EURASIAN JOURNAL OF EDUCATIONAL RESEARCH

CONTENTS
İçindekiler

REVIEWERS of the 79th ISSUE
79. Sayı Hakemleri

Eurasian Journal of Educational Research
www.ejer.com.tr

Analysis of Gender Roles in Primary School (1st to 4th Grade) Turkish Textbooks*

Derya ARSLAN OZER[1], Zeynep KARATAS[2], Ozge Ruken ERGUN[3]

A R T I C L E I N F O	A B S T R A C T
	Purpose: Textbooks can directly or indirectly affect gender stereotypes. In this study, text characters and pictures of textbooks were examined to find out how the messages were given about gender. This study examined text characters and the pictures in the primary school Turkish textbooks respectively in terms of Bem's Sex-Role Inventory and gender stereotypes. **Research Methods:** This paper explored the roles of text characters and pictures in the primary school Turkish textbooks through document analysis. The material for this study consisted of twelve primary school Turkish textbooks from the 1st to the 4th grades. Textbooks were selected out of the ones used in Burdur in the 2015-2016 academic years. A "Text Character Evaluation Form" arranged in line with Bem's Sex-role Inventory and "Picture Evaluation Form" was used in the study.

Results: The texts included male characters with feminine and masculine roles, and female characters with feminine and masculine roles. The first, second, third and fourth grade textbooks had more male pictures than female pictures. All colors were used for both males and females in the textbooks examined in this study. Dress, skirt, shirt and t-shirt were mostly worn by the females in the textbooks. Male clothes consisted of trousers, shirt and t-shirt.

Implications for Research and Practice: It is observed that text characters have androgynous roles, and it can be argued that using more texts with characters possessing such qualities may have a positive effect on children's opinions about gender.

---

* This study was partly presented at the 3rd International Eurasian Educational Research Congress in Muğla, 31 May – 03 June, 2016

[1] Corresponding Author, Burdur Mehmet Akif Ersoy University, TURKEY, e-mail: deryaarslan@mehmetakif.edu.tr, ORCID: https://orcid.org/0000-0002-4555-5435

[2] Burdur Mehmet Akif Ersoy University, TURKEY, e-mail: zeynepkaratas@mehmetakif.edu,tr, ORCID: 0000-0002-4532-6827

[3] Burdur MAKU, TURKEY E-mail: ozgeruken@gmail.com, ORCID: https://orcid.org/0000-0002-1514-8599

Introduction

It is necessary to determine how the practices within the education system affect gender perception of children so that the child can be comfortable with both his/her own sex and the opposite sex. Textbooks are carrying the gender's model for girls and boys. This study aims to reveal how the gender roles are described both in pictures and texts of textbooks and to come up with suggestions in this regard.

Textbooks have become a research topic in education, sociology and many other scientific fields as they are one of the oldest lesson materials, easily accessible and cheaper than most materials (Esen, 2003; Yanpar, 2006), and serve as a significant tool for preventing social conflicts (Pingel, 2010). The information and messages in textbooks can have a positive or negative effect on the cognitive, affective and social development of children through implicit or explicit images and writings (Sever, 2003). The characters and situations in books introduce children to what the world may look like through others' eyes, and offer a chance to construct their own views of self and the world (Mendoza, & Reese, 2001, 1). A textbook is and must be a medium for obtaining correct information and protecting the right to obtain correct information. As an objective communication medium, a textbook must be structured as to protect both the right to receive education and the right to access correct information (Cotuksoken, 2003) as well to be in line with the student's level (Ozbay, 2006).

Some information and messages given in textbooks are related to gender either implicitly or explicitly. It must be ensured through all kinds of media, types of literature (Gol, 2011) and textbooks that children understand both males and females, who have many differences in biological and/or sociological aspects according to the gender role approaches, but are first of all humans and valuable beings in all circumstances. There have been many studies in this field which examine textbooks in terms of *gender* (Chick & State Altoona, 2006; Elgar, 2004; Gharbavi & Mousavi, 2012; Gumusoglu, 2008; Guneş, 2008; Kırbasoglu Kılıç & Eyup, 2011; Sano, Lida, & Hardy, 2001; Worland, 2008; Yaylı & Kitis Çınar, 2014; Yıldız, 2013)*, female figure* (Ozkan, 2013), *female characters* (Vannicopulou, 2004), *sexism* (Asan Tezer, 2010), the *gender concept* (Cubukcu & Sivaslıgil, 2007) and *inclusion of female authors* (Kuşçu Küçükler, 2014). In addition, female characters in children's books (Cınar, 2015), genders of parents (Anderson, Hamilton, 2005), gender roles in children's literature (Poarch, Monk-Turner, 2001) are fields of interest as well.

Biology-based differences are accepted as "sex", while social differences between males and females, stemming from socio-cultural reasons, which are created by males and change across time, culture and even family, are regarded as "gender" (Bhasin, 2014; Dokmen, 2015; Giddens, 2013). Images related to both genders exist in everything ranging from dietary habits to dressing style, from toys to books, cartoons, computer games as well as in textbooks first encountered by students at school (Yaylı & Kitis Cınar, 2014). Some behaviors and traits expected by society from males and females as a group are called gender stereotypes (Dokmen, 2015). According to Cüceloglu (1993), the stereotypical perceptions of people about behavior and trait differences between the genders may not be related to reality at all, but people believe

stereotypes as if they were true. Some studies have been carried out with the aim of examining gender stereotypes in children's books (Anderson & Hamilton, 2005; Catalcalı Soyer, 2009; Lee & Collins, 2009; Paterson & Lach, 1990;). Catalcalı Soyer (2009) states that children meet stereotypes of the adult world through children's books and become candidates for the adult world via these stereotypes.

Studies dealing with sex, gender etc. in textbooks examine pictures (Esen & Baglı, 2002; Özkan, 2013; Yaylı & Kitis Cınar, 2014) and pictures and texts (Asan Tezer, 2010; Kılıç Kırbasoglu & Eyup, 2011). The number of male-female pictures (Özkan, 2013); the people, actions, places and objects with which adult figures are depicted (Esen & Baglı, 2002); roles of males and females in texts and images, gender roles, roles within and outside the family, professional roles, housework roles and personality roles (Kırbasoglu Kılıc & Eyup, 2011); main and side characters of texts and pictures; individuals, actions and places where people come together are analyzed according to the sexes. One of the frequently used tools for determining sex roles in quantitative research is the Sex-Role Inventory developed by Bem.

This study uses the Sex Role Inventory in order to evaluate the sex roles of text characters. Bem's (1974) inventory is based on the gender schema theory developed by herself. Children realize their own sex. Then, as they learn the content of the gender schema, they learn which qualities are related to their own sex. Bem developed the Sex-Role Inventory with the aim of measuring the perception of sex roles. This is a frequently used tool in psychology and other fields because it measures masculine and feminine sex roles separately, it can measure androgyny and has psychometric properties (Holt & Ellis, 1998). The main point of Bem's notions is that masculinity and femininity are two separate dimensions. Sex role orientations of individuals are determined by considering the level at which they possess the qualities in these two separate dimensions. Individuals with high femininity and low masculinity have feminine sex role orientation while individuals with high masculinity and low femininity have masculine sex role orientation (Demirtas, 2000). According to Bem (1974), mixed or androgynous self-perception may enable the individual to break free from traditional masculine and feminine sex role orientations and to perform both masculine and feminine behaviors freely. This study examines the roles of text characters in terms of feminine, masculine and social acceptability adjectives in Bem's Sex-Role Inventory, and pictures in terms of gender stereotypes in primary school Turkish books. With this aim, the main research questions of this study were determined as follows:

1. How are the gender roles of text characters in primary school Turkish Textbooks (1st to 4th grades) depicted in terms of feminine, masculine and social acceptability adjectives in Bem's Sex-Role Inventory?

2. How are the number of males and females, the colors used for males and females, the clothes of males and females, the environment where males and females are depicted, the division of labor, professions and physical attributes expressed in pictures according to gender stereotypes?

Method

Research Design

Qualitative research model was used in this research. This paper examines respectively the text characters and text pictures in terms of Bem's Sex-Role Inventory and gender stereotypes in the primary school Turkish textbooks from the 1st to the 4th grades through document review. Document review is a method of qualitative research (Yıldırım & Şimsek, 2005).

Research Sample

The materials in this study consisted of primary school Turkish textbooks from the 1st to the 4th grades. Textbooks were selected out of the ones used in Burdur in the 2015-2016 academic years. The references section provides information about the books reviewed.

Primary school Turkish student books consisted of three books and the first part of each book was the textbook, and the second part was the workbook. The first parts of the first and second books consisted of a section for teaching reading and writing only in the first grade. Four sets of Turkish textbooks used from the 1st to the 4th grades of primary school, which were equal to twelve books, (4x3), were reviewed during the study. A total of twelve textbooks for four grades, twenty-four themes included in those twelve textbooks, and one hundred and eighteen texts included in those themes were reviewed. Themes about Atatürk and the texts within that theme were not included in the study. In addition, as the listening texts were not in the student books, they were also excluded from the research.

Research Instruments and Procedures

A "Text Character Evaluation Form" arranged in line with Bem's Sex-role Inventory and "Picture Evaluation Form" were used in the study. Firstly, the research about the textbooks were reviewed. Some studies include only picture reviews (Esen & Baglı, 2002; Ozkan, 2013) while others examine both pictures and texts together (Asan Tezer, 2010; Kırbaşoglu Kılıc & Eyup, 2011). This study also reviews text characters and pictures together.

Text character evaluation form. Bem developed a scale for testing sex roles. It consisted of three sub-scales of masculinity (20 adjectives), femininity (20 adjectives) and social acceptability (20 adjectives) with a total of sixty items. The short version of Bem's form (20 feminine and 20 masculine) was adapted to Turkish by Kavuncu (1987), and the validity and reliability check was performed by Dokmen (1999). The remaining twenty adjectives which were related to social acceptability were translated into Turkish by the second researcher, and checked by an expert from the English Teaching Department. The final text evaluation form consisted of the theme title, text title, character name, roles, feminine, masculine and social acceptability adjectives.

Picture evaluation form. Firstly, the literature was examined (Esen & Baglı, 2002; Kırbasoglu Kılıc & Eyup, 2011; Ozkan, 2013), and an item pool was prepared based on

the literature. The form took its final shape according to expert opinions and a pilot study. The picture evaluation form consisted of the number of males and females, colors used for males and females, clothes, division of labor, professions and physical appearances of the characters. The third Turkish book of the 3rd grade was examined separately by three researchers in terms of both text characters' roles and pictures. The inter-rater reliability score (Miles, Huberman, 1994) of the picture evaluation form was determined to be .81.

Data Analysis

The text character evaluation form and picture evaluation form were used to evaluate the text characters and pictures, respectively. The data were evaluated through descriptive analysis, which is a qualitative research analysis method. The pictures were examined by the first and the second researcher. Frequency was used for evaluation.

The process of text character analysis is given below:

1. The first and second researchers read the texts together and determined the personality roles of the text characters.

2. The determined roles were compared with sixty adjectives consisting of twenty masculine, twenty feminine, and twenty social acceptability adjectives developed by Bem. The text character evaluation form was used.

3. The comparison revealed the sex role of the character as feminine, masculine or socially acceptable. A sample evaluation is shown in Table 1.

Table 1

Text Character Evaluation Form

Character	Personality Role in Text	According to Bem's Sex-Role		
		Masculine	*Feminine*	*Social Acceptability*
Big Ant	- Mentor - Sensitive	Acting like a leader	- Compassionate Sensitive about other's needs	-
Three Young Men	Arrogant			Arrogant

One of the characters in the text, "Big Ant", was determined to be a leader and sensitive as shown in Table 1. According to Bem's Sex-role Inventory, the mentor is the leader (masculine) and compassionate (feminine). The responsive of being sensitive is to be sensitive to wishes of the others (feminine). The other character in the other text, "Three Young Man", was arrogant. According to Bem's Sex-role Inventory, arrogant is the arrogance of social acceptability. After this analysis, the characters' sex role could be easily seen, if she\he acted feminine, masculine or\and in a social acceptability role.

The number of males and females, colors used for males and females, clothes, division of labor, professions, and physical appearances of the characters in textbooks' pictures were evaluated via descriptive analysis. Picture evaluation form was used.

Results

Findings about Gender Roles of Text Characters in terms of Bem's Sex-Role

Gender roles of the text characters in the textbooks according to Bem's Sex-Role is shown in Table 2.

Table 2

Gender Roles of the Text Characters in the Textbooks According to Bem's Sex-Role

Book	Sex	n	Feminine	Masculine	Social Acceptability
First grade	Female	11	9	-	7
	Male	8	11	5	1
	No sex	9	6	7	7
	Couple	-	-	-	-
Second grade	Female	5	7	-	4
	Male	12	11	-	7
	No sex	20	7	14	12
	Couple	1	-	2	-
Third grade	Female	14	24	11	12
	Male	12	24	28	10
	No sex	17	22	7	13
	Couple	2	1	-	1
Fourth grade	Female	7	5	6	2
	Male	12	8	11	3
	No sex	6	4	3	3
	Couple	1	-	2	-

As can be seen in Table 2, eleven female characters in the first grade Turkish textbook had nine feminine and seven social acceptability roles. Eight male characters in the texts had eleven feminine, five masculine, and one social acceptability roles. The texts in the books had characters that were explained without stating any sex. Some characters were depicted as females or males in the pictures although their sexes were not stated in the text. Nine characters whose sexes were not stated in the texts had six feminine, seven masculine, and seven social acceptability roles. The second grade Turkish textbook depicted five female characters with seven feminine and four social acceptability roles. Twelve male characters in the texts had eleven feminine and seven social acceptability roles. Twenty characters whose sexes were not stated in the texts had seven feminine, fourteen masculine, and twelve social acceptability roles. The books also had a couple (e.g. parents) of characters, and these characters had masculine roles.

The third grade Turkish textbook depicted fourteen female characters with twenty-four feminine, eleven masculine, and twelve social acceptability roles. Twelve male

characters in the texts had twenty-four feminine, twenty-eight masculine, and ten social acceptability roles. Seventeen characters whose sexes were not stated in the texts had twenty-two feminine, seven masculine, and thirteen social acceptability roles. The fourth grade Turkish textbook depicted seven female characters with five feminine, six masculine, and two social acceptability roles. Twelve male characters in the texts had eight feminine, eleven masculine, and three social acceptability roles. Six characters whose sexes were not stated in the texts had four feminine, three masculine, and three social acceptability roles. The books also had a couple (e.g. parents) of characters, and these characters had masculine roles.

Findings about the Evaluation of Pictures in terms of Gender Stereotypes

The number of males and females, the colors used for males and females, the clothes of males and females, the environment where males and females were depicted, the division of labor, professions, and physical attributes were examined in the textbooks' pictures.

Table 3

The Number of Males and Females

Grade	Female	Male	Total
First	216	263	479
Second	132	270	402
Third	173	244	417
Fourth	155	225	380

Primary school Turkish textbooks from the 1st to the 4th grades were examined, and numbers of female and male pictures were determined as shown in Table 3. The first grade text books had two hundred and sixteen female and two hundred and sixty-three male pictures; the second grade textbooks had one hundred and thirty-two female and two hundred and seventy male pictures; the third grade textbooks had one hundred and seventy-three female and two hundred and forty-four male pictures; and the fourth grade textbooks had one hundred and fifty-five female and two hundred and twenty-five male pictures.

The colors used in the female and male pictures in the textbooks were examined. The most used colors in the first grade textbooks were green (n = 52), white (n = 52), blue (n = 45), red (n = 41), pink (n = 31), purple (n = 25), orange (n = 25), yellow (n = 22), and brown (n = 19) for females, while the most used colors for males were green (n = 110), white (n = 73), brown (n = 49), blue (n = 47), grey (n = 36), orange (n = 34), yellow (n = 28), red (n = 24), and purple (n = 19). The most used colors in the second grade textbooks were white (n = 68), green (n = 20), yellow (n = 17), blue (n = 17), checkered (n = 16), grey (n = 15), orange (n = 14), red (n = 13), pink (n = 12), and black (n = 12) for females, while the most used colors for males were white (n = 120), brown (n = 80), green (n = 62), red (n = 44), blue (n = 34), black (n = 25), grey (n = 23), and orange (n = 21). The most used colors in the third grade textbooks were blue (n = 33), green (n = 27), orange (n = 23), purple (n = 21), pink (n = 21), red (n = 10), and brown

(n = 5) for females, while the most used colors for males were brown (n = 64), green (n = 56), blue (n = 51), orange (n = 37), yellow (n = 28), purple (n = 27), white (n = 20), red (n = 17), and purple (n = 7). The most used colors in the fourth grade textbooks were white (n = 38), blue (n = 33), pink (n = 13), orange (n = 11), yellow (n = 10), black (n = 8), and red (n = 7) for females, while the most used colors for males were blue (n = 34), white (n = 30), grey (n = 25), green (n = 18), brown (n = 16), orange (n = 13), yellow (n = 10), and red (n = 9). Female and male characters wore different colors. Examples of pictures are given below in Figure 1 and 2.

Figure 1. Colors (Third Grade)   **Figure 2.** Colors (First Grade)

The clothes of males and females in the primary school Turkish textbooks were examined. The clothes worn the most in the first grade textbooks were in the following order: dress, shirt, t-shirt, and skirt for females; shirt, trousers, t-shirt, and vest for males. The most frequently worn clothes in the second-grade textbooks were in the following order: skirt, t-shirt, shirt, trousers, blouse, and dress for females; shirt, trousers, t-shirt, shalwar, and sweater for males. The clothes worn in the third grade textbooks were in the following order: skirt, dress, shirt, and trousers for females; shirt, trousers, military uniform, and shalwar for males. The most frequently worn clothes in the fourth-grade textbooks were in the following order: shirt, blouse, trousers, t-shirt, and shalwar for females; trousers, shirt, jacket, and t-shirt for males. It was observed that females were depicted wearing dresses and skirts in general while males were depicted with trousers and shirts. Examples of pictures are given below in Figure 3-4.

Figure 3. Clothes (Third Grade)   **Figure 4.** Clothes (Fourth Grade)

Table 4

The Places Where Males and Females are Depicted in Textbooks

Grade	Place	Female	Male
First	Indoor	44	30
	Outdoor	76	116
Second	Indoor	8	24
	Outdoor	87	166
Third	Indoor	25	37
	Outdoor	59	129
Fourth	Indoor	15	24
	Outdoor	62	70

Table 4 shows the amount of males and females depicted in and out of the home in the primary school Turkish textbooks from the 1st to the 4th grades. Males and females were depicted more outside their homes in all textbooks. The number of pictures showing males at home was more than the ones showing females at home.

