, A., & Özel, M. E. (2013). İlköğretimde matematik kaygısına karşı "gerçekçi matematik eğitimi" yaklaşımı ve "geometri bahçesi"nin rolü [A cure for mathematics anxiety: "geometry garden" in the context of "realistic mathematics education" in elementary schools]. *Çağ Üniversitesi Sosyal Bilimler Dergisi [Cag University Journal of Social Sciences]*, *10*(1). 144-151.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic Reviews and Meta-Analysis*. UK: Oxford University Press.

Maloney, E., & Block, S. (2012). Math anxiety: who has it, why it develops, and how to guard against it. *Trends in Cognitive Science, 16*(8), 404-406.

Millar, J. (2004). 'Systematic Reviews for Policy Analysis', In S. Becker and A. Byrman (eds.), *Understanding Research for Social Policy and Practice: Themes, Methods and Approaches*, Bristol: Policy Press

Oksal, A., Durmaz, B., & Akın, A. (2013). SBS'ye hazırlanan öğrencilerin sınav ve matematik kaygılarının bazı değişkenler açısından incelenmesi [An investigation into exam and maths anxiety of students preparing for SBS]. *Cumhuriyet International Journal of Education, 2*(4), 47-62.

Özdemir, E., & Gür, H. (2011). Matematik kaygısı-endişesi ölçeğinin (MKEÖ) geçerlik ve güvenirlik çalışması [Validity and reliability study of mathematics anxiety apprehension survey (MASS)]. *Eğitim ve Bilim [Education and Science]*, *36*(161), 39-50.

Özdemir, E., & Sezginsoy Şeker, B. (2017). Prospective primary teachers' mathematics anxiety-apprehension and its causes. *International Education Studies*, *10*(11), 1-22.

Pamuk, M., & Karakaş, S. (2011). Sosyal bilimler öğrencilerinde matematik kaygısı: uzaktan eğitim ve kampüs öğrencileri üzerine bir çalışma. *Ekonometri ve İstatistik*, 14, 19–37.

Peker, M. & Ertekin, E. (2011). The relationship between mathematics teaching anxiety and mathematics anxiety. *The New Educational Review*, 23(1), 213-226.

Peker, M. & Şentürk, B. (2012). İlköğretim 5.sınıf öğrencilerinin matematik kaygılarının bazı değişkenler açısından incelenmesi [An investigation of 5th grade studens' math anxiety in terms of some variables]. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi [Dumlupınar University Journal of Social Sciences]*, 34, 21-32.

Richardson, F.C., & Suinn, R.M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology, 19,* 551-554. http://dx.doi.org/10.1037/h0033456

Sarıçam, H., & Oğurlu, Ü. (2015). Metacognitive awareness and math anxiety in gifted students. *Cypriot Journal of Educational Sciences*, *10*(4), 338-348.

Şimşek, H., Şahinkaya, N., & Aytekin, C. (2017). İlköğretim öğrencilerinin matematik kaygılarının ve matematik dersine yönelik tutumlarının çeşitli değişkenler açısından incelenmesi [Investigation of the anxieties and attitudes of elementary school students towards mathematics with various variables]. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED) [Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education]*, *11*(2), 82-108.

Şad, S. N., Kış, A., Demir, M., & Özer, N. (2016). Matematik başarısı ile matematik kaygısı arasındaki ilişki üzerine bir meta-analiz çalışması [Meta-analysis of the relationship between mathematics anxiety and mathematics achievement]. *Pegem Eğitim ve Öğretim Dergisi [Pegem Journal of Education and Instruction]*, *6*(3), 371-392, http://dx.doi.org/10.14527/pegegog.2016.019

Taşdemir, C. (2013). Meslek yüksekokulu öğrencilerinin matematik kaygı düzeylerinin bazı değişkenler açısından incelenmesi [Evaluation of vocational school student's mathematics anxiety levels in term of some variables]. *BEÜ Fen Bilimleri Dergisi [BEU Journal of Science]*, *2* (2), 154-162.

Taşdemir, C. (2015). Ortaokul ögrencilerinin matematik kaygı düzeylerinin incelenmesi [To investigate the mathematic anxiety levels of secondary students]. *Batman Üniversitesi Yaşam Bilimleri Dergisi [Batman University Journal of Life Sciences]*, *5*(1), 1-12.

Torgerson, C. (2003). *Systematic Reviews*. London: Continuum International Publishing Group.

Uysal, F., & Selışık, A. (2016). Lise öğrencilerinin matematik kaygı düzeylerinin çeşitli değişkenlere göre incelenmesi [An investigation about high school students' mathematics anxiety level according to some variables]. *AKU Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science]*, 9(1), 146-164.

Ünlü, M., Ertekin, E., & Dilmac, B. (2017). Predicting relationships between mathematics anxiety, mathematics teaching anxiety, self-efficacy beliefs towards mathematics and mathematics teaching. *International Journal of Research in Education and Science (IJRES), 3*(2), 636-645. DOI: 10.21890/ijres.328096

Wu, S.S., Willcutt, E.G., Escovar, E., & Menon, V. (2014). Mathematics achievement and anxiety and their relation to internalizing and externalizing behaviors. *Journal of Learning Disabilities, 47*(6), 503-514.

Yenilmez, K., Girginer, N., & Uzun, Ö. (2007). Mathematics anxiety and attitude level of students of the faculty of economics and business administrator; The Turkey Model. *International Mathematical Forum*, 2(41), 1997-2021.

Yılmaz, H. R., & Bindak, R. (2016). Ortaokul öğrencilerinde matematik başarısının matematik kaygısı, sınav kaygısı ve bazı demografik değişkenlerle ilişkisinin incelenmesi [Investigation of math success with math anxiety, test anxiety and some demographic variables in secondary school students]. *MSKU Eğitim Fakültesi Dergisi [MSKU Journal of Education], 3*(2), 30-42.

Yurt, E., & Kurnaz, A. (2015). Özel yetenekli öğrencilerin matematik öz-yeterlik kaynaklarının matematik kaygıları üzerindeki etkilerinin incelenmesi [An investigation of the effects of the mathematics sources of self-efficacy on talented students'mathematics anxiety]. *Pegem Eğitim ve Öğretim Dergisi [Pegem Journal of Education and Instruction], 5*(4), 347-360, http://dx.doi.org/10.14527/pegegog.2015.019

Zengin, Y. (2017). GeoGebra yazılımının matematik kaygısı ve matematik öğretme kaygısına etkisinin incelenmesi [The investigation of the effect of geogebra software on mathematics anxiety and mathematics teaching anxiety]. *YYÜ Eğitim Fakültesi Dergisi [YYU Journal of Education Faculty], 14*(1), 908-939. http://dx.doi.org/10.23891/efdyyu.2017.34