Another aspect reviewed in the textbooks was the co-operation in the pictures. The primary school (1st to 4th grade) textbooks had the following pictures: a mother looking after her child (n = 19), a woman in the kitchen (n = 3) (Figure 6), a grandmother looking after children (n = 2), a woman doing shopping (n = 1), a woman cleaning (n = 1), a women doing the laundry (n = 1), a woman doing handicraft (n = 1), women by the fountain (n = 1), men working outdoors (n = 8), man riding a bicycle (n = 2), and a father looking after his children (n = 6). The textbooks had more home-related pictures of females. Furthermore, there were pictures depicting males and females together doing the following: playing games (n = 17), working (n = 8), taking a holiday(n = 5),

catching fish (n = 2), wandering (n = 2), watching television, doing sports, swimming in the sea, planting saplings, talking, flying a kite, walking, collecting garbage, and riding bikes. Also there were activities performed as a family such as reading books (n = 1). Moreover, there were family pictures and pictures showing parents and children together (n = 4). In one picture, the mother was looking after her child while the father was helping his daughter with her lessons (Figure 5). Examples of pictures were given below in Figure 5-6.

Figure 5. Co-operation (Fourth Grade)   **Figure 6.** Co-operation (Second Grade)

Occupations of males and females were examined in the Turkish textbooks. Females were depicted as students (n = 16), farmers (n = 8), teachers (n = 5), workers (n = 4), sportswomen (n = 3), soldiers (n = 3), scientists (n = 3), nurses (n = 2), greengrocers, doctors, waitresses, receptionists, and astronauts. Males were depicted while dealing with many occupations such as soldiers (n = 39), shepherds (n = 10), farmers (n = 9), students (n = 8), astronauts (n = 6), scientists (n = 4), sultans (n = 4), sportsmen (n = 4), architects (n = 3), apprentices (n = 3), engineers (n = 3), beekeepers (n = 2), bakers (n = 2), doctors (n = 2), trainers (n = 2), ice cream sellers, football players, grooms, greengrocers, stallholders, drummers, barbers, teachers, canteen owners, referees, policemen, insurance brokers, matadors, waiters, chauffeurs, and newsmen etc. Examples of pictures are given below in Figure 7 and 8.
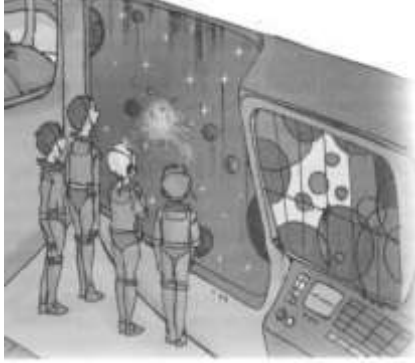
Figure 7. Occupations (Fourth Grade)

Figure 8. Occupations (Fourth Grade)

Physical appearances were examined in the primary school Turkish textbooks in terms of weight, hair color, and hairstyle, skin color for females; and hair color, hair length, and skin color for males. Females and males were depicted mainly as slim with the exception of a few overweight females and males. Females in the first, second, third and fourth grade Turkish textbooks had black, brown, red or yellow hair. Their hair was depicted with a knob, double or single pony tails. Haircuts were usually pageboy style; and short hair was rare. Females were depicted with normal height and weight. Very few overweight people were included, and they were old people. Females in the pictures were mostly light skinned. Males' hair was mostly neither very short nor very long in the first, second, third and fourth grade textbooks. The hair colors were black, brown, white, and red. Males were depicted as light skinned with a shaven beard. There were also bald and bearded men. Examples of pictures are given below in Figure 9-10.

Figure 9. Physical appearances (Third Grade)

Figure 8. Physical appearances (Second Grade)

Discussion, Conclusion and Recommendations

Textbooks are one of the most important, indispensable, cheap, and easily accessible education materials. It is easy for a teacher to ask students to have textbooks instead of other materials so that the students can follow the course, and their influence on individuals is important. This fact makes text books a significant topic of research. Expressions included in textbooks affect stereotypes of societies. One of these stereotypes is gender stereotypes. This study aims to evaluate the text characters and pictures in the primary school Turkish textbooks from the 1st to the 4th grades in terms of Bem's Sex-role Inventory and gender stereotypes, respectively.

Textbooks are an important factor in the educational dimension. The part played by textbooks, and the number of people reached by them cannot be denied. The qualities of the characters in primary school Turkish textbooks, the distribution of sex roles, and their effect on gender stereotypes within society must be evaluated. The importance of androgyny and "being stuck in traditional gender roles" must be emphasized and considered especially at the "educational" dimension (Demirtas, 2002). The texts included male characters with feminine and masculine roles, and female characters with feminine and masculine roles. It was observed that the characters had androgynous roles and it can be argued that using more texts with characters possessing such qualities may have a positive effect on children's opinions about gender. One of the most significant properties of some texts was that the sexes of the characters were not stated. This may enable the reader to envisage the characters based on his/her own sex, and break down social gender stereotypes. However, the fact that the designers of the text images depicted the sexes not specified in the texts as male in some sections reinforces gender stereotypes.

The number of males was more than the number of females in the primary school Turkish textbooks. There are other studies that have revealed the same results in different textbooks (Chick & Altoona, 2006; Cubukcu & Sivaslıgil, 2007; Gharbavi, & Mousavi, 2012; Lee & Collins, 2009; Yaylı & Kitis Cınar, 2014). Yaylı and Kitis Cınar (2014) state that characters are depicted in harmony with gender stereotypes. It was seen that colors that were in contrast with gender stereotypes were used in the textbooks. Similar colors can be used for males and females. Cinar (2015) states that females are depicted in pink. All colors were used for both males and females in the textbooks examined in this study. It was seen that males were shown with pink, green, yellow etc. in traditional texts such as Kaloghlan, Nasreddin Hodja etc.

Dress, skirt, shirt, and t-shirt were mostly worn by the females in the textbooks. Male clothes consisted of trousers, shirt, and t-shirt although there were texts including traditional clothing as well. Cinar (2015) stated that the children's books he examined included an apron, ribbon, skirt, and dress as the most common clothing materials reflecting the appearance of females. Although females use trousers frequently in daily life, they were depicted mostly with skirts and dresses in the pictures.

As there were more males in the pictures, all settings, including individual settings, favor the male. The reason was that the number of males was more than the number

of females in the books, and this affected the number of pictures showing males in and out of their homes. On the other hand, the number of females out of their homes had increased in direct proportion to their number. It could be said that males were seen at home more frequently. Poarch & Monk-Turner (2001) state that pictures showing females out of their homes have increased in award-winning children's books.

The textbooks had more home-related pictures of females. It was seen that the division of labor between males and females was not reflected sufficiently in the textbooks. Anderson and Hamilton (2005) state that mothers establish more relationships with their children, feed their children, and express their feelings more than the fathers in children's books. According to Poarch & Monk-Turner (2001), even award-winning children's books more frequently depict female characters doing housework.

It was seen that females were depicted as students, farmers, workers, teachers, sportswomen, scientists, soldiers, nurses, waitresses, receptionists, and astronauts. Males were depicted dealing with many other different jobs. Various studies have revealed that males and females deal with traditional occupations (Lee, & Collins 2009; Yaylı & Kitis Cinar, 2014). Cubukcu and Sivaslıgil (2007) state that females are depicted as students, teachers, nurses etc. while males are depicted doing many different jobs in the 7th grade English textbook. The job varieties of females had increased in the examined textbooks but they could not keep up with the variety in male occupations.

Apart from the sex differences between female and male pictures, males and females were not much different among themselves in terms of physical appearance. This may limit children's male or female perception to certain types, and cause them to act with prejudice, negative and exclusionary attitudes towards individuals with different physical qualities. This will cause stereotypes on physical appearance.

In recent years, violence against women in Turkey is getting more frequent. Children can live or watch this violence, and all these affect their gender perception. Moreover, this situation makes families nervous especially who have girls. The solution of this problem is showing the similarities of both women and men by means of textbooks, television etc. Textbooks constitute a significant medium when the subject of sex is discussed through similarities instead of differences. It is emphasized in the textbooks of various countries (Australia, the USA, Japan, Iran etc.) that male and female sexes must be included equally in pictures and texts. The Ministry of National Education and the authors of textbooks have a major responsibility for realizing the aim of raising gender awareness and overcoming inequality (Kuscu Kucukler, 2014). Teachers may select supplementary books after reviewing them in terms of gender stereotypes even though they cannot choose the textbooks. Publishing houses that are sensitive about this subject should be praised and rewarded by the Ministry of National Education. Speaking about negative examples reinforces the negative behavior, but more people can be made aware of good examples. Texts including androgynous characters can be placed in textbooks. Furthermore, sex roles of text characters may be examined with adjectives other than those included in Bem's Inventory.

References

Anderson, D. A., & Hamilton, M. (2005). Gender role streotyping of parents in children's picture books: The invisible father. *SexRoles*, *52* (3/4), 145-151.

Aslan, G. (2015). Öğretmen adaylarının toplumsal cinsiyet algılarına ilişkin metaforik çözümleme [A Metaforic analysis regarding gender perception of preservice teachers]. *Eğitim ve Bilim*, *40*, 363-384.

Asan Tezer, H. (2010). Ders kitaplarında cinsiyetçilik ve öğretmenlerin cinsiyetçilik algılarının saptanması [Gender differences in textbooks and teachers' genderperceptions]. *Fe Dergi*, *2*(2), 65-74.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155-162.

Bhasin, K. (2014). *Understanding gender: Genderbasics*. New Delhi: RajPress.

Chick, A. K., & Altoona P. S. (2006). Gender balance in K-12 American history textbooks. *Social Studies Research and Practice*, *1*(3), 284-290.

Cüceloğlu, D. (1993). *İnsan ve davranışı* (Dördüncü basım). İstanbul: Remzi Kitabevi.

Catalcali Soyer, A. (2009). Okul öncesi dönem çocuk hikâye kitapları: Stereotipler ve kimlikler [Storybooks of thepre-schoolkids: streotypes and identities]. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *1*(1), 13-27.

Cinar, P. (2015). Resimli çocuk kitaplarında giyimiyle kadın karakterler [Female characters in children's pictorial books]. Ankara University *Journal of Faculty of Educational Sciences*, *48*(1), 1-18.

Çotuksoken, B. (2003). Giriş (1-22). *Ders kitaplarında insan hakları: Tarama sonuçları*. İstanbul: Tarih Vakfı Yayınları.

Cubukçu H., & Sivaslıgil, P. (2007). 7. Sınıf İngilizce ders kitaplarında cinsiyet kavramı [The concept of gender in seventh grade eltcoursebooks]. *Dil Dergsi*, *137*, 7-17.

Demirtas, H. A. (2000). Cinsiyet rolü kalıpyargıları, androjenlik ve diğer cinsiyet rolü yönelimleri [Gender role stereotypes, androgyny and gender role orientations]. *Kadın/Woman*, 1-16.

Dokmen, Z. Y. (2015). *Toplumsal cinsiyet: Sosyal psikolojik açıklamalar* (6. Basım). İstanbul: Remzi Kitabevi.

Dokmen, Z. Y. (1999). Bem cinsiyet rolleri envanteri kadınsılık ve erkeksilik ölçekleri Türkçe formunun psikometrik özellikleri [Turkish version of the Bem sex role inventory abstract]. *Kriz Dergisi*, *7*(1), 27-40.

Elgar, A. G. (2004). Science textbooks for lower secondary schools in Brunei: issues of gender equity. *International Journal of Science Education*, *26* (7), 875-894.

Esen, Y. (2003). Okul bilgisi ve ders kitapları. In *Ders kitaplarında insan hakları: İnsan haklarına duyarlı ders kitapları için* (pp. 5-19). İstanbul: Tarih Vakfı Yayınları.

Esen,Y., & Bağlı, M. T. (2002). İlköğretim ders kitaplarındaki kadın ve erkek resimlerine ilişkin bir inceleme [A research on male and female figures in the elementary school textbooks]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 35*(1-2), 143-154.

Giddens, A. (2013). *Sosyoloji.* İstanbul: Kırmızı.

Gharbavi, A., & Mousavi, S. A. (2012). A content analysis of textbooks: Investigating gender bias as a social prominence in Iranian highschool English textbooks. *English Linguistics Research*, *1*(1), 42-49.

Gol, M. (2011). *Çocuk edebiyatında cinsiyet rolleri: Mustafa Ruhi Şirin ve Ayla Çınaroğlu'nun eserleri [Gender roles in childrenliterature: Mustafa Ruhi Şirin's and Ayla Çınaroğlu's Works]* (Master Dissertation). Afyon Kocatepe University The Institute of Social Sciences Department of Sociology, Afyon.

Gumusoglu, F. (2008). Ders kitaplarında toplumsal cinsiyet [Gender in school textbooks]. *Toplum ve Demokrasi, 2*(4), 39-50.

Güneş, Ö. (2008). Ders kitaplarında toplumsal cinsiyet ayrımı (1990- 2006) [Gender discrimination in school books (1990-2006). *Toplum ve Sosyal Hizmet, 19*(2), 81-95.

Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem Sex-Role inventory. *Sex Roles*, *39*(11/12), 929-941.

Kavuncu, N. (1987). *Bem cinsiyet rolü envanteri'nin Türk toplumuna uyarlanması çalışması* (Master Dissertation). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Kırbaşoğlu Kılıç, L., & Eyüp, B. (2011). İlköğretim Türkçe ders kitaplarında ortaya çıkan toplumsal cinsiyet rolleri üzerine bir inceleme [An investigation on societal sexual roles formed in elementary school Turkish school books]. *ODÜ Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi, 2*(3), 129-148.

Kuscu Küçükler, N. (2014). Toplumsal cinsiyet ve edebiyat öğretimi: ortaöğretim Türk edebiyatı ders kitaplarında kadın yazarlar [Gender and teaching literature: Women writers in secondary school Turkish Literature textbooks]. *Journal of Research in Education and Teaching*, *3* (2), 195-202.

Lee J. F. K., & Collins, P. (2009). Australian English-language text books: The gender issues. *Gender and Education*, *21*(4), 353-370.

Mendoza, J., & Reese, D. (2001). Examining multicultural Picture books fort he early childhood classroom: Possibilities and pitfalls. *Early Childhood Research & Practice*, *3*(2), 1-21.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (Second edition). California: SAGE Publications.

Ozbay, M. (2006). *Türkçe özel öğretim yöntemleri I.* Ankara: Öncü Kitap.

Ozkan, R. (2013).İlköğretim Ders Kitaplarında Kadın Figürü [The women figüre in primary school coursebooks]. *The Journal of Social Science Studies, 6*(5), 617-631.

Paterson, S. B., & Lach, M. A. (1990). Gender stereotypes in children's boks: Their prevalence and influence on cognitive and affective development. *Gender&Education*, *2*(2), 185-198.

Pingel, F. (2010). *UNESCO guidebook on textbook research and textbook revision* (Second revised and updated edition.). Paris: United Nations Educational Scientific and Cultural Organization and the GeorgEckert Institute for International Textbook Research Celler.

Poarch, R., & Monk-Turner, E. (2001). Gender roles in children's literature: A review of non-award-winning "easy-to-read" books. *Journal of Research in Childhood Education*, *16*(1), 70-76.

Sano, F., Lida, M., & Hardy, T. (2001). *Gender representation in Japanese EFL textbooks*. PACS at JALT Conference Proceedings, 22-25 November, Kitakyushu, Japan.

Sever, S. (2003). *Çocuk ve edebiyat*. Ankara: Kök Yayıncılık.

Vannicopulou, A. (2004). Resimli çocuk kitaplarında kadın karakterler (Çev. Dilek Ciftci Yesiltuna). *Ege Eğitim Dergisi*, *5*(2), 65-73.

Worland, J. (2008). Girls will be girls…and soon: Treatment of gender in preschool boks from 1960 through 1990. *Children and Libraries*, Spring, 42-46.

Yanpar, T. (2006). *Öğretim teknolojileri ve materyal tasarımı* (7. Baskı). Ankara: Anı Yayıncılık.

Yaylı, D., & Kitiş Çınar, E. (2014). Ortaokul ders kitapları görsellerinde toplumsal cinsiyet [Social gender in middle school Turkish coursebook visuals]. *Turkish Studies*, *9* (5), 2075-2096.

Yildirmm, A., & Şimşek, H. (2005). *Sosyal bilimlerde nitel araştırma yöntemleri* (Beşinci Baskı). Ankara: Seçkin Yayıncılık.

Yildiz, M. (2013). İlkokul ve ortaokul din kültürü ve ahlak bilgisi kitapları görsellerinin toplumsal cinsiyet açısından incelenmesi [Analysis visual components of elementary and middle school religious culture and moral knowledge textbooks in terms of gender]. *Dini Araştırmalar*, *16* (42), 143-165.

Çalışmada İncelenen Kaynaklar

Demir, E., Bozbey, S., Oğan, M., Özkara, M., & Aktaş, A. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 1.* Ankara: Özgün Matbaacılık.

Demir, E., Bozbey, S., Oğan, M., Özkara, M., & Aktaş, A. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 2.* Ankara: Özgün Matbaacılık.

Demir, E., Bozbey, S., Oğan, M., Özkara, M., & Aktaş, A. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 3.* Ankara: Özgün Matbaacılık.

Hengirmen, M., Öziş, N., & Öngay, N. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 1.* Ankara: Engin Yayınevi.

Hengirmen, M., Öziş, N., & Öngay, N. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 2.* Ankara: Engin Yayınevi.

Hengirmen, M., Öziş, N., & Öngay, N. (2015). *İlköğretim Türkçe 3. Sınıf ders kitabı ve çalışma kitabı 3.* Ankara: Engin Yayınevi.

Temur, Ö. D. (2015). *İlköğretim Türkçe 1. Sınıf ders kitabı ve çalışma kitabı 1.* Ankara: Yıldırım Yayınları.

Temur, Ö. D. (2015). *İlköğretim Türkçe 1. Sınıf ders kitabı ve çalışma kitabı 2.* Ankara: Yıldırım Yayınları.

Temur, Ö. D. (2015). *İlköğretim Türkçe 1. Sınıf ders kitabı ve çalışma kitabı 3.* Ankara: Yıldırım Yayınları.

Temur, Ö. D., Aksu, S., Aktaş, S., & Aksu, D. (2015). *İlköğretim Türkçe 2. Sınıf ders kitabı ve çalışma kitabı 1.* Ankara: Gizem Yayıncılık.

Temur, Ö. D., Aksu, S., Aktaş, S., & Aksu, D. (2015). *İlköğretim Türkçe 2. Sınıf ders kitabı ve çalışma kitabı 2.* Ankara: Gizem Yayıncılık.

Temur, Ö. D., Aksu, S., Aktaş, S., & Aksu, D. (2015). *İlköğretim Türkçe 2. Sınıf ders kitabı ve çalışma kitabı 3.* Ankara: Gizem Yayıncılık.

İlkokul (1-4. Sınıf) Türkçe Ders Kitaplarında Toplumsal Cinsiyet Rollerinin İncelenmesi

Atıf:

Özet

Problem Durumu: Cinsiyete toplumsal bakış açısı, çocuklar tarafından doğrudan ya da dolaylı yollarla kazanılmaktadır. Ders kitapları, çocukların toplumsal cinsiyet kalıp yargılarını açık ya da örtük etkileme gücüne sahip önemli, ucuz, kolay ulaşılabilir materyallerdir. Bu çalışmada resimler ve metinler incelenerek, çocuklara nasıl mesajlar verildiği belirlenmektedir. Metinlerdeki karakterlerin cinsiyet rolleri Bem'in geliştirdiği envanterde yer verilen kadınsı, erkeksi ve sosyal kabul edilirlikle ilgili sıfatlara göre sınıflanmaktadır. Her kadın ve erkek duruma göre kadınsı ya da erkeksi roller gösterebilir. Androjen kimlik, cinsiyetler arasındaki benzerlikleri farklılıklardan daha fazla ortaya koyabilir. Bireyler kendi cinsiyeti ve karşı cinsi anlarsa toplumsal birliktelik güçlenecektir.

Araştırmanın Amacı: Bu çalışmada ilkokul 1-4. sınıf Türkçe ders kitaplarındaki metin karakterlerinin cinsiyet rolleri BEM'in Cinsiyet Rolleri Envanteri ve resimler ise toplumsal cinsiyet kalıp yargıları açısından incelenmiştir.

Araştırmanın Yöntemi: İlkokul 1-4. sınıf Türkçe ders kitaplarındaki metin karakterlerinin cinsiyet rolleri BEM'in Cinsiyet Rolleri Envanteri, resimler ise

toplumsal cinsiyet kalıpyargıları açısından incelemek amacıyla doküman incelemesi yapılmıştır. Çalışmada ilkokul Türkçe 1-4. Sınıf on iki kitap bu araştırmanın çalışma materyalini oluşturmaktadır. Burdur ilinde 2015-2016 eğitim öğretim yılında okutulan kitaplar arasından seçim yapılmıştır. Araştırmada metinlerdeki karakterleri değerlendirmek için BEM'in Cinsiyet Rolleri Envanterine göre düzenlenmiş "Metin Değerlendirme Formu" ve resimleri değerlendirmek için "Resim Değerlendirme Formu" kullanılmıştır.

Metin karakterlerini değerlendirme formu. Bem, cinsiyet rollerini test etmek için, bir ölçek geliştirmiştir. Erkeksilik (20 madde), kadınsılık (20 madde) ve sosyal kabul edilirlik (20 madde) olmak üzere üç alt ölçekten, toplam 60 maddeden oluşmaktadır. Bem'in kısa formu (20 kadınsı, 20 erkeksi) Türkçe'ye Kavuncu (1987) tarafından uyarlanmış, geçerlik güvenirlik çalışması Dökmen (1999) tarafından yapılmıştır.

Resim Değerlendirme Formu. Alanda yapılan çalışmalardan yola çıkılarak madde havuzu oluşturulmuştur. Uzman görüşleri ve ön uygulama çalışmaları ile forma son şekli verilmiştir. Resim değerlendirme formu; kadın-erkek sayısı, kadın ve erkeklerde kullanılan renkler, giyilen kıyafetler, işbölümü, meslekler, karakterlerin fiziksel görünüşleri, aile yapısı (çekirdek aile, geniş aile, tek ebeveynli aile ve parçalanmış aile) maddelerinden oluşmaktadır. 3. Sınıf Türkçe 3. Kitabı üç araştırmacı tarafından ayrı ayrı hem resim hem de metin açısından incelenmiştir. Resim Değerlendirme Formu, puanlayıcılar arası güvenirlik puanı (Miles, Huberman, 1994).81 olarak bulunmuştur.

Metindeki karakterlerin cinsiyet rollerini incelenmek amacıyla, Bem'in ölçeğinde yer verdiği kadınsı, erkeksi ve sosyal kabul edilirlik sıfatlarından (Dökmen, 1999, Holt, Ellis, 1998, Monto, 1993) yararlanılmıştır. Envanterin geçerlik çalışması, değişen toplumsal koşullardan dolayı Holt ve Ellis (1999) tarafından tekrarlanmış ve envanterde kullanılan cinsiyet rolleri ile ilgili sıfatların geçerli olduğu belirlenmiştir. Türkçe'ye kısa form çevrildiği için kalan sosyal kabul edilirlikle ilgili 20 sıfat ikinci araştırmacı tarafından Türkçe'ye çevrilmiş, İngilizce Öğretmenliği Anabilim Dalından bir uzman tarafından da kontrol edilmiştir. Form tema adı, metin adı, karakter adı, roller, kadınsı, erkeksi ve sosyal kabul edilirlik maddelerinden oluşmaktadır. Veriler, nitel araştırma analiz yöntemlerinden betimsel analiz ile değerlendirilmiştir.

Araştırmanın Bulguları: Metinlerdeki karakterlerde kadınsı, erkeksi ve sosyal kabul edilirlik rolleri incelenmiştir. Üçüncü sınıf Türkçe Kitabında 14 kadın karakter; 24 kadınsı, 11 erkeksi, 12 sosyal kabul edilirlik rollerine sahip olarak betimlenmiştir. Metinlerdeki 12 erkek karakter; 24 kadınsı, 28 erkeksi ve 10 sosyal kabul edilirlik rollerine sahiptir. Metinlerde cinsiyeti belirtilmeyen 17 karakter, 22 kadınsı, 7 erkeksi, 13 sosyal kabul edilirlik rollerine sahiptir. Araştırmada resimlerde birinci sınıf ders kitabında 216 kadın, 263 erkek, ikinci sınıf kitabında 132 kadın, 270 erkek, üçüncü sınıf kitabında 173 kadın, 244 erkek, dördüncü sınıf kitabında 155 kadın, 225 erkek resmi bulunmaktadır. Kadınlar; öğrenci (n=16), çiftçi (n=8), öğretmen (n=5), işçi (n=4), sporcu (n=3), asker (n=3), bilim insanı (n=3), hemşire (n=2), manav, doktor, garson, resepsiyon görevlisi, astronot olarak resimlenmiştir. Erkekler; asker (n=39), çoban (n=10), çiftçi (n=9), öğrenci (n=8), işçi (n=8), astronot (n=6), bilim adamı (n=4), sultan (n=4), sporcu (n=4), mimar (n=3), çırak (n=3), mühendis (n=3), arıcı (n=2), fırıncı (n=2),

doktor (n=2), antrenör (n=2), dondurmacı, futbolcu, seyis, manav, pazarcı, davulcu, berber, öğretmen, kantinci, hakem, polis, sigortacı, matador, garson, şoför, spiker gibi erkekler bir çok mesleği yapmaktadır.

Araştırmanın Sonuçları ve Önerileri: Metinlerde erkek karakterler kadınsı ve erkeksi roller, kadın karakterler de kadınsı ve erkeksi roller göstermektedirler. Karakterlerin androjen roller gösterdiği görülürken, bu özelliklere sahip karakterlerin olduğu daha fazla metne yer verilmesinin çocukların cinsiyet ile ilgili düşüncelerini olumlu etkileyebileceği söylenebilir. Bazı metinlerin en önemli özelliklerinden biri de karakterlerin cinsiyetinin belirtilmemiş olmasıdır. Bu da okuyucunun, kendi cinsiyetini temel alarak, karakterleri zihninde canlandırmasını sağlayabilir. İlkokul Türkçe ders kitaplarında erkek sayıları kadın sayılarından sayıca daha fazladır. Erkeklerin sayıları resimlerde fazla olduğu için, bireylerin bulunduğu bütün ortamlar erkeklerin lehinedir. Kitaplarda kadın ve erkek resimlerinde bütün renklerin kullanıldığı görülmektedir. Erkeklerin kadınlara göre daha farklı mesleklerde resmedildiği görülmektedir. Ders kitaplarında kadınların ev ile alakalı daha fazla resmi bulunmaktadır.

Anahtar Sözcükler: Ders kitapları, Toplumsal cinsiyet, Androjenlik, Bem.

Predictive Role of Narcissism and Family Relations on Decision-making Characteristics of Secondary School Students*

Durmus UMMET[1], Halil EKSI[2], Ahmet ERDOGAN[3]

ARTICLE INFO	ABSTRACT
	Purpose: This study aimed at examining the effects of narcissism and family relations on secondary school students' decision-making characteristics. The study also investigated whether or not students' decision-making characteristics differentiated according to gender. **Research Methods:** Participants were secondary school students who were studying at different secondary schools in Istanbul, and who were selected randomly on a voluntary basis. Formed by a relational screening model, data were collected using the Adolescent Decision-Making Questionnaire for

determining decision-making characteristics, the Childhood Narcissism Scale for determining narcissism characteristics, and the Family Relationship Scale for Children for determining family relationships. Regression analysis was used to determine whether or not students' narcissism and interfamily relations predicted their decision making characteristics, and independent samples t test was used to determine whether or not students' decision making characteristics differentiated according to gender.

Results: The findings showed that narcissism characteristics of secondary school students significantly predicted self-esteem, vigilance, and complacency characteristics in decision-making. Both discouraging and supportive relations of families significantly predicted self-esteem, vigilance, complacency, panic, and cop-out characteristics in decision-making. It was also found that gender was a significant factor.

Implications for Research and Practice: Both narcissism and family relations of students significantly predicted their decision making characteristics; also there were significant differences according to gender variances in decision making. The results suggest that families should be educated regarding their attitudes on raising children, and that educators and teachers should work together to support adolescents' decision-making process in a positive direction. Moreover, different studies about this issue should be conducted.

---

[1] Marmara University, TURKEY, e-mail: dummet@marmara.edu.tr, ORCID: https://orcid.org/0000-0002-8318-9026

[2] Marmara University University, TURKEY, e-mail: h.eksi70@gmail.com, ORCID: https://orcid.org/0000-0001-7520-4559

[3] Ministry of National Education, TURKEY, e-mail: ahmeterdogan.psk@gmail.com, ORCID: https://orcid.org/0000-0003-0374-4104

Introduction

The last years of middle school are a period when students begin to undergo physical, emotional, social and other changes, and when entering adolescence is strongly emphasised. This period, which can also be regarded as a significant period of change in a person's life, is considered in the present study within the context of several variables. Various topics related to adolescence have received attention from educators, psychologists, philosophers, sociologists and parents (Eksi, 1990). Adolescence, from the Latin *adolescere* (to grow up), is characterised as the age between the end of childhood and the transition to young adulthood (Arnett, 2000; Schvaneveldt & Adams, 2001; Yavuzer, 1993). It is a particular stage in which physical growth; hormonal, sexual and social development; and emotional, personal and mental changes occur in an individual, beginning at puberty, and is considered to end when physical growth stops. In addition to these changes, society today assigns certain roles to young people who are approaching adulthood (Adams, 1995; Dacey & Kenny, 1994; Kulaksizoglu, 1998). In short, adolescence is a developmental period whose significance has been emphasised by researchers in the field with motions towards these dynamics.

The families, communities and educational institutions of adolescent individuals hold many expectations of these young people. As they are no longer children due to the development that occurs specifically in this period, adolescents are expected to start making more critical decisions, whether related to their own life or to their environment. Many research studies have investigated how decisions made during adolescence can have an impact on an individual's entire life (Arnett, 2004; Atkinson, Atkinson, Smith, Bem, & Nolen-Hoeksema, 1999; Colakkadioglu, 2003; Deniz, 2006; Ersever, 1996; Kilicci, 2000; Kurt, 2003; Mann, 1998; Naftel & Driscoll, 1993). These studies, whether conducted in Turkey or in other cultures, share the common perspective that the healthy choices young individuals make, have a positive influence on their lives in the period that follows. In summary, it can be said that the path to success in life comes from the ability to make healthy decisions based on knowing which choices are likely to produce beneficial results and which are not (Byrnes, 1998; Oztemel, 2012). The act of making decisions, as mentioned earlier, is demonstrated in every area and every moment of an individual's life.

The processes of decision-making and problem-solving resemble each other conceptually. Decision-making, which starts after a situation emerges that requires a decision, has been identified as a process that establishes how one will behave, what one will do, and when one will act in the face of this situation. Both concepts are also considered complex processes in which an individual determines the alternative that can obtain the most useful result from among the identified and evaluated options (Kasik, 2009; Miller, 2000; Phillips, Pazienza & Ferrin, 1984). Deciding, thus, does not refer to a single moment or phase but to a process. First, individuals encounter a crisis or situation, and then they set a path for themselves by taking advantage of the options that appear in front of them. Afterwards, they are affected by the results of the choice they make, which is either positive or negative (Charles, 2000; Colakkadioglu, 2003; Ozolins & Stenstrom, 2003). In other words, deciding is not a linear phenomenon that

starts and ends in a moment; rather, it is a complex action with internal stages, and it functions more in the process of a situation.

Some studies have attempted to explain the styles individuals use when making decisions, and the degree to which these styles affect their levels of self-esteem and anxiety. Janis and Mann (1977) established the conflict-theory model of decision-making, which addressed the styles used in decision-making. The styles in this approach are classified into five types: unconflicted adherence, unconflicted change, defensive avoidance, hypervigilance, and vigilance. One can argue that, regardless of whichever decision style an individual uses, the family is the first factor in influencing how the forms of behaviour related to an individual's decision-making are shaped (Noone, 2002; Rollinson, 2002). The family plays a key role in allowing adolescents to participate in decisions and in developing their competence while they are going through adolescence (Brown & Mann, 1990). The family, which plays the most significant role in an individual's development of personality and behaviour, is also critical for the individual to enjoy success while going through an important process such as puberty (Kulaksizoglu, 1998; Ladd & Petit, 2002). In a way, the family is the source for the way in which adolescents gain freedom, express themselves in society and, therefore, overcome difficulties they face (Morgan, 1991). According to Satir (2001), the worth that family members place on each other is high in healthy families, and each family member is supported in making his/her own decisions. As in all other developmental periods, it is also important for a family to be encouraging for an adolescent to acquire the skills of decision-making and autonomy (Cuhadaroglu, 2004). The effect of the family on adolescents' decision-making behaviour is encountered in most positions on child-raising as well as in personality development, which is shaped by this (Buss & Plomin, 1984). All these explanations emphasise the role of family in shaping the behaviours adolescents exhibit related to decision-making.

The critical role of the family in shaping a child's personality traits is well known and universally accepted. In this regard, narcissism is a phenomenon which reportedly appears to have increased in frequency these days (Twenge & Campbell, 2009), and where the effect of the family is a rather significant variable in its development (Kernberg, 2012). Narcissism means a person takes pleasure from and feels sexual desire towards his or her own body (Ozaydin, 1984). A narcissistic person is more interested in him/herself than others, and feels self-admiration. Narcissism in individuals occurs during infancy. Under normal conditions, a baby gradually begins to distinguish itself within a short period after birth as an entity separate from its mother/caregiver. Being able to experience this process in a healthy way depends on the mother providing the baby with enough love and trust. However, a mother's failure to satisfy a child's requirements for love and trust may cause the child to experience self-defensive anger, because his/her development remains stuck at a primitive level without separate boundaries of self (Atay, 2010; Masterson, 2006). Such fear and anger experienced in childhood may be another source in the formation of narcissism (Fromm, 1999). In fact, while narcissistic individuals have a deep lack of confidence in infrastructure, they think of themselves as unique and great. This also

causes them to feel strong and to act as if they are better than everyone else (Freud, 2007; Fromm, 1994; Kohut, 2004). In light of all this revealed information, it can be said that personality is one of the most important variables that provides direction for the behaviour of an individual, whose foundation begins within the family environment. As a determining factor in much behaviour, an individual's personality traits lie beneath his/her decision-making behaviours. It can be argued that narcissism, which in this context influences individual behaviour as a personality trait, provides direction in decision-making. With motion from this rationale, the current study may be considered as original research insofar as it may help experts who work with adolescent individuals develop different points of view about their behavioural dynamics.

This study, which aims at examining narcissism in middle school students, and the predictive role of family relationships in decision-making, seeks to address the following sub-problems:

i) To what degree do students' narcissistic qualities predict their decision-making behaviours?

ii) To what degree do students' family relations predict their decision-making behaviours?

iii) Do students' decision-making behaviours differ according to gender?

Method

Research Design

Because the present research aimed at investigating the effect of middle school students' narcissistic characteristics and patterns of family relations on their decision-making behaviours, the relational screening model was selected among quantitative research methods. This model attempts to detect whether or not variables jointly change and, if so, how this change occurs (Karasar, 1999).

Research Sample

Participants were 300 students in the seventh- and eighth-grade classrooms of five different middle schools in Istanbul, with 157 (52.3%) male and 143 (47.7%) female students, whose average age was 13.20. Of these students, 158 (52.6%) were in the seventh grade, and 142 (47.4%) were in the eighth grade. Stratified sampling, a probability-based sampling type, was used in the study, and this sampling offers a variety that guarantees that subgroups in the universe will be represented (Greasley, 2008). Because students in the seventh and eighth grades in middle school have more experience in making decisions, such as deciding on the type of high school, the fifth and sixth grades were not included in the study group.

Research Instruments and Procedures

Adolescent decision-making questionnaire. The Adolescent Decision-Making Questionnaire, developed by Mann, Harmoni and Power (1989) for preadolescents between the ages of 13 and 15 and adapted into Turkish by Colakkadioglu and Gucray

(2007), was examined for its coefficients of construct validity, correlations of subscales with each other, criterion-related validity, Cronbach's alpha of internal consistency, item-total score correlation and test-retest reliability. The result of confirmatory factor analysis performed for the scale's construct validity showed the scale, as adapted into Turkish for a population between 13 and 15 years old, to be in accordance with the original five sub-dimensions of the scale (self-esteem in decision-making, vigilance, complacency, panic and cop-out), and all its items were placed in the relevant subscale. Additionally, the scale's Cronbach alpha coefficient for the sub-dimensions of self-esteem, vigilance, complacency, panic and cop-out were found to be .84, .85, .83, .76 and .77, respectively, and the test-retest consistency was found to be .85, .79, .85, .67 and .78, respectively (Colakkadioglu, 2012).

Childhood narcissism scale. Validity and reliability of the Turkish version of this scale (the original scale was made by Thomaes, Stegge, Bushman, Olthof, & Denissen, 2008) was performed by Akın, Sahin and Gulsen (2015). As a result of the confirmatory factor analysis applied to the scale's structural validity, it was found appropriately to have one dimension (χ^2 = 49.88, *df* = 35, *p* = 0.04920, *RMSEA* = .042, *NFI* = .91, *IFI* = .97, *CFI* = .97, *GFI* = .96, *SRMR* = .050). The scale, made using a four-point Likert-type evaluation (0 = definitely not true; 1 = not true; 2 = partially true; 3 = definitely true), consisted of a total of 30 items. The internal consistency of reliability for the scale was found to be .72. The corrected item-total correlations of the scale range between .29 and .52 (Akın, et al., 2015).

Family relationship scale for children. This scale, developed by Demirtas-Zorbaz and Korkut-Owen (2013), was developed for children based on healthy family characteristics as identified by the McMaster Model, and by Krysan, Moore and Zill (1990) by looking at how children perceive family functions. Through explanatory factor analysis, the scale was reduced from 56 items to a two-factor 20-item structure using varimax rotation. This structure was tested through confirmatory factor analysis, and the scale's indices were found to have significant consistency. The scale had a two-factor structure, Supportive Family Relations (10 items) and Inhibitive Family Relations (also 10 items). Cronbach's alpha for the scale, which gave two separate scores according to the theoretical structure, was found for the two separate groups as .82 and .84 in the first sub-dimension, and .76 and .78 in the second dimension.

Data Analysis

The appropriate post-application data were entered into the programme SPSS-21, and the sub-dimension and total scores were calculated in line with the specifications of the scales. Parametric and non-parametric analyses were performed in accordance with the purpose after a series of calculations had been carried out to determine the suitability of the data for statistical analyses. At this point, Pearson's correlational analysis was first used for the purpose of being able to see the relationship among variables. Afterwards, multiple regression analysis was used to be able to see whether or not students' narcissistic qualities, with supportive internal family relations and discouraging internal family relations, were able to predict the sub-dimensions of their decision-making characteristics. Here, the two separate and inversely related variables of internally supportive family relations and internally discouraging family relations were analysed independently alongside narcissism. Also, because the decision-

making scale did not have a single total score, the sub-dimensions of decision-making were handled separately as dependent variables, and the analyses were made in accordance with this. Lastly, independent-samples *t*-test was used to determine whether decision-making characteristics differed according to students' gender.

Results

The mean and standard deviation values resulting from Pearson's correlational analysis of the relationships among dependent and predictor variables are shown in Table 1. Based on these, while no significant relationship was detected between internally supportive family relations and narcissism, a significant negative relationship was detected between internally discouraging family relations and narcissism. While a significant positive relationship was found for narcissism with self-esteem in decision-making, vigilance and complacency; no significant relationship was found for narcissism with panic or cop-out. While the scores for internally discouraging family relations had a significant negative relationship with self-esteem in decision-making, they had a significant positive relationship with complacency, panic and cop-out. However, no significant relationship existed between vigilance and internally discouraging family relations. Lastly, internally supportive family relations had a significant positive relationship with vigilance and self-esteem in decision-making, and a significant negative relationship with complacency, panic and cop-out.

Table 1

Means, Standard Deviations and Intercorrelations for Narcissism and Internally Discouraging/Supportive Family Relations with Subscales of Decision-making (N = 300)

	M	**SD**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1.	16.40	5.21	1							
2.	14.94	4.21	.082	1						
3.	25.49	4.13	.055	.449**	1					
4.	17.08	3.00	.143*	.293**	.316**	1				
5.	18.16	3.44	.245**	.067	.291**	.517**	1			
6.	10.48	2.94	.261**	.187**	−.142*	−.302**	−.118*	1		
7.	13.14	3.68	.102	.215**	−.184**	−.321**	−.061	.445**	1	
8.	10.91	3.41	.047	.278**	−.190**	−.279**	−.172**	.428**	.268**	1

*p < .05, **p < .001, ***p < .000
(1. Narcissism, 2. Discouraging relations, 3. Supportive relations, 4. Self-esteem, 5. Vigilance, 6. Complacency, 7. Panic, 8. Cop-out)

Through multiple regression analysis, both the features of narcissism and internally discouraging family relations were identified as having a significant predictor effect on the students' characteristics of self-esteem in decision-making according to the standardised *ß*-coefficients. Narcissism and discouraging relations together were found to explain 11% of the variance in the scores for self-esteem in

decision-making. The performed analyses showed the qualities of narcissism to be a significant predictor in the characteristics of students' vigilance, a sub-dimension of decision-making; however, internally discouraging family relations were not shown as being a significant predictor. Narcissism and discouraging relations together were found to explain 7% of the variance in the scores for vigilance. The performed analyses showed that both students' narcissistic features and their internally discouraging family relations were significant predictors of the characteristic of complacency, another sub-dimension of decision-making. Narcissism and discouraging relations together were found to explain 10% of the variance in the scores for complacency. However, while students' narcissistic qualities were identified as not being a significant predictor in the characteristic of panic, internally discouraging family relations were identified as a significant predictor of panic. Narcissism and discouraging relations together were found to explain 5% of the variance in scores for panic. Lastly, students' qualities of narcissism were not identified as being a significant predictor of the characteristic of cop-out; however, internally discouraging family relations were found as being a significant predictor of cop-out. Narcissism and discouraging relations together were found to explain 8% of the variance in the scores for cop-out (see Table 2).

Table 2

Simultaneous Multiple Regression Analysis for Narcissism and Discouraging Family Relations as Predictors of Decision-making Characteristics (N = 300)

Dependent Variables	Independent Variables	B	SE	β	t	R	R²	f	p
1	Narcissism	.097	.032	.168	3.070	.338	.114	19.107	.000
	Discouraging relations	−.219	.039	−.307	−5.600				
2	Narcissism	.166	.037	.252	4.483	.260	.068	10.765	.000
	Discouraging relations	−.072	.046	−.088	−1.561				
3	Narcissism	.140	.031	.248	4.475	.310	.096	15.732	.000
	Discouraging relations	.116	.039	.166	3.003				
4	Narcissism	.060	.040	.085	1.504	.231	.053	8.384	.000
	Discouraging relations	.182	.050	.208	3.672				
5	Narcissism	.016	.037	.024	.436	.279	.078	12.521	.000
	Discouraging relations	.223	.045	.276	4.932				

(1. Self-esteem, 2. Vigilance, 3. Complacency, 4. Panic, 5. Cop-out)

Through the multiple regression analysis that was performed, both the qualities of narcissism and internally supportive family relations were identified as having a significant predictive effect on students' characteristics of self-esteem in decision-making according to the standardised β-coefficients. Narcissism and supportive relations together were found to explain 12% of the variance in scores for self-esteem in decision-making. The performed analyses showed both narcissistic features and internally supportive family relations as being significant predictors of students' qualities of vigilance, a sub-dimension of decision-making. Narcissism and supportive relations together were found to explain 14% of the variance in scores for vigilance. The performed analyses showed both students' narcissistic qualities and internally supportive family relations as significant predictors of the characteristic of complacency, another sub-dimension of decision-making. Narcissism and supportive relations together were found to explain 9% of the variance in the scores for complacency. Thus, in terms of the analyses, both students' narcissistic features and internally supportive family relations were identified as significant predictors of the characteristic of panic. Narcissism and supportive relations together were found to explain 5% of the variance in scores for panic. Lastly, students' narcissistic features were found as not being a significant predictor of cop-out, whereas internally supportive family relations were found as being a significant predictor of cop-out. Narcissism and supportive relations together were found to explain 4% of the variance in scores for cop-out (see Table 3).

Table 3
Simultaneous Multiple Regression Analysis for Narcissism and Internally Supportive Family Relations as Predictors of Decision-making Characteristics (N = 300)

Dependent Variables	Independent Variables	B	SE	β	t	R	R²	f	p
1	Narcissism	.073	.031	.126	2.305	.340	.116	19.461	.000
	Supportive relations	.225	.040	.309	5.661				
2	Narcissism	.152	.036	.229	4.250	.370	.137	23.582	.000
	Supportive relations	.232	.045	.278	5.151				
3	Narcissism	.152	.031	.270	4.880	.305	.093	15.222	.000
	Supportive relations	−.112	.039	−.157	−2.840				
4	Narcissism	.080	.040	.113	1.988	.216	.047	7.257	.000
	Supportive relations	−.169	.051	−.190	−3.352				
5	Narcissism	.080	.040	.113	1.988	.199	.040	6.121	.000
	Supportive relations	−.169	.051	−.190	−3.352				

(1. Self-esteem, 2. Vigilance, 3. Complacency, 4. Panic, 5. Cop-out)

The results of the performed independent group *t*-test (see Table 4) showed the difference between male and female students' arithmetic means for their scores on self-esteem in decision-making to be statistically significant ($t = -1,390$; $p < .000$). According to this, female students exhibited greater self-esteem behaviour in decision-making than male students. However, students showed no significant change in their decision-making behaviours of vigilance, complacency, panic or cop-out according to gender ($p > .050$).

Table 4

Independent Group t-*test Related to Difference of Self-Esteem in Decision-making According to Gender*

Characteristic	Gender	*n*	$\bar{\chi}$	SS	$SH_{\bar{\chi}}$	t Test		
						SD	t	p
Self-Esteem in Decision-making	Male	157	16.85	2.58	.206	298	−1.390	.000
	Female	143	17.33	3.40	.284			

Discussion, Conclusion and Recommendations

According to the results of the present study, the characteristics of vigilance, complacency, and self-esteem in the processes of decision-making stand out in adolescents who show the quality of narcissism. The most significant feature of narcissism concerns how the individual's features are exaggeratedly brought forward, and their interest in others is reduced (Timuroglu & Iscan, 2008). This finding on the qualities of narcissism overlaps with those from previous studies. Because narcissistic individuals are more concerned with their own views and inclinations than the views and thoughts of those in the outside world, having high levels of inclination towards self-esteem can be interpreted as an expected situation. This is because self-esteem in decision-making reflects high confidence they feel towards themselves in decision-making situations (Janis & Mann, 1979). The behaviours of vigilance and complacency that narcissistic individuals exhibit are compatible with the findings in the literature. According to Masterson (2006), narcissistic individuals act while thinking about their future because they place importance on their own future and do not think about others as much as themselves. In this context, the tendency to be complacent and more careful in decision-making becomes more important to these individuals. A similar finding appears in Kocakula's (2012) study, which identified a positive relationship between the sub-dimensions of decision-making and narcissism.

The concept of narcissism has also been considered separately as non-threatening narcissism and pathological narcissism. According to Rozenblatt (2002), non-threatening (normal) narcissism is the experience of feeling one's self in harmony with one's near surroundings and those in it, and feeling able to meet its expectations. Normal narcissistic individuals are those over-interested in their own successes who always want more (Fromm, 1994). People defined as normal narcissists can in fact be treated as individuals with quite high self-esteem and self-confidence. Tazegul (2013) revealed a positive relationship between self-respect and narcissism. From this perspective, the findings of the present study appear to overlap with those from other

studies that have indicated the positive effect of self-respect on the behaviour of decision-making (Burnett, 1991; Burnett, Mann, & Beswick 1989; Mann et al., 1989; Phillips et al., 1984).

Students' internally supportive family relations were found as significantly predicting decision-making characteristics of self-esteem and vigilance in a positive direction, and of complacency, panic and cop-out in a negative direction. In parallel, students' internally discouraging family relations were found as predicting the characteristic of self-esteem in decision-making in a negative direction, and of complacency, panic and cop-out in a positive direction. These findings clearly show that, when parents raise their children supportively, with a healthy attitude that acquaints them with freedom, children's decision-making behaviours are expressed in a more desirable way. The findings of the current study coincide with those of other studies showing the positive effects of being raised in an environment where the family is healthy and supportive of children's decision-making behaviours (Brown & Mann, 1991; Dulger, 2009; Eldeleklioglu, 1996; Gucray, 1998; Mann et al., 1989; Wilks, 1986). Schvaneveldt and Adams (1983) determined that the family has a significant impact on the autonomy of an adolescent individual's decision-making. However, Tatlilioglu's (2014) study found the level of self-esteem in decision-making within all demeanours of the family to be highest in those raised in democratic families. All these findings indicate the importance of the family's role in adolescents' decision-making behaviour.

When investigating the decision-making levels of all the sub-dimensions according to gender, female students were found as exhibiting to a greater extent the behaviour of self-esteem in decision-making than male students. In parallel with the findings of this study, some previous studies that examined gender and decision-making (Gucray, 1998; Izgar, 2003; Koksal, 2003) showed girls as being more likely than boys to demonstrate healthy decision-making behaviours. This finding is thought to be related to the developmental period in which girls and boys of this age find themselves. In this period, girls undergo changes related to puberty before boys, and in one respect gain adult-like qualities before boys (Berkem-Guvenc, 1996; Kulaksizoglu, 1998; Ozbay & Ozturk, 1992; Yorukoglu, 1993). This situation establishes the significance of their self-esteem appearing higher than boys and, therefore, their self-esteem in decision-making. Other studies on this subject (Gucray, 1995; Gulbahce & Kartol, 2014; Mau, 2000) indicated contrarily that boys' levels of healthy behaviour in decision-making are further ahead of girls'. The results of other studies (Avsaroglu & Ure, 2007; Leaper, 1998; Tiryaki, 1997) showed no significant difference in decision-making behaviour according to the variable of gender. In fact, these findings suggest that gender may not be the only determinant variable in decision-making behaviours and that other variables may impact decision-making behaviours when introduced to the session.

The study has clearly identified that family attitudes towards child-rearing are one of the most important variables affecting children's personality features and behaviours. These findings also confirm this known fact. In this context, the importance of families being more educated and informed on this process is clear. However, children's qualities of narcissism appear as an effective variable on their decision-making behaviours. Because the literature, again, often emphasises the

importance of family in narcissistic qualities being acquired, it is also important to conduct future studies with families.

This study was conducted with middle school students in Istanbul. The social environment in which an individual lives is an effective variable on one's behaviours. Therefore, further studies conducted with adolescents living in different regions would lead to better understanding of this subject. Additionally, it is important to support the findings of this study, which was conducted as quantitative research, with qualitative studies that can reveal the causes of this issue in a greater depth.

References

Adams, J. F. (1995). *Ergenligi anlamak.* [Understanding adolescence]. (Translate: B. Onur & et. all). Ankara: Imge Yayinevi.

Akin, A., Sahin, M., & Gulsen, M. (2015). Cocukluk cagi narsisizm olcegi: Gecerlik ve guvenirlik calismasi [Childhood narcissism scale: A study of validity and reliability]. *Dicle Universitesi Ziya Gokalp Egitim Fakultesi Dergisi, 24*, 203-215.

Arnett, J. J. (2000). Emerging adulthood: a theory of development from the late teens through the twenties. *American psychologist, 55*, 469-480.

Arnett, J. J. (2004). *Emerging adulthood: The winding road from the late teens through the twenties*. New york: Oxford university.

Atkinson, A. L., Atkinson, R. C., Smith, E. E., Bem, D. J., & Nolen-Hoeksema, S. (1999). *Psikolojiye giris.* [Introduction to psychology]. (Translate: Y. Alagon). Ankara: Arkadas Yayinevi.

Atay, S. (2010). *Calısan narsisist.* [Working narsisist]. Istanbul: Namar yayinevi.

Avsaroglu, S. (2007). *Universite ogrencilerinin karar vermede ozsaygi, karar verme ve stresle basa cıkma stillerinin benlik saygisi ve bazı degiskenler acisindan incelenmesi.* [The study of styles of coping with stress, decision-making and self-esteem of university students on decision-making in terms of self-esteem and some variables]. (Yayinlanmamis Doktora Tezi). Selcuk Universitesi, Konya.

Berkem-Guvenc, G. (1996). Erkek ve kiz universite ogrencilerinin aile ici etkilesime iliskin algilari ile toplumsal cinsiyet rolune iliskin tutumlari arasindaki iliski. [The relationship between male and female college students' perceptions of family interaction and gender role attitudes]. *3p dergisi, 4* (1), 34-40.

Brown, J. E., & Mann L. (1991). Decision-making competence and self-esteem: A comparision of parents and adolescent. *Journal of Adolescence, 14,* 363- 371.

Brown, J. E., & Mann, L. (1990). The relationship between family structure and process variables and adolescent decision making. *Journal of Adolescence, 13*, 25-37.

Burnett, P. C. (1991). Decision-making style and self-concept. *Australian Psychologist, 26,* 55-58.

Burnett, P. C., Mann, L., & Beswick, G. (1989). Validation of the flinders' decision making questionnaire on course decision making by students. *Australian Psychologist, 24,* 285-292.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits. Hillsdale*. New jersey: Lawrence Erlbaum Associates.

Byrnes, J. P. (1998). *The nature and development of decision-making: A self regulation model.* USA: Earlbaum Manwah.

Charles, G. M. (2002). *Psikolojiyi anlamak (psikolojiye giris).* [Understanding psychology (Introduction to psychology). (Translate: H. B. Ayvasik, M. Ayil). Ankara: Turk Psikoloji Birligi Yayinlari no: 23.

Colakkadioglu, O. (2012). Ergenlerde karar verme olcegi'nin ortaogretim ogrencileri icin gecerlik ve guvenirlik calismasi. [The reliability and validity study of adolescent decision making questionnaire for the high school students]. *Mustafa Kemal Universitesi Sosyal Bilimler Enstitusu Dergisi , 9* (19), 387-403.

Cuhadaroglu, F. (2004). *Ergen ve ruhsal sorunlar. Durum saptama calismasi.* [Adolescence and mental problems. Situation detection work]. Ankara: Turkiye Bilimler Akademisi Raporlari.

Dacey, J. S., & Kenny, M. (1994). *Adolescent development.* USA: Brown & Benchmark Publishers.

Demirtas-Sorbaz, F., & Korkut-Owen, F. (2013). Cocuklar icin aile iliskileri olceginin gelistirilmesi. [Developing family relationship scale for children]. *Turk Psikolojik Danısma ve rehberlik Dergisi, 4*(39), 58-67.

Deniz, M. E. (2006). The relationships among coping with stress, life satisfaction, decision-making styles and decision self-esteem: an investigation with turkish university students. *Social Behaviour and Personality, 34,* 1161-1170.

Dulger, O. (2009). *Ergenlerde algilanan sosyal destek ile karar verme davranislari arasindaki iliski.* [Correlation between perceived social support and decision making behavior of adolescents]. Yuksek Lisans Tezi, Marmara Universitesi, Istanbul.

Eldeleklioglu, J. (1996). *Karar stratejileri ile ana-baba tutumlari arasindaki iliski.* [Correlation between parental attitudes and decision making strategies]. Yayimlanmamis Doktora Tezi. Gazi Universitesi, Ankara.

Ersever, O. H. (1996). *Karar stratejileri ile anne baba tutumlari arasindaki iliski.* [Correlation between decision making strategies and parental attitudes]. (Yayimlanmamis Doktora Tezi). Gazi Universitesi, Ankara.

Eksi, A. (1990). *Cocuk, genc, ana babalar.* [Child, young, parents].Ankara: Bilgi Yayinevi

Freud, S. (2007). *Narsizm uzerine ve Schreber vakasi,* [On narcissism and Schreber case]. (Translate: Buyukkal, & Tura). Istanbul: Metis Yayinlari.

Fromm, E. (1994). *Sevginin ve siddetin kaynagi* [Source of love and violence]. (Translate: Salman & Icten). Istanbul: Payel Yayinlari.

Fromm, E. (1999). *Ozgurlukten kacis.* [Escape from freedom]. (Translate: Budak) Ankara: Oteki Yayinlari.

Greasley, P. (2008). Quantitative data analysis using SPSS. An Introduction for health & Social science. Maidenhead: McGraw-Hill/Open University Press.

Gucray-S. S. (1995). Karar verme davranıslari olceginin gecerlik ve guvenirligi. [Reabiliyt and validity of making decision behaviors scale]. *Cukurova Universitesi Egitim Fakultesi Dergisi, 2*(4), 60- 68.

Gucray, S. S. (1998). Bazi kisisel degiskenler, algilanan sosyal destek ve atilganligin karar verme stilleri ile iliskisi. [The relationsip between some individual variables, percieved social support, assertivenes and making decision sytles]. *Psikolojik Danısma ve Rehberlik Dergisi, 11,* 7-17.

Gulbahce, A., & Kartol, A. (2014). Ilkogretim sekizinci sinif ogrencilerinin karar verme stratejilerinin bazı degiskenler acisindan incelenmesi. [The investigation of 8th grade student's decision making strategies in terms of some variables]. *Bayburt Universitesi Egitim Fakultesi Dergisi, 9*(2), 132-143.

Izgar, G. (2003). *Okul yoneticilerinin karar verme stratejileri ve yeterlikleri.* [Desicion strategies and adequacies of school administrators]. (Yuksek Lisans Tezi). Selcuk Universitesi, Konya.

Janis, I., & Mann, L. (1979). *Decision making. A psychological analysis of conflict, choice, and commitment.* New York: The free press, a division of macmillan inc.

Karasar, N. (1999). *Bilimsel arastırma yontemleri.* [Scientific research methods]. Ankara: Nobel Yayinevi.

Kasik, D. Z. (2009). *Ergenlerde karar verme stilleri ve algilanan sosyal destek duzeylerinin sosyal etkinlik beklentisi ve bazi degiskenler acisindan incelenmesi.* [Adolesent self-esteem and decision-making style of the decision with the perceived level of social support, social competence and level of expectation in terms of some variable treated as a comparative analysis]. (Yuksek Lisans Tezi). Selcuk Universitesi, Konya.

Kernberg, O. F. (2012). *Sinir durumlar ve patolojik narsisizm* [Boundary conditions and pathological narcissism]. (Translate: Atakay). Istanbul: Metis Yayinlari.

Kilicci, Y. (2000). *Okulda ruh sagligi.* [Mental health in school]. Ankara: Ani Yayincilik.

Kocakula, O. (2012). *Narisistik ve obsesif kompusuf kisilik bozukluklarının karar sureclerine etkisi.* [The effects of narcissistic and obsessive compulsive personality disorders on decision making process]. (Yuksek Lisans Tezi). Adnan Menderes Universitesi, Aydın.

Kohut, H. (2004). *Kendiligin cozumlenmesi.* [Solution of self]. (Translate: Atbasoglu, Buyukkal & Iscan) Istanbul: Metis Yayinlari.

Koksal, A. (2003). *Ergenlerde duygusal zeka ile karar verme stratejileri arasindaki iliski.* [The Relationship between emotional intelligence and decision making strategies in Adolescence]. (Yuksek Lisans Tezi). İstanbul Universitesi, Istanbul.

Kulaksizoglu, A. (1998). *Ergenlik psikolojisi.* [Adolescence psychology]. Istanbul: Remzi Kitapevi.

Kurt, U. (2003). *Karar verme surecinde yoneticilerin kisilik yapilarinin etkileri.* [The effects of the managers' personalities to the decision making process]. (Yuksek Lisans Tezi). Baskent Universitesi, Ankara.

Ladd, G. W., & Pettit, S. (2002). *Parenting and the development of children's peer relationships.* (In Bornstein, M. H. (ed.) *Handbook of Parenting*: Vol. 5: *Practical Issues in Parenting*, Erlbaum, Mahwah, NJ, pp. 269–309.

Leaper, C. (1998). Decision-making processes between friends: Speaker and partner gender effects. S*ex Roles, 39*(112), 125-133.

Mann, L. (1989). Becoming a better decision maker. *Australian psychologist, 24*(2), 141-155.

Mann, L., Harmoni, R., & Power, P. (1989). Adolescent decision-making: The development of competence. *Journal of Adolescence, 12*(3), 265-278.

Masterson, J. F. (2006). *Narsisistik ve borderline kisilik bozuklukları.* [Narcissistic and borderline personality disorders]. (Translate: Acil) Istanbul: Litera Yayincilik.

Mau, W. C. (2000). Cultural differences in career in career decision-making styles and self-efficacy. *Journal of Vocational Behavior. 57,* 365-378.

Miller, D. C. (2000). *Adolescents' decision making in academic and social contexts: A self-regulation perspective.* (Doctoral Dissertation). University of Maryland, College park.

Morgan, T., C. (1991). *Psikolojiye giris.* [Introduction to psychology]. (Translate: Arıcı). Ankara: Hacaettepe Universitesi Psikoloji Bolumu Yayinlari.

Naftel, M. I., & Driscoll, M. (1993). Problem solving and decision making in eighth-grade class. *Clearing house, 66*(3), 177-181.

Noone, J. (2002). Concept analysis of decision making. *Nursing forum, 3* (3), 21-32.

Ozolins, A. R., & Stenstrom, U. (2003). Validation of health locus of control patterns in Swedish adolescent. *Adolescence, 38*(152), 650-658.

Ozaydin, S. (1984). *Psikiyatri.* [Psychiatry]. Istanbul: Sanal Yayinlari.

Ozbay, H., & Ozturk, E. (1992). *Genclik.* [Youth]. Istanbul: Iletisim Yayinlari.

Oztemel, K. (2012). Kariyer kararsizligi ile mesleki karar verme oz yetkinlik ve kontrol odagi arasındaki iliskiler. [Relationships between career ındecision career decision making self efficacy and locus of control]. *Gazi Universitesi Egitim Fakultesi Dergisi, 32*(2), 459-472.

Phillips, S. D., Pazienza, N. Y., & Ferrin, H. H. (1984). Decision making styles and problem solving appraisal. *Journal of Counseling Psychology, 31*(4), 497-502.

Rollinson, D. (2002): *Organizational behavior and analysis an integrated approach.* Essex: Pearson Education ltd.

Rozenblatt, S. (2002). *In defence of self: The relationship of self-esteem and narcissim to aggressive behavior. (*Doctoral Dissertation). Long Island University, USA.

Satir, V. (2001). *İnsan yaratmak: Aile terapisinin basyapıtı*. [Creating people: The masterpiece of family therapy]. İstanbul: Beyaz Yayinlari.

Schvaneveldt, J. D., & Adams, G. R. (2001). Adolescents and the decision-making process. *Theory into Practice, 22*(2), 98-104.

Tatlilioglu, K. (2014). Universite ogrencilerinin karar vermede oz-saygı duzeyleri ile karar verme stilleri arasindaki iliskinin bazi degiskenlere gore incelenmesi. [The examination of relation between undergraduates' self-esteem level and decision making style in deciding in terms of some variants]. *The Journal of Academic Social Science, 2*(1), 150-170.

Tazegul, U. (2013). Bayan Badmintoncularin benlik saygi ve narsisizm duzeyleri arasindaki iliskinin incelenmesi. [Investigating the relationship between self-esteem and narcissism level of female badminton players]. *The Journal of Academic Social Science Studies, 6*(8), 983-996.

Timuroglu, K., & Iscan, O. F. (2010). *Isyerinde narsisizm ve is tatmini iliskisi*. [The relationship between narcissism in workplace and job satisfaction]. Ataturk Universitesi Yayinlari.

Tiryaki, M. G. (1997). *Universite ogrencilerinin karar verme davranıslarının bazı degiskenler acisindan incelenmesi*. [Decision-making strategies of university students according to some variables]. (Yuksek Lisans Tezi). Hacettepe Universitesi, Ankara.

Twenge, J. M. (2010). *Asrın vebası: Narsisizm illeti*. [Century Plaggue: Narcissim ill]. (Translate: Merter). Istanbul: Kaknus Yayinlari.

Yavuzer, H. (1993). *Cocuk psikolojisi.* [Child psychology]. Istanbul: Remzi

Wilks, J. (1986). The relative importance of parents and friends in adolescent decision making. *Journal of Youth and Adolescence,15*(4), 323-334.

Yorukoglu, A. (1993). *Genclik cagi: Ruh sagligi ve sorunlar*. [Youth age: Mental health and problems]. Istanbul: Ozgur Yayinlari.

Ortaokul Öğrencilerinde Narsisizm ve Aile İlişkilerinin Karar Verme Özellikleri Üzerindeki Yordayıcı Rolü

Özet

Problem Durumu: Ergenliğe giriş sürecinin yaşandığı ortaokul yıllarında birey fiziksel, duygusal, toplumsal vb. birçok değişim ve gelişim evresinden geçer ve çocukluk döneminde başlayan kişilik yapılanması bu evrede belirginleşmeye başlar. Ergenlik dönemi, genç bireyin yaşamının ileriki zamanlarında belirleyici olacak bir takım kararların alındığı da bir yaşam evresidir. Araştırmacılar ergenlik döneminde beklenen gelişim görevlerini sıralarken üst sıralara, kişilik önceliklerini şekillendirme, duygusal bağımsızlığını kazanma, önemli kararlarını kendi başına verme, kendi yaşına özgü bir yaşam felsefesi geliştirme ve geleceği konusunda kendi isteklerine odaklanma gibi özellikleri koymaktadırlar. Karar alma basit bir durum olmayıp birçok değişkenin devreye girdiği karmaşık bir olgu olarak ele alınmaktadır. Her gelişim döneminde olmakla beraber ergenlik döneminde de ailenin ergen üzerindeki etkisi oldukça fazladır. Bir yandan aileler diğer yandan öğretmenler bu çağ çocuğundan bazı eğitsel, kişisel ve mesleki kararlar almasını bekler. Çocuğun bu kararlarını alırken sergilediği davranışlarda aile ilişkilerinin ve yine temelleri ailede atılan narsisistik özelliklerin etkili olduğu düşülmektedir. Aileler çoğunlukla farkında olmadan çocuklarının kişilik gelişimlerinde birçok olumsuzluğa neden olabilmektedirler. Aslında doğru rehberlik çalışmalarıyla aileler bilinçlendirildiğinde çocuklarını daha sağlıklı bir şekilde yetiştirmeleri mümkündür. Bu noktada, etkili okul rehberlik ve psikolojik danışma hizmetlerine büyük görev düşmektedir.

Araştırmanın Amacı: Bu araştırmanın amacı, ortaokul öğrencilerinde narsisizm ve aile ilişkilerinin karar verme özellikleri üzerindeki etkisini incelemektir. Ayrıca öğrencilerin karar verme özelliklerinin cinsiyet değişkenine göre farklılaşıp farklılaşmadığı da araştırmada incelenmiştir.

Araştırmanın Yöntemi: Araştırmanın örneklemini İstanbul'daki 5 farklı ortaokulun yedinci ve sekizinci sınıflarında öğrenim gören, 157'si (%52.3) erkek, 143'ü ise (%47.7) kız öğrencilerden oluşan ve yaş ortalaması=13.20 olan toplam 300 kişi oluşturmaktadır. Bu öğrencilerin 158'i (%52.6) yedinci sınıfta, 142'si ise (%47.4) sekizinci sınıfta öğrenim görmektedir. Araştırmada örneklem belirleme yöntemi olarak olasılığa dayalı örneklem türlerinden biri olan tabakalı örnekleme kullanılmıştır. Ortaokullarda yedinci ve sekizinci sınıftaki öğrencilerin lise kararı alma gibi daha fazla karar yaşantısı söz konusu olduğu için beşinci ve altıncı sınıflar çalışma grubuna dahil edilmemiştir. İlişkisel tarama modeline göre yapılandırılan araştırmanın verileri; *Kişisel bilgi formu*, karar verme özelliklerini belirlemek amacıyla *Ergenlerde karar verme ölçeği*, narsisizm özelliklerini ortaya koymak için *Çocuklar için narsisizm ölçeği* ve aile ilişkilerini belirlemek amacıyla *Çocuklar için aile ilişkileri ölçeği*

kullanılarak toplanmıştır. Verilerin istatistiksel analizlere uygunluğunu belirlemek için bir dizi hesaplamalar yapıldıktan sonra amaca göre parametrik ve non-parametrik analizler yapılmıştır. Bu noktada öncelikle değişkenler arasındaki ilişkileri görebilmek amacıyla Pearson Korelasyon Analizi kullanılmıştır. Ardından öğrencilerin narsisizm özellikleri ile aile içi destekleyici ve aile içi engelleyici ilişkilerinin karar verme özelliklerinin alt boyutlarını yordayıp yordamadığını görebilmek için Çoklu Regresyon Analizi kullanılmıştır. Burada aile içi içi destekleyici ve engelleyici ilişkiler birbirinin tersi olan iki ayrı değişken olduğundan narsisizm ile birlikte ayrı ayrı analize tabi tutulmuşlardır. Ayrıca karar verme ölçeğinden, ölçeğin yapısı gereği tek bir toplam puan alınamadığı için karar vermenin alt boyutları ayrı ayrı bağımlı değişken olarak ele alınmış ve analizler buna göre yapılmıştır. Son olarak karar verme özelliklerinin öğrencilerin cinsiyetlerine göre farklılaşıp farklılaşmadığını belirlemek için Bağımsız Grup t Testi kullanılmıştır.

Bulgular: Yapılan istatistiksel analizlere göre; öğrencilerin karar vermede öz saygı özelliklerinde hem narsistik özelliklerinin hem de ailelerindeki engelleyici ilişkilerin anlamlı düzeyde yordayıcı etkisinin olduğu belirlenmiştir. Yapılan analizler karar vermenin alt boyutlarından biri olan öğrencilerin ihtiyatlı seçicilik özelliklerinde, narsistik özelliklerinin anlamlı bir yordayıcı olduğunu ancak ailelerindeki engelleyici ilişkilerin anlamlı bir yordayıcı olmadığını göstermiştir. Yapılan analizler karar vermenin bir diğer alt boyutu olan umursamazlık özelliğinde, öğrencilerin hem narsistik özelliklerinin hem de ailelerindeki engelleyici ilişkilerin anlamlı yordayıcılar olduğunu göstermektedir. Diğer yandan analizlere bakıldığında, karar vermede panik özelliğinde öğrencilerin narsistik özellikleri anlamlı bir yordayıcı olarak bulunmazken ailelerindeki engelleyici ilişkilerin anlamlı bir yordayıcı olduğu belirlenmiştir. Son olarak karar vermede sorumluluktan kaçma özelliğinde öğrencilerin narsistik özelliklerinin anlamlı bir yordayıcı olmadığı ancak ailelerindeki engelleyici ilişkilerin anlamlı bir yordayıcı olduğu görülmüştür. Diğer yandan öğrencilerin karar vermede öz saygı özelliklerinde hem narsistik özelliklerinin hem de ailelerindeki destekleyici ilişkilerin anlamlı düzeyde yordayıcı etkisinin olduğu belirlenmiştir. Yapılan analizler karar vermenin alt boyutlarından biri olan öğrencilerin ihtiyatlı seçicilik özelliklerinde, hem narsistik özelliklerinin hem de ailelerindeki destekleyici ilişkilerin anlamlı yordayıcılar olduğunu göstermiştir. Yapılan analizler karar vermenin bir diğer alt boyutu olan umursamazlık özelliğinde, öğrencilerin hem narsistik özelliklerinin hem de ailelerindeki destekleyici ilişkilerin anlamlı yordayıcılar olduğunu göstermektedir. Yine analizlere bakıldığında, karar vermede panik özelliğinde öğrencilerin hem narsistik özelliklerinin hem de ailelerindeki destekleyici ilişkilerin anlamlı yordayıcılar olduğu belirlenmiştir. Son olarak karar vermede sorumluluktan kaçma özelliğinde öğrencilerin narsistik özelliklerinin anlamlı bir yordayıcı olmadığı ancak ailelerindeki destekleyici ilişkilerin anlamlı bir yordayıcı olduğu görülmüştür. Cinsiyet değişkenine dair bulgulara bakıldığında ise, kızların karar vermede öz saygı puanlarının erkeklerinkinden anlamlı düzeyde daha yüksek olduğu belirlenmiştir.

Sonuç ve Öneriler: Araştırma bulguları, öğrencilerin hem narsisistik özelliklerinin hem de aile ilişkilerinin karar verme davranışlarında etkili değişkenler olduğunu göstermiştir. Ayrıca öğrencilerin cinsiyetlerine göre karar verme davranışlarında farklılaşmalar olduğu belirlenmiştir. Açıkça bilinmektedir ki, ailenin çocuk

yetiştirmedeki tutumları çocukların kişilik özelliklerini ve davranışlarını etkileyen en önemli değişkenlerden biridir. Bu araştırma bulguları da bilinen bu gerçeği doğrular niteliktedir. Bu çerçevede, ailelerin bu süreç hakkında daha fazla eğitilmesi ve bilgilendirilmelerinin önem arz ettiği açıktır. Diğer yandan, çocukların karar verme davranışlarında narsistik özelliklerinin de etkili bir değişken olduğu görülmüştür. Alan yazını yine narsistik özelliklerin kazanılmasında ailenin önemine sıklıkla vurgu yaptığı için, ailelerle yürütülecek olan çalışmaların önemi tekrar karşımıza çıkmaktadır. Diğer araştırmalar için; bu araştırma İstanbul'daki ortaokul öğrencileri ile yürütülmüştür. Bireyin içinde yaşadığı sosyal çevre onun davranışları üzerinde etkili bir değişkendir. Bu yüzden farklı bölgelerde yaşayan ergenler ile yürütülecek diğer çalışmalar konunun daha anlaşılır olmasına yardımcı olacaktır. Ayrıca nicel araştırma yöntemi ile yürütülen bu araştırmanın bulgularını, konunun nedenlerini daha derin bir şekilde ortaya koyabilecek nitel araştırmalarla desteklemenin önemli olduğu düşünülmektedir.

Anahtar Kelimeler: Karar verme, narsisizm, aile ilişkileri, ergenlik, kişilik.

Assessment with and for Migration Background Students-Cases from Europe*

Funda NAYIR[1], Martin BROWN[2], Denise BURNS[3], Joe O'HARA[4], Gerry MCNAMARA[5], Guri NORTVEDT[6], Guri SKEDSMO[7], Silje Kristin GLOPPEN[8], Eline F. WIESE[9]

ARTICLE INFO	ABSTRACT
	Purpose: The purpose of this research as part of an Erasmus+ funded project tilted Aiding Culturally Responsive Assessment in Schools (www.acras.eu) is to provide an exploratory analysis of survey responses related to culturally responsive assessment policies, professional development and practices that were administered to school principals in four European countries (Austria, Ireland, Norway and Turkey). **Research Methods:** The research method used in this study was a quantitative comparative analysis. A purposeful sampling strategy was adopted based on geographical spread of the participants in Austria, Ireland and Turkey. The survey was also administered to all principals in four out of eighteen Norwegian counties. Descriptive analysis and nonparametric analysis were used.

Findings: There is evidence to suggest that the foundations for culturally responsive assessment practices are beginning to take shape albeit varying degrees of difference in each country. However, the survey results also indicate the need for training and professional development, and this study also implies that not enough emphasis is being placed on culturally responsive assessment despite the rhetoric that espouses interculturalism.

Implications for Research and Practice: The research points the need for upskilling in culturally responsive leadership as well as the development of an overarching culturally responsive assessment framework and toolkit that can be used by policy makers and schools to allay the various interpretations of what it means to satisfy the assessment needs of teachers and students with migrant backgrounds.

---

[1] Çankırı Karatekin University, TURKEY, fnayir@yahoo.com, orcid.org/0000-0002-9313-4942

[2] EQI, DCU Institute of Education, IRELAND, martin.brown@dcu.ie, orcid.org/0000-0002-5436-354X

[3] EQI, DCU Institute of Education, IRELAND, denise.burns@dcu.ie, orcid.org/0000-0002-4009-7666

[4] EQI, DCU Institute of Education, IRELAND, joe.ohara@dcu.ie, orcid.org/0000-0003-1956-7640

[5] EQI, DCU Institute of Education, IRELAND, gerry.mcnamara@dcu.ie, orcid.org/0000-0002-9725-930

[6] University of Oslo, NORWAY, guri.nortvedt@ils.uio.no, orcid.org/0000-0002-4699-9089

[7] University of Oslo, NORWAY, guri.skedsmo@ils.uio.no, orcid.org/0000-0001-7151-8630

[8] University of Oslo, NORWAY, s.k.gloppen@ils.uio.no, orcid.org/0000-0002-5090-5866

[9] University of Oslo, NORWAY, e.f.wiese@ils.uio.no, orcid.org/0000-0002-1655-7642

Introduction

Countries throughout Europe have witnessed significant changes in patterns of migration. According to the United Nations Department of Economic and Social Affairs (UN DESA), the number of international migrants in the world reached 258 million in 2017; with a 49% increase in migration since 2000 (2017a). Of this total migrant stock, 14 percent are under 20 years of age (2017b). However, given that these figures relate to first-generation migrants (a person born in a country other than the country of residence), the net global migrant population is considerably higher if other migrant categories such as second-generation migrants (native-born with at least one foreign-born parent) and native-born persons with migration backgrounds are considered (Eurostat, 2017). Whereas global patterns of migration have changed where for example, many countries in Europe such as Ireland have become migrant receiving countries, what has remained constant over the last eighty years are the rates of and motives for migration such as increased safety and better living conditions (Borjas, 1995; Castles, Haas & Miller, 2013; Cavalli Sforza & Pievani, 2012; Czaika & Haas, 2014; Geddes & Scholten, 2016).The major differences between migration in this era and the past is that 'human beings are engaged in much faster, more distant, and more frequent migrations than ever before' (Chung & Griffiths, 2018, p.64). Certainly, these fluctuating patterns of migration have resulted in heightened awareness among various national and transnational policy makers such as the Organisation for Economic Co-operation and Development (OECD) on how best to sustain and meet the needs of culturally diverse populations; particularly when vital elements for social cohesion such as education are concerned. As stated by OECD (2016a), 'Integration policies, and extra support targeted towards immigrant families and children, can make a significant difference in how immigrant students fare in their host communities' (p.16). Thus far; however, with numerous descriptions of cultural neglect such as the inattention given to linguistic interdependence (Cummins, 1979), the lack of culturally responsive leadership in schools (Brown, McNamara, O'Hara, Hood, Burns & Kurum, 2017), and the supremacy of standardized testing (Brown, McNamara & O'Hara, 2016; Padilla, 2001; Young, McNamara, Brown & O'Hara, 2018,); educational outcomes for migrant children have not always been as uniformly positive compared to that of their native counterparts (Brown, 2007; Griner & Stewart, 2013). The OECD Programme for International Student Assessment (PISA) starkly illustrates this point.

Acknowledging that there are varying conceptions of quality and indicators for educational achievement such as access to education, participation, progression, and youth unemployment (Brown et al. 2016; Caspersen, Smeby & Olaf Aamodt, 2017; Scheerens, Luyten & van Ravens, 2011); analysis of PISA test scores (OECD, 2000 – 2017) demonstrates that, in most OECD countries, there have been constant achievement gaps between migrant and non-migrant children in the areas of Reading, Mathematics and Science. Bilgili, Volante & Klinger (2018), in reference to PISA 2015 (OECD, 2016a), also state that the majority of first and second-generation migrant students performed worse than those students without a migrant background. On the

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

41

other hand, with reasons for optimism, there is a growing body of research that provides strategies for schools to become, what interchangeably referred to as, 'culturally relevant' (Ford & Kea, 2009; Ladson-Billings, 1995) and 'culturally responsive' (Gay, 2010); Villegas & Lucas, 2002) learning environments. As stated by Bledsoe and Donaldson (2015): 'the call for cultural responsiveness has reached a deafening crescendo' (p. 7).

Ford and Kea (2009) define culturally responsive classrooms as 'student centered and, by design and default, culture centered. A student-centered classroom does not exist if culture is ignored or disregarded in any way' (p.6). The authors identify five components of culturally responsive classrooms that need to be considered for equitable participation with and for students with a migration background: 'Curriculum' (Banks, 2006; 2008); 'Philosophy' (Ford & Kea, 2009); 'Instruction' (Gay, 2010); 'the Learning Environment' (Rothstein-Fisch & Trumbull, 2008); and 'Assessment' (Hood, 1998). This paper focuses on one core component of cultural responsiveness, often referred to in the literature as culturally responsive assessment (Slee, 2010; Smith-Maddox, 1998).

Researchers argue that adapting and being sensitive to people from different cultures is a critical 21st century skill that every student needs to survive and succeed in the world (Earley & Mosakowski 2004; Harris 2006; Tan, 2004). In education, Brown (2013) among others suggests that it is necessary for schools and external agencies to consider reciprocal methods of assessment and evaluation that take cognizance of cultural variations that exist. Be that as it may, most practicing teachers and those external agencies responsible for curriculum development come from a culture of the majority and have been trained in the majority culture (Gay, 2010); often making implicit assumptions about how to conduct assessment regardless of the cultural variations that exist in schools (Rothstein-Fisch & Trumbull, 2008). Culturally responsive teachers, on the other hand, are aware of cultural and social diversities and embed culturally sensitive approaches in their practices (Ford & Moore, 2013). In other words, being culturally responsive means being respectful of, taking cognizance of, the social and cognitive cultural variations that exist. Culturally responsive assessment can, therefore, be described as assessment that utilizes strategies to acknowledge and respect learners' cultural backgrounds and approaches to learning as they strive for academic success. A review of the literature suggests that there are certain interconnected social, cognitive and affective dimensions that affect the quality and merit of assessment strategies that have the potential to be culturally fair.

1. *Researching the self*. For teachers involved in cultural responsiveness, the process is 'deliberate and self-exploring and requires new thinking and practices' (Hood, Hopson & Frierson, 2015, p.xv). Researching the self requires: knowing one's biases; adapting; assisting and leading (Hofstede, 2009).

2. *Multicultural validity*. Multicultural validity is defined as 'the accuracy and trustworthiness of understandings and actions across simple, intersecting dimensions

of cultural difference' (LaFrance, Kirkhart & Nichols, 2015, p. 57). Multicultural validity addresses five elements of qualitative and quantitative assessment:

a)   Theory (rationales support the inferences and actions based on assessment)

b)   Methodology (design and measurement)

c)   Relationships (among all forms of life including interactions among people)

d)   Experience (congruence with the lived experience of participants)

e)   Consequences (justice of outcomes)

3. *Construct validity*. For students who may be in the process of learning the language of the test instrument, construct validity may be a serious validity concern if the 'linguistic complexity unnecessarily interferes' with the ability to demonstrate their knowledge in the situation where language factors are 'unrelated to the measured construct' such as mathematics (DeBacker, Van Avermaet & Slembrouck, 2016).

4. *Language*. First language is the source of the student's identity, and depth of concepts in a second language occurs via the conceptual knowledge in the first language. Furthermore, 'Conceptual knowledge developed in one language helps to make input in the other language comprehensible' (Cummins, 2000). Also, the development of proficiency in communication skills takes about eighteen months to two years, and the acquisition of academic proficiency requires five to seven years (Hancock, 2017).

5. *Thinking styles*. There are two styles of thinking suggested by Bennett (2009, p.134) that teachers need to be aware of when considering how to embed culturally responsive assessment into practice; Style A (Analytically logical, Abstract, Objective, Dialectic and Doubting) and Style B (Holistic, Metaphorical, Subjective, Integrative, Believing).

These core dimensions for culturally responsive assessment have influenced the increased use of specific methods of assessment that are deemed to have the potential of being culturally fair. In the United States for example, culturally responsive methods of assessment such as peer and creativity assessment are increasingly being used with indigenous youth (Demmert, 2001; Nelson-Barber & Trumball, 2007) and other ethnic minority students (Aceves & Orosco, 2014; Qualls, 1998). However, in Europe, a review of the literature suggests that culturally responsive assessment practices are less prevalent and consequently less discussed. In fact, with rare exceptions (e.g., Mitakidou, Karagianni &Tressou, 2015) very few studies in Europe have examined assessment strategies that teachers use to integrate cultural responsiveness into their student assessments and to compare the relative merit of these strategies. Nor have they looked at the challenges to assessing students with a migration background. To fill the lacuna of research in this area, this paper as part of an Erasmus+ funded project titled Aiding Culturally Responsive Assessment in Schools (ACRAS.eu) provides an exploratory analysis of research that was conducted over a two-year period in four European countries in order to ascertain: (1) the

supports available to schools to enhance their assessment of migrant students; (2) the perceived competence of staff to assess migrant students; and (3) the challenges and methods of assessment used by teachers to assess culturally diverse students. The countries chosen for this study have complementary patterns of migration. The population of Ireland, a country with a strong tradition of emigration, is now made up of approximately 420,000 non-Irish people (Central Statistics office, 2017), a significant rise in non-nationals, given that in 2002, there were approximately 224,000 non-nationals resident in this country (Darmody & Smyth, 2018). Dissonant, albeit similar rates of increased migration have occurred in Turkey. In 2000, Turkey hosted approximately 3,000 refugees and in 2016, this number increased to 3.1 million (UN DESA,2017b). Similar figures related to the percentage of migrants resident in Austria (15.2 %) and Norway (10.6%) exist (Eurostat, 2017).

The first part of this paper provides a review of the literature on assessment methods that have the potential to be culturally fair together with a review of assessment policies and practices in four countries. Leading on from a description of the research method that was used in the study, a comparative analysis of assessment practices derived from a survey that was administered to secondary school principals in four countries is described. The paper concludes with a discussion of research findings derived from the preceding phases of the study, and suggests how these findings have wider implications for the future development of assessment policies and practices in Europe and elsewhere.

Methods of Assessment that Have the Potential to be Culturally Fair

It has been well documented that limiting the availability of certain methods of assessment such as standardized testing can have unintended consequences on student achievement (Caldas & Bankston, 1997; Stobart, 2008). This is of course not to say that there are no benefits to using standardized tests, but rather, to acknowledge antecedent variables that can affect the overall interpretation of assessment results. Certainly, there are benefits to standardized testing. In some OECD Countries for example, standardized tests are used as tools to ensure that schools and teachers provide a high-quality education to students (Morris, 2011). They can also be used to monitor national progress (Shewbridge, Jang, Matthews & Santiago, 2011) as well as providing a right of entry into tertiary education. Nonetheless, there are also concerns related to the validity of using test results for the assessment of students with a migration background (O'Connor,1989; Padilla, 2001). Concerns relate to, for example, the linguistic complexity of the test instrument that can negatively affect the overall assessment of students who are in the process of learning the test language (De Backer, Van Avermaet & Slembrouck, 2016). According to Menken (2010), 'testing research is conclusive that a content-area test administered to an ELL [English Language Learner] in English is unlikely to render a true portrait of what the student knows and is able to do because language impacts the results' (p.123). Other factors in parallel with the limitations of standardized tests for all students (e.g. the educational level of attainment of parents) can also affect assessment results. Yet, in addition to the acknowledged limitations of standardized tests; for students with a migration

background, one of the most significant desiderata for attainment in standardized tests relates to that of the family proximity to the language of the classroom (Randen, 2015; Stevenson & Willot, 2007). Indeed, Kim and Zabelina (2015) suggest that because standardized tests are acculturated to the knowledge, values and socially dominant language of instruction, if the same assessments are used, this can result in unfairness towards students with a migration background.

In addressing the issue of fairness, a mounting body of research suggests that fairness in assessment means providing a range of assessment strategies to allow for the assessment of students with a migration background (Castagno & Brayboy, 2008; Espinosa, 2005). These strategies include the use of multiple methods of assessment to provide additional opportunities for students to demonstrate their learning (Castagno & Brayboy, 2008; Qualls, 1998) as well as using locally validated formative assessments (Tichá & Abery, 2016). Culturally responsive methods of assessment include:

1. *Creativity assessment*. Creativity assessment is defined as 'producing something that is novel and useful' (Kim & Zabelina, 2015, p. 136). Hempel and Sue-Chan (2010) and Kim and Zabelina (2015) recommend that including creativity assessment can address cultural bias. Kim and Zabelina also state 'creativity assessment shows few differences across gender or ethnicity' (p. 136).

2. *Performance-based assessment*. The six characteristic features of performance-based assessment proposed by (Baker, O'Neil & Linn, 1993) are as follows: (a) uses open-ended tasks, (b) focuses on higher order or complex skills, (c) employs context-sensitive strategies, (d) often uses complex problems requiring several types of performance and significant student time, (e) consists of either individual or group performance, (f) may involve a significant degree of student choice (p. 1211)

There are, however, issues to consider with performance-based assessment such as: (a) validity, reliability issues, (b) difficulties in construction, (c) use of resources and time, and (d) the design and purpose of the assessment.

3. *Peer-assessment*. Peer assessment has the potential for assessment that is more learner-centric, flexible, and culturally responsive as these approaches help ethnic groups, including immigrants, to take ownership of their educational progress and assessment, and in a wider sense, involvement and inclusion in society. Products to be assessed can include writing, oral presentations, portfolios, test performance, or other skilled behaviors (Topping, 2009, p.20). However, Reynolds and Trehan (2000) among others caution that the movement towards a more learner-centric mode of assessment brings about its own challenges such as the shift in power relations between students and teachers. Simply to exchange one situation of power relations (tutor-student) with another (student-student) does not of itself guarantee equality. It raises a new set of complex power relations which need to be understood (Reynolds & Trehan, 2000, p.274)

4. *Self-assessment*. Self-Assessment is used as 'a formative, awareness building tool which is grounded in 'learning to learn' and student reflective practice' (Taras, 2001, p.606). As with peer assessment, there are many benefits of facilitating self-assessment

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE

45

Eurasian Journal of Educational Research 79 (2019) 39-68

practices such as increased confidence (Topping, 2009) and a better understanding of standards required (Hanrahan & Isaacs, 2001). However, challenges towards the implementation of self-assessment need to be considered. These challenges include lack of confidence by students to be objective in their assessments (Cheng & Warren, 1997) as well other challenges such as a lack of training required for the assessment task (Wood & Kurzel, 2008).

Nonetheless, etic perspectives on strategies of assessment that have the potential to be culturally fair can act as a powerful catalyst for classroom practice. It must also be noted that akin to all aspects of education; while schools and teachers have a responsibility for the implementation of culturally responsive assessment, they are also, as with any aspect of education, dependent on the vagaries of policies and supports that allow for the flourishing of such practices. To concur with Schapiro (2009), there is a need to ask if education policies and concomitant classroom practices do in fact 'improve the students' access to quality education, stimulate equitable participation in schooling, and lead to learning outcomes at a par with native peers' (p.33). It is this issue of assessment policies and practices in the case study countries that forms the next part of the paper.

Assessment Policies and Practice in four European Countries

In this section, summary descriptions of assessment policies and practices together with strategies to support the assessment of migrant students in each of the countries under investigation are described.

Assessment policies and practice in Austrian secondary schools: Entry to secondary education in Austria is based on the assessment of students at primary school level. The grades obtained at the end of primary school and at lower secondary school (8th year) are used to assign students to lower secondary school types and also to different types of upper secondary education or vocational training, respectively. Based on these assessments, students enter either a four-year Neue Mittelschule (NMS; i.e. lower 'practical' secondary school) or an eight-year Gymnasium (AHS; a traditional 'academic' secondary school). To complete the nine years of compulsory education, NMS-graduates can also attend a one-year Polytechnische Schule which prepares students for vocational training starting at grade 10. Whilst assessment in Austrian secondary schools is mainly teacher-led where the performance of students is continuously assessed throughout the school year using various instruments (e.g. tests, oral participation, homework, schoolwork, etc.). At the end of upper secondary level, students are assessed via a Leaving Certificate examination referred to as "Matura" which licenses for entry to tertiary education based on their performance in these examinations. Until recently, "Matura" examinations were set by individual schools. However, for increased transparency within the Austrian education system, from 2015 (2016 in upper vocational schools) a new scheme is in place which includes centrally set written examinations, regionally moderated oral examinations, and a 'research paper' written by individual students.

Extensive strategies and supports to enhance assessment with and for migrant students exist in the Austrian education system. At a policy level, intercultural learning is an 'educational principle' in Austrian schools. However, according to Fillitz (2002), many teachers do not consider it important or do not know how to put it into practice. The Ministry of Education have also set up The Federal Center for Interculturality, Migration and Multilinguality (BIMM, located at the University of Education Graz). This center provides relevant material and publications, organizes conferences and sustains networks among relevant actors of all Austrian teacher education institutions. The Austrian Center for Language Competences also develops materials and offers courses for teachers in the field of languages as well as strategies for whole-school development in the context of linguistic diversity. In the last number of years, the Federal Ministry has also begun the process of providing professional development workshops and courses for the purpose of having at least one competent teacher with responsibility for the further development of language-sensitive subject-teaching.

At a school level, students with first languages other than German are exempted from assessment for twelve months, and if necessary up to 24 months maximum. A language course or a language starting group is also available to students who have just started to learn German. For those students who are resident in the country for between 12 – 24 months, further courses in German as a second language are available for up to 5-6 hours per week. However, these supports in all cases are dependent on the organizational feasibility of the school management. Teaching in migrant student's first language is also offered as an optional subject with classes limited to a maximum of five students.

Assessment policies and practice in Irish Secondary schools. The first phase of secondary education in Ireland consists of three years of what is referred to as the junior cycle. This is followed by an optional transition year and two years of senior cycle. At both junior and senior cycle level, students are awarded national certificates which are equivalent to levels 2 and 3 of the European Qualifications Framework. Certificates are awarded based on national examinations that are set and administered by the State Examinations Commission. Assessment at junior and senior cycle level is undergoing reform. Derived from the OECD's definition and selection of key competencies, i.e. 'Use tools interactively (e.g. language, technology), Interact in heterogeneous groups, Act autonomously' (Rychen & Salganik 2003, p.5); the new junior cycle curriculum places a significant emphasis on key skills for the world of work (National Council for Curriculum and Assessment (NCCA), 2009). Within this assortment of competencies, the new junior cycle emphasizes the centrality of assessment for learning (Poole, Brown, McNamara, O'Hara, O'Brien & Burns, 2018), and from 2022 it will include two classroom-based assessments in all subjects which will be reported on separately from the results of the state examinations. Proposals for reform of the Leaving Certificate have also been suggested where a greater emphasis is also being placed on key skills and assessment for learning (Burns, Devitt, McNamara, O'Hara & Brown, 2018; NCCA, 2005). However, given the initial resistance to junior cycle reform in Ireland (Brown, McNamara & O'Hara, 2016), that is teachers devising and grading student

assessments, reform efforts at this level have been delayed for the purpose of system level acceptance of new junior cycle assessment arrangements. Finally, as a unique situation to Ireland, almost all students in Ireland sit what is referred to as a Mock Junior and Leaving Certificate examination in rehearsal and preparation for the actual certificate examinations that occur at the end of each cycle. After each cycle is complete, almost all secondary schools in Ireland evaluate, as a measure of quality, the collective attainment for each state examination subject in comparison to the national average subject score for all secondary schools (Brown, McNamara, O'Hara & O'Brien, 2017).

At a policy level, the aspiration to respect diversity has been enshrined in Irish educational legislation since 1998 (Government of Ireland, 1998). The NCCA has also published guidelines for schools on intercultural education. These guidelines acknowledge the ways that cultural or language factors can give rise to errors in assessment and contained within, and recommendations are made on how to minimise potential errors for assessment (NCCA, 2006). In 2010, the Department of Education and Skills also published a strategy for intercultural education (DES, 2010). Both of these publications were prompted by the sudden migrant flow in Ireland that commenced in the late 1990s. At a school level, the main targeted support mechanism for students with a migration background is the provision of extra support for students who are learning English as an additional language (EAL). More recently, allocation of EAL supports have been subsumed with learning supports (for students with special educational needs) (DES, 2012).

Assessment policies and practice in Norwegian schools. In Norway, compulsory education comprises primary (grades 1 – 7) and lower secondary education (grades 8 – 10). Compulsory education is inclusive, without streaming or tracking. Teaching should be adapted to the individual student (Lovdata, 2016). In total, 97% of the students attend public schools (NDET, 2017), mostly the neighborhood schools. Education from the primary to the post-secondary level are mainly funded (94%) at the local level (OECD, 2016b). Although schools are locally funded, and Norway has a decentralized education system, there is still a national curriculum and an education act regulating school practices. As such, national guidelines for assessment are included in the Education Act (Lovdata, 2016).

At a policy level, the Norwegian education system is based on assessment for learning that has now become a national policy. In addition, assessment of learning in the form of grading is part of the mandatory assessment practices at the secondary level. Grades range from 1-6, with 6 as the highest degree. The same grading system is used throughout secondary education (Lovdata, 2016). Students are mainly graded by their subject teacher with two exceptions; 1) national tests which are computer based with automated scoring and 2) exams which are scored and graded by independent external scorers. Students apply for further schooling (upper secondary level and tertiary level education) mainly based on teachers' grading, although the outcome of national exams are included on school leaving certificates and averaged with teacher grades. Lower secondary students take one written and one oral exam.

At a school level, assessment for learning is the main strategy to support assessment with migrant students. Specially adapted Norwegian instruction for students with a migration background in core areas of the curriculum such as literacy and numeracy are also provided. Students with special needs also receive special education or adapted education within the local school.

Assessment policies and practice in Turkish secondary schools. Whilst assessment in Turkey is teacher-led, national examinations exist in primary (middle) education, at grade 8. This examination determines admission into the type of secondary school and are administered by the Ministry of National Education (MoNE). According to the results obtained in these examinations, students are placed in science high schools, social science high schools, project schools, and vocational or technical Anatolian high schools. Schools other than the ones that take the students who have completed MoNE examinations, take students without examination (MoNE, 2018). At secondary level, the MoNE also assesses students via two national examinations. Furthermore, if students wish to enter a university, there is another national examination referred to as the university entrance examination that is administered by the Assessment, Selection and Placement Centre, a unit of the ministry of Education (OSYM). At a policy level, given the sudden influx of refugees into Turkey, almost all resources to support migrant students are allocated to those students with refugee status. Training and course activities are available to all refugee children living in asylum centers and attending public schools. In these schools, there are approximately 12,759 teachers, of which 1024 are citizens of the Republic of Turkey, and 11,735 are Syrians (AFAD, 2016).

Method

Research Design

The research method used in this study was a quantitative comparative analysis of assessment strategies and challenges to assess students with a migration background in the countries under investigation.

Research Sample

A purposeful sampling strategy was used in this study based on geographical spread of the participants in Austria (n = 100), Ireland (n = 120) and Turkey (n =120). The survey was also administered to all principals in four out of eighteen Norwegian counties with a total of 29 responses. In this regard, given the low response rate in the Norwegian sample, caution is advised when interpretations are made in relation to analyses of Norwegian data.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

49

Research Instruments and Procedures

All questionnaire responses received a score for bi-polar response alternatives. This score did not apply to questions that provided nominal data asking for yes/no responses. The questionnaire was also translated into the official language of each country. The validity of the translations was also checked by subject field experts in each country.

Data Analysis

Descriptive analysis was used to examine the mean and standard deviation of the responses given in each country. Non-parametric analysis of variance was used to see if there was any significant difference between countries. Kruskal Wallis analysis was used since the number of groups was not equal and there were under 30 participants in the groups. For all Kruskal-Wallis ANOVA tests, α was set at 0.05. When the variance between the groups was significant, Mann Whitney U test was performed for paired comparisons. Bonferroni correction was made to mitigate Type-1 errors in Mann Whitney U tests (Tabachnik & Fidel, 2007). Bonferroni correction was determined with p (significance level) / k (number of groups), and since the number of variables was four, the significance level was found to be .0125. The reliability of the questionnaire was checked using Cronbach's alpha coefficient on all scaled items that had an ordinal measurement scale. All item subscales had a Cronbach's alpha coefficient greater than 0.7 and, in this regard, would be considered statistically reliable.

Results

This part of the paper provides an analysis of questionnaire responses and is divided into four sections: (1) School and student profile characteristics; (2) Supports to enhance teacher's assessment of migrant students; (3) Strategies used by schools to assess students with a migration background; and (4) challenges for the assessment of students with a migration background.

School and Student Profile Characteristics

School and student profile characteristics are presented in Table 1.

Table 1

Student Profile Characteristics

Variables	Answer Choices	Austria		Ireland		Turkey		Norway		Total	
		n	%	n	%	n	%	n	%	n	%
Student Population/ School size	30-500	99	99.0	43	35.8	57	47.5	25	86.2	224	60.7
	501-1000			76	63.3	42	35.0	4	13.8	122	33.1
	1001 - 1500			1	.8	15	12.5			16	4.3
	1501 - 2500					6	5.0			6	1.6
	Missing Value	1	1.0							1	.3
% of student population with migration backgrounds and whose language in most cases is different from that of the language of the classroom	None	10	10.0	2	1.7	11	9.2	3	10.3	26	7.0
	1% to 20%	65	65.0	111	92.5	103	85.8	21	72.4	300	81.3
	21% to 40%	11	11.0	4	3.3	3	2.5	4	13.8	22	6.0
	41% to 60%	5	5.0	1	.8	1	.8	1	3.4	8	2.2
	61% to 80%	6	6.0	1	.8					7	1.9
	More than 80%	3	3.0	1	.8	2	1.7			6	1.6

Regarding student profile characteristics, more than 81% of schools had a migrant population of between 1% and 20% (Table 1). Table 2 shows the responses of participating schools to whether any policy was followed regarding the evaluation of the academic achievement of migrant students.

Table 2

Policy on Assessment for Migrant Student

Variables	Answer Choices	Austria		Ireland		Turkey		Norway		Total	
		n	%	n	%	n	%	n	%	n	%
Policy on assessment for migrant student	Yes	15	15.0	7	5.8	27	22.5	10	34.5	59	16,0
	No	76	76.0	113	94.2	93	77.5	19	65.5	301	81,6

In Table 2, somewhat surprisingly, given the percentage of schools who had students with a migrant background, more than 80% of schools in all countries did not have a policy on assessment of students with a migration background.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

51

Supports to Enhance Teacher's Assessment of Migrant Students

The responses given by participants related to professional development opportunities that are available for staff at their school to help them reflect on their own cultural backgrounds, experiences, and expectations of students with migrant backgrounds are provided in Table 3.

Table 3

Professional Development Opportunities – Integration of culturally relevant materials into assessment practices

Item	Answer Choices	Austria		Ireland		Turkey		Norway		Total	
		n	%	n	%	n	%	n	%	n	%
Are professional development opportunities available for staff in your school to enhance their skill in integrating culturally relevant materials into assessment practices?	Yes	47	47.0	22	18.3	36	30.0	17	58.6	122	33.1
	No	37	37.0	98	81.7	84	70.0	12	41.4	231	62.6
	Missing Value	16	16.0							16	4.3

As can be seen from Table 3, when Principals were asked if professional development opportunities were available for staff to enhance their skills towards the integration of culturally relevant materials into assessment practices, almost 59% of respondents in Norway and 47% of respondents in Austria expressed that professional development opportunities were available. However, these values were considerably lower for Turkey (30%) and Ireland (18%).

The responses given by participants related to professional development opportunities available to staff that help them reflect on their own cultural backgrounds, experiences, and expectations of students are presented in Table 4.

Table 4

Professional Development Opportunities - Experiences, and Expectations of Students with a Migration Background

Item	Answer Choices	Austria		Ireland		Turkey		Norway		Total	
		n	%	n	%	n	%	n	%	n	%
Are professional development opportunities available for staff at your school to help them reflect on their own cultural backgrounds, experiences, and expectations of students with migrant backgrounds?	Yes	40	40.0	16	13.3	76	63.3	9	31	141	38.2
	No	38	38.0	104	86.7	44	36.7	19	65.6	205	55.6
	Missing Value	22	22.0					1	3.4	23	6.2

As can be seen in Table 4, resonating with the lack of professional development opportunities to integrate culturally relevant materials into assessment practices; except for Turkey (63.3%), professional development opportunities that allowed staff to help them reflect on their own cultural backgrounds, experiences, and expectations of students with migrant backgrounds were also limited in Austria (40%) , Ireland (13.3%), and Norway (31%).

The extent to which teachers have adequate training to diagnose, support, communicate and assess students with migration backgrounds is shown in Table 5.

Table 5

The Extent to Which Teachers Have Adequate Training to Diagnose, Support, Communicate and Assess Students with Migration Backgrounds

	Austria				Ireland				Turkey				Norway		
Item	n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s	
(a) diagnosing the diverse needs of students	86	2.29	.99	3	120	2.02	.64	2	120	2.91	1.05	1	29	2.59	.68	2
(b) support individual student's learning and assessment needs	86	2.64	.93	1	120	2.44	.66	1	120	2.90	.98	2	29	3.07	.59	1
(c) communicate with culturally diverse students and their parents or guardians	86	2.23	.95	4	120	1.53	.71	3	120	2.85	1.02	3	29	2.31	.89	4
(d) assess students with migrant backgrounds	86	2.34	.95	2	120	1.38	.59	4	120	2.63	1.09	4	28	2.39	.83	3
	2.37				2.84				2.82				2.59			

Regarding the extent to which teachers had adequate training to diagnose, support, communicate and assess students with migration backgrounds, principals in Austria, Ireland and Norway perceived item b (support individual student's learning and assessment needs) as being the most adequate training provided in these countries. In the case of Turkey, item a (diagnosing the diverse needs of students) and b (support individual student's learning and assessment needs) had similar values. On the other hand, principals in Austria and Norway perceived item c (communicate with

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

53

culturally diverse students and their parents or guardians) as being the least adequate training provided. In the case of Ireland and Turkey, principals perceived item d (assess students with migrant backgrounds) as being the least adequate training provided. Kruskall Wallis analysis results are presented related to the extent to which teachers have adequate training to diagnose, support, communicate and assess students with migration backgrounds in Table 6.

Table 6

Kruskall Wallis Analysis Results – The Extent to Which Teachers Have Adequate Training to Diagnose, Support, Communicate and Assess Students with Migration Backgrounds

Item	Country	n	Mean Rank	s	χ^2	P	Difference U
(a) diagnosing the diverse needs of students	1. Austria	86	163.45	3	49,752	,000	1-3
	2. Ireland	120	137.95				2-3
	3. Turkey	120	223.34				2-4
	4. Norway	29	199.26				
(b) support individual student's learning and assessment needs	1. Austria	86	173.56	3	24.152	,000	2-3
	2. Ireland	120	148.07				2-4
	3. Turkey	120	200.23				
	4.Norway	29	223.02				
(c) communicate with culturally diverse students and their parents or guardians	1. Austria	86	183.17	3	97.85	,000	1-3
	2. Ireland	120	112.19				1-2
	3. Turkey	120	236.54				2-3
	4.Norway	29	192.76				2-4
(d) assess students with migrant backgrounds	1. Austria	86	202.14	3	100.11	,000	1-2
	2. Ireland	120	105.88				2-3
	3. Turkey	120	223.58				2-4
	4.Norway	28	211.29				

In Table 6, analysis of variance using the Kruskal Wallis test revealed a significant difference for all items a [$\chi^2_{(3)}$= 49.752; p<.05], b [$\chi^2_{(3)}$ = 24.152; p<.05], c [$\chi^2_{(3)}$= 97.85; p<.05] and d [$\chi^2_{(3)}$= 100.11; p<.05]). Further analysis using the Mann Whitney U test to find between groups variance also revealed a significant difference between groups for all items. For item a, principals in Austria are of the view that they have less opportunities to diagnosing the diverse needs of students compared to Turkey and principals in Ireland are of the view that have less opportunities to diagnosing the diverse needs of students compared to Turkey and Norway. For item b, principals from Ireland thought that there was less training provided to diagnose, support, communicate and assess students with migration backgrounds when compared to Turkey and Norway. For item c, principals from Ireland thought that there was less training provided to diagnose, support, communicate and assess students with migration backgrounds when compared to Austria, Turkey, and Norway. Also, principals from Austria thought that there was less training provided to diagnose, support, communicate and assess students with migration backgrounds when compared to Turkey. For item d, principals thought that there was less training in Ireland compared to Austria, Turkey, and Norway.

Strategies Used by Schools to Assess Students with a Migration Background

Principals' views on the extent to which assessment practices are used in schools is provided in Table 7.

Table 7

Assessment Practices That are Used in Schools

Item	Austria				Ireland				Turkey				Norway			
	n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s	
(a)This school assesses students' intellectual and academic strengths and weaknesses, and development needs	83	2.93	.89	4	120	3.62	.63	1	120	2.82	1.03	2	29	3.72	.53	1
(b)Teachers in this school use a range of assessment strategies that provide students with migrant backgrounds with opportunities to demonstrate their mastery and skills	82	3.12	.78	2	120	3.22	.69	2	120	2.78	1.00	3	29	3.45	.63	2
(c)Teachers in this school utilise culturally appropriate assessment tools for assessing migrant students	83	2.41	.92	7	120	2.35	.68	7	120	2.58	1.05	7	29	2.21	.77	7
(d)Teachers in this school use a wide range of assessment tools (for example, portfolio, oral presentations, mapping tests, project work) for assessment of students with migrant backgrounds	83	3.34	.72	1	120	3.02	.71	3	120	2.61	1.02	6	29	3.38	.56	3

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

55

Table 7 Continue

Item	Austria				Ireland				Turkey				Norway			
	n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s		n	$\bar{x}$	s	
(e)Staff in this school frequently collaborate to implement the best practices for assessment of students with migrant backgrounds	83	2.88	.90	5	120	2.65	.67	5	120	2.64	1.04	5	27	2.92	.78	4
(f)Teachers utilize information from several sources, including families, in assessing students' achievements	84	2.62	.90	6	120	2.65	.69	5	120	2.85	.97	1	29	2.52	.99	6
(g)Assessment data for students with migrant backgrounds is consistently used to inform teaching/learning	84	2.94	.87	3	120	2.76	.64	4	120	2.75	.99	4	28	2.64	.95	5
	3.37				3.38				3.17				3.47			

As can be seen from Table 7, in the case of Austria, item d (Teachers in this school use a wide range of assessment tools (for example, portfolio, oral presentations, mapping tests, project work for assessment of students from migrant backgrounds) was applied the most. In the case of Ireland and Norway, item a (This school assesses students' intellectual and academic strengths and weaknesses, and development needs) was applied the most. For Turkey, item f (teachers utilize information from several sources, including families, in assessing students' achievements) was applied the most. However, while principals in all four countries to a large extent were of the view that their schools assessed students' intellectual and academic strengths and weaknesses, and development needs; and teachers used a range of assessment strategies that provided students with migrant backgrounds with opportunities to demonstrate their mastery and skills; differences were found in the extent to which these practices occurred.

On the other hand, and worryingly in terms of the migration population of students in all countries, the least used assessment item for all countries was item c (Teachers in this school utilise culturally appropriate assessment tools for assessing migrant students). In this regard, when comparing results derived from all items, one could infer that, whilst assessment strategies occured in schools, they were applied to the general population of the school with limited differentiation in terms of culturally responsive assessment practices.

Kruskall Wallis analysis results were presented related to levels of implementation of assessment practices in schools for migrant students in Table 8.

Table 8

Kruskall Wallis Analysis Results Related to Assessment Practices Used in Schools

		n	Mean Rank	s	χ2	P	Difference U
(a)This school assesses students' intellectual and academic strengths and weaknesses, and development needs	1. Austria	83	143.96	3	67,36	,000	1-2
	2. Ireland	120	222.27				1-4
	3. Turkey	120	139.44				3-2
	4. Norway	29	233.57				3-4
(b)Teachers in this school use a range of assessment strategies that provide students with migrant backgrounds with opportunities to demonstrate their mastery and skills	1. Austria	82	179.80	3	18,027	,000	3-2
	2. Ireland	120	190.02				3-4
	3. Turkey	120	149.17				
	4. Norway	29	218.26				
(c)Teachers in this school utilise culturally appropriate assessment tools for assessing migrant students	1. Austria	83	175.80	3	6,617	,085	
	2. Ireland	120	166.25				
	3. Turkey	120	192.97				
	4. Norway	29	152.78				
(d)Teachers in this school use a wide range of assessment tools (for example, portfolio, oral presentations, mapping tests, project work) for assessment of students with migrant backgrounds	1. Austria	83	214.98	3	36,816	,000	1-2
	2. Ireland	120	176.72				1-3
	3. Turkey	120	139.75				2-3
	4. Norway	29	217.53				4-3
(e)Staff in this school frequently collaborate to implement best practices for assessment of students with migrant backgrounds	1. Austria	83	193.95	3	6,298	,098	
	2. Ireland	120	163.79				
	3. Turkey	120	170.35				
	4. Norway	27	193.74				
(f)Teachers utilize information from several sources, including families, in assessing students' achievements	1. Austria	84	170.21	3	6,299	,098	
	2. Ireland	120	167.96				
	3. Turkey	120	194.62				
	4.Norway	29	161.14				
(g)Assessment data for students with migrant backgrounds is consistently used to inform teaching/learning	1. Austria	84	193.24	3	4,069	,254	
	2. Ireland	120	169.84				
	3. Turkey	120	175.38				
	4. Norway	28	159.59				

Analysis of variance using the Kruskal Wallis test further revealed that there was a significant difference between the principals' views for items a [$\chi^2_{(3)}$= 67.36; p<.05], b [$\chi^2_{(3)}$= 18.027; p<.05] and d [$\chi^2_{(3)}$= 36.817; p<.05] . Further analysis of variance using the Mann Whitney U test to find between groups variance revealed significant differences for item a, with this item reflecting principals' views less in Austria and Turkey compared to Ireland and Norway. For item b, principals had fewer opinions in Turkey compared to Ireland and Norway. Finally, in terms of culturally appropriate assessment methods, there was a significant difference for item d with this item reflecting fewer opinions of principals in Turkey compared to principals in the other countries and in Ireland, compared to Austria. Principals' views on teachers' use of assessment techniques that have the potential to be culturally fair are presented in Table 9.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE

57

Eurasian Journal of Educational Research 79 (2019) 39-68

Table 9

Principals' Views on Teachers' Use of Assessment Techniques That Have the Potential to Be Culturally Fair

Item	Austria				Ireland				Turkey				Norway			
	n	x̄	s		n	x̄	s		n	x̄	s		n	x̄	s	
(a)Self-Assessment	84	3.13	.72	4	120	3.35	.56	3	120	2.80	1.00	4	28	3.35	.67	3
(b)Peer assessment	84	2.32	.85	3	120	3.28	.56	5	120	2.76	1.00	5	29	3.06	.75	5
(c) Portfolios	83	3.12	.90	5	120	2.91	.75	7	120	2.57	1.03	8	29	2.48	.87	10
(d)Students write assessment items	83	2.44	.81	8	120	2.95	.67	6	120	2.40	1.06	9	29	2.51	.57	9
(e)Students writing assessment criteria	83	1.09	8.7	10	120	2.68	.66	9	120	2.21	1.07	10	29	2.82	.71	7
(f)Oral assessment	83	2.48	8.8	7	120	3.58	.60	1	120	3.04	1.02	1	29	3.58	.50	2
(g)Oral presentations	81	3.55	.59	1	120	3.34	.64	4	120	2.84	1.05	2	29	3.62	.49	1
(h) Project work	83	3.38	.71	2	120	3.51	.59	2	120	2.82	1.03	3	29	3.10	.55	4
(i)Artistic/Dramatical performances	83	2.54	.97	6	120	2.45	.79	10	120	2.69	1.01	6.5	29	3.00	,53	6
(j)Designing and developing Individualised Learning Plans	83	2.32	.95	9	120	2.84	.72	8	120	2.69	1.03	6.5	29	2.62	.94	8
	2.64				3.09				2.68				3.01			

When principals views on teachers' use of assessment techniques that have the potential to be culturally fair are examined (Table 9); in the case of Austria and Norway, the most used assessment technique was item g (oral presentations). In the case of Ireland and Turkey, the most frequently used assessment technique was item f (oral assessment). On the other hand, the least used assessment item in Austria and Turkey was item e (students writing assessment items). In the case of Ireland, the least used assessment item was item i (artistic/dramatical performances). Finally, in the case of Norway, the least used assessment item was item c (Portfolios).

Kruskall Wallis analysis results related to principals' views on teachers' use of assessment techniques that have the potential to be culturally fair are presented Table 10.

58

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

Table 10

Kruskall Wallis Analysis – Principals' Views on Teachers' Use of Assessment Techniques That Have the Potential to Be Culturally Fair

Item		n	Mean Rank	s	χ2	P	Difference U
(a) Self Assessment	1. Austria	84	174.57	3	21.64	.000	2-3
	2. Ireland	120	200.68				
	3. Turkey	120	147.55				
	4.Norway	28	202.73				
(b) Peer assessment	1. Austria	84	119.18	3	59.18	.000	1-2
	2. Ireland	120	220.75				1-3
	3. Turkey	120	168.92				1-4
	4.Norway	29	196.86				2-3
(c) Portfolios	1. Austria	83	211.03	3	21.38	.000	1-3
	2. Ireland	120	183.16				1-4
	3. Turkey	120	154.67				
	4.Norway	29	140.43				
(d) Students write assessment items	1. Austria	83	156.66	3	25.64	.000	1-2
	2. Ireland	120	212.25				2-3
	3. Turkey	120	158.05				2-4
	4.Norway	29	161.71				
(e) Students writing assessment criteria	1. Austria	83	139.34	3	34.66	.000	1-2
	2. Ireland	120	207.13				1-4
	3. Turkey	120	160.07				2-3
	4.Norway	29	224.09				3-4
(f) Oral assessment	1. Austria	83	178.98	3	19.94	.000	2-3
	2. Ireland	120	198.79				
	3. Turkey	120	148.08				
	4.Norway	29	194.78				
(g) Oral presentations	1. Austria	81	208.86	3	34.36	.000	1-3
	2. Ireland	120	179.62				2-3
	3. Turkey	120	139.20				3-4
	4.Norway	29	215.48				
(h) Project work	1. Austria	83	193.70	3	35.94	.000	1-3
	2. Ireland	120	207.21				2-3
	3. Turkey	120	140.31				2-4
	4.Norway	29	149.97				
(i) Artistic/Dramatical performances	1. Austria	83	170.66	3	12.52	.006	2-4
	2. Ireland	120	158.00				
	3. Turkey	120	188.47				
	4.Norway	29	220.26				
(j) Designing and developing Individualised Learning Plans	1. Austria	83	144.92	3	13.15	.004	1-2
	2. Ireland	120	193.49				1-3
	3. Turkey	120	182.63				
	4.Norway	29	171.21				

In Table 10, analysis of variance using the Kruskal Wallis test revealed a significant difference for all items a [$\chi^2_{(3)}$= 21.64; p<.05], b [$\chi^2_{(3)}$= 59.18; p<.05], c [$\chi^2_{(3)}$= 21.38; p<.05], d [$\chi^2_{(3)}$= 25.64; p<.05], e [$\chi^2_{(3)}$= 34.66; p<.05], f [$\chi^2_{(3)}$= 19.94; p<.05], g [$\chi^2_{(3)}$= 34.36; p<.05], h [$\chi^2_{(3)}$= 35.94; p<.05], i [$\chi^2_{(3)}$= 12.52; p<.05] and j [$\chi^2_{(3)}$= 13.15; p<.05]). Further analysis using the Mann Whitney U test to find between groups variance also revealed significant differences between groups for all items. For item a (Self-Assessment), there

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

59

was a significant difference for this item reflecting teachers' views more in Ireland compared to Turkey (U=5059.500, p<0.0125). There was also a significant difference for item b (Peer Assessment) which was used less in Austria compared to all countries (Ireland (U=2038.500, p<0.0125), Turkey (U=3739.500, p<0.0125), and Norway (U=663.00, p<0.0125)).

Regarding item c, this item was used more frequently in Austria than in Turkey (U=3486.00, p<0.0125) and Norway (U=721.00, p<0.0125). For item d, this item was used more frequently in Ireland compared to all other countries (Austria (U=3320.500, p<0.0125), Turkey (U=5143.00, p<0.0125), and Norway (U=1166.500, p<0.0125)). In the case of item e, this item was used less frequently in Austria compared to Ireland (U=2928.00, p<0.0125) and Norway (U=617.500, p<0.0125). For item f, this item was used less frequently in Turkey compared to Ireland (U=5166.500, p<0.0125). There was also a significant difference for item g with this item being used less frequently in Turkey compared to all other countries (Austria (U=2999.500, p<0.0125), Ireland (U=5433.00, p<0.0125), and Norway (U=1011.00, p<0.0125)). In the case of item h, this item was used less frequently in Turkey compared to Austria (U=3490.00, p<0.0125) and Ireland (U=4516.00, p<0.0125), and less frequently in Norway than in Ireland (U=1109.00, p<0.0125). There was also a significant difference for item i, with this item being used more frequently in Norway than in Ireland (U=1041.00, p<0.0125). Finally, item j was used less frequently in Austria than in Ireland (U=3550.00, p<0.0125) and Turkey (U=3971.00, p<0.0125).

Challenges for the Assessment of Students with a Migration Background

The answers given by the principals in terms of how fairly and sensitively the students with culturally and linguistically different backgrounds were assessed in the classroom were given in Table 11.

Table 11

Fairness and Sensitivity of Classroom Assessment

	Austria			Ireland			Turkey			Norway		
Item	n	$\bar{x}$	S	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
To what extent is classroom assessment in your school conducted with fairness and sensitivity towards students from culturally and linguistically diverse backgrounds?	82	3.93	.83	120	3.26	.87	120	3.06	1.17	28	3.17	1.02

Finally, table 12 shows that, the majority of principals from Turkey (90%), Norway (83%), Ireland (74%), and Austria (57%) raised concerns about the extent to which high stakes tests were culturally inclusive of students with migrant backgrounds.

60

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

Table 12

Central Examinations

		Austria		Ireland		Turkey		Norway		Total	
Item	Answer Choices	n	%	n	%	n	%	n	%	n	%
Is it the observation of your school that high stakes tests have been designed to be inclusive of students with migrant backgrounds and have been standardized on populations of students similar to your students?	Yes	25	25,0	31	25,8	12	10,0	4	13,8	72	19,5
	No	57	57,0	89	74,2	108	90,0	24	82,8	278	75,3
	Missing Value	18	18,0					1	3,4	19	5,1

Discussion, Conclusion and Recommendations

With significant changes in migration patterns throughout Europe and elsewhere, coupled with the central importance of providing equitable education opportunities to cater for the workforce needs of knowledge-based economies; governments have, in response, attempted to put in place mechanisms and supports to cater for the varying educational needs of students such as those students with a migration backgrounds. Embedded through acts of legislation and various curriculum specifications, there has also been a significant trend towards the promotion of diversity in schools together with a range of complementary assessment for learning strategies, that not only judges but also supports students' learning. A review of the literature suggests that these assessments for learning strategies also have the potential to be culturally fair.

However, against a background of different systems of assessment and accreditation, overall findings from this research suggests that there is a lack of policy at a school level for the assessment of students with migration backgrounds (Table 2). Certainly, it would be reasonable to suggest that many schools do not see the need for a specific school policy on the assessment of students with migrant backgrounds given that, parity of equity to achieve desired learning outcomes as well as respect for diversity is already enshrined in government mandated policy and practice.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

61

It is also evident that, while an appreciation for cultural diversity exists in theory, this study also found that supports for cultural diversity were not necessarily translated into school practice. Except for Norway, the majority of principals in the other three countries thought that professional development opportunities were not readily available for staff to enhance their skill in integrating culturally relevant materials into assessment practices (Table 3). Furthermore, there also appeared to be a belief among a significant number of principals in all countries that the training provided to teachers was not adequate to allow teachers in their schools to diagnose, support, communicate and assess students with migrant backgrounds (Tables 5 and 6). As such, there is a need to evaluate the scope of culturally responsive professional development opportunities available to schools. There is also a need to ascertain the extent to which curriculum specifications at initial teacher education level do in fact place any emphasis on strategies to assess students with migrant backgrounds.

Additionally, whilst principals thought that a range of culturally responsive assessment strategies were used by teachers, there did not appear to be a clearly defined distinction between assessment strategies and tools that were used for students with migration backgrounds and the general population of students in each of the countries. This was apparent when one examined the limited extent to which teachers utilised culturally appropriate assessment tools for assessing migrant students (Tables 6 and 7) and the significant extent to which teachers used assessment techniques that had the potential to be culturally fair such as self and oral assessment (Tables 9 and 10). It is no surprise therefore that, with the limited culturally responsive training provided to schools, coupled with the application of various assessment techniques to the entire student population, the majority of principals thought that due to the frequent use of assessment for learning techniques, their schools conducted classroom assessment with fairness and sensitivity towards culturally and linguistically diverse backgrounds (Table 11). Finally, with cause for concern, the majority of principals were also of the view that high stakes tests had not been designed to be inclusive of students with migrant backgrounds and had not been standardized on populations of students similar to their students.

In conclusion, from a review of policy documents and curriculum specifications in each country, there is evidence to suggest that the foundations for culturally responsive assessment practices are beginning to take shape albeit varying degrees of difference in each country. The survey results also indicate the need for training and professional development and implies that not enough emphasis is being placed on culturally responsive assessment despite the rhetoric that espouses interculturalism. There are undoubtedly many reasons for this that requires further investigation such as the belief that because various assessment for learning techniques are becoming a common feature of classroom practice, assessment needs of culturally diverse students are being met. The research also points to the need for upskilling in culturally responsive leadership, and as a starting point, the development of an overarching culturally responsive assessment framework and toolkit that can be used by policy

makers and schools in order to allay the various interpretations of what it means to satisfy the assessment needs of teachers and students with migration backgrounds.

References

Aceves, T. C., & Orosco, M. J. (2014). *Culturally responsive teaching* (Document No. IC-2). Retrieved from University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center website: http://ceedar.education.ufl.edu/tools/innovation-configurations/ on 10 May 2017

AFAD (2016). *Suriye afet raporu (Syria disaster report)* (Retrieved from, https://www.afad.gov.tr/tr/2372/Afet-Raporu-Suriye, in 22.11.2016).

Baker, E. L., O'Neil, H. F.Jr., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist American Psychological Association*, 48 (12), 1210-1218

Banks, J. A. (2006). *Diversity in american education: Foundations, curriculum and teaching.* Boston: Allyn & Bacon.

Banks, J. A. (2008). *Teaching strategies for ethnic studies.* Boston Allyn & Bacon.

Bennett, J. M. (2009). Cultivating intercultural competence: A process perspective. In Deardorff, D. K. (Ed.), *The SAGE handbook of intercultural competence* (pp. 121-140). Thousand Oaks, CA: Sage.

Bilgili O., Volante L., & Klinger, D. (2018). Immigrant student achievement and the performance disadvantage. In Volante L., Klinger D., Bilgili O. (Eds.), *Immigrant student achievement and education policy- Policy implications of research in education* (pp. 3-150). Springer, Cham.

Bledsoe, K., & Donaldson, S. I. (2015). Culturally responsive theory driven evaluation. In S. Hood, R. Hopson & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 3-28). Charlotte, NC: Information Age Publishing Inc.

Borjas, G. J. (1995). The economic benefits from immigration. *Journal of economic perspectives*, *9*(2), 3-22

Brown, M., McNamara, G., & O'Hara, J. (2016). Quality and the rise of value-added in education: The case of Ireland. *Policy Futures in Education*, 14*(6)*, 810-829

Brown, M., McNamara, G., O'Hara, J., Hood, S., Burns, D., & Kurum, G. (2017) 'Evaluating the impact of distributed culturally responsive leadership in a disadvantaged rural primary school in Ireland'. *Educational Management and Administration*. Retrieved from https://doi-org.dcu.idm.oclc.org/10.1177%2F1741143217739360 on 15 April 2018

Brown, M., McNamara, G., O'Hara, J., & O'Brien, S. (2017) *Inspectors and the Process of Self-Evaluation in Ireland.* In: Baxter J. Eds) School Inspectors. Accountability and Educational Improvement. Springer, Cham. doi: https://doi.org/10.1007/978-3-319-52536-5_4

Brown, M. (2013) *Inquiry into the education and training ınspectorate* (ETI). Northern Ireland Assembly, Northern Ireland. doi: https://doi.org/10.13140/rg.2.2.17129.80486

Brown, M. (2007). Educating all students: Creating culturally responsive teachers, classrooms, and schools. *Intervention in school and clinic*, 43(1), 57-62.

Burns, D., Devitt, A., McNamara, G., O'Hara, J. ,& Brown, M. (2018). Is it all memory recall? An empirical investigation of intellectual skill requirements in Leaving Certificate examination papers in Ireland. *Irish Educational Studies*, 37(3), 351-372. doi: https://doi.org/10.1080/03323315.2018.1484300

Caldas, S.J., & Bankston, C. (1997) Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research*, 90(5), 269-277. Retrieved from

Caspersen, J., Smeby, J.C., & Olaf Aamodt, P. (2017). Measuring learning outcomes. *European Journal of Education*, *52*(1), 20-30.

Castagno, A.E., & Brayboy, B.M.J., (2008). Culturally responsive schooling for Indigenous youth: A review of the literature. *Review of educational research*, *78*(4), 941-993.

Castles, S., De Haas, H., & Miller, M.J., 2013. *The age of migration: International population movements in the modern world*. London: Palgrave Macmillan.

Cavalli Sforza, L. L., & Pievani, T. (2012) *Homo sapiens-The great history of human diversity*. Codice, Turin

Central Statistics Office. (2017). *Census 2016 summary results.* Dublin : Central Statistics Office.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-239.

Chung, R.Y.N., & Griffiths, S.M. (2018). Migration and health in the world: a global public health perspective. *Public Health*, *158*, 64-65.

Cummins, J., (1979). Linguistic interdependence and the educational development of bilingual children. *Review of educational research*, *49*(2), 222-251

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the Crossfire*. Clevedon: Multilingual Matters. eBook.

Czaika, M., & De Haas, H. (2014). The globalization of migration: Has the world become more migratory ?. *International Migration Review*, 48(2), 283-323.

De Backer, F., Van Avermaet, P., & Slembrouck, S. (2017). Schools as laboratories for exploring multilingual assessment policies and practice. *Language and Education*, *31*(3), 217-230. http://dx.doi.org.dcu.idm.oclc.org/10.1080/09500782.2016.1261896

Demmert, Jr, W.G. (2001). *Improving Academic Performance among Native American Students: A Review of the Research Literature.* Retrieved 01 January 2018, from http://www.eric.ed.gov./ERICDocs/data/ericdocs2/content_storage_01/0 000000b/80/0d/de/70.pdf.

Department of Education and Skills. (2012). *Circular 09/2012, Staffing arrangements in post-primary schools for the 2012/13 school year* [Online]. Available from: https://circulars.gov.ie/pdf/circular/education/2012/09.pdf [Accessed 12 March 2016].

Department of Education and Skills. (2010). *Intercultural Education Strategy, 2010 – 2015.* Retrieved from http://www.education.ie/en/Schools-Colleges/Information/Intercultural-Education-Strategy/Intercultural-Education-Strategy.html#sthash.m6ty0EoB.dpuf on 01 February 2018.

Earley, P. C., & Mosakowski, E. (2004). *Cultural intelligence*. Harvard Business Review.

Eurostat. (2017) Migration and migrant population statistics. Available from:http://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Non-national_population_by_group_of_citizenship,_1_January_2017_.png [Accessed 01 February 2018]

Espinosa, L. M. (2005). Curriculum and assessment considerations for young children from culturally, linguistically, and economically diverse backgrounds. gay*Psychology in The Schools*, 42(8), 837-853.

Fillitz, T. (2002). *Interkulturelles lernen. zwischen institutionellem rahmen, schulischer praxis und gesellschaftlichem kommunikationsprinzip*. Innsbruck: Studienverlag.

Ford, D. Y., & Kea, C. D. (2009). Creating culturally responsive instruction: For students' sake and teachers' sake. *Focus on Exceptional Children*, (41), 1-18.

Ford, D.Y., & Moore, J.L., (2013). Understanding and reversing underachievement, low achievement, and achievement gaps among high-ability African American males in urban school contexts. *The Urban Review*, *45*(4), 399-415.

Gay, G. (2010). *Culturally responsive teaching: Theory, research and practice*. (3rd ed.). Multicultural Education Series. New York: Teachers College Press, Columbia University.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

65

Geddes, A., & Scholten, P. (2016). *The politics of migration and immigration in Europe*. London: Sage.

Government of Ireland. (1998). *Education Act.* Dublin: Stationery Office.

Griner, A.C., & Stewart, M.L., (2013). Addressing the achievement gap and disproportionality through the use of culturally responsive teaching practices. *Urban Education*, *48*(4), 585-621.

Hancock, A. (2017). The challenges of multilingualism for schools. Conference presentation. Minority language pupils and the curriculum: closing the achievement gap. *Marino Institute of Education*. 25 April 2017

Hanrahan, S.J., & Isaacs, G. (2001). Assessing self-and peer assessment: The students' views. *Higher Education Research & Development*, *20*(1), 53-70.

Harris, M.M. (2006). Cultural skill: An emerging construct for the 21st century. *Global Forum*, 43 (3), 43-47.

Hempel, P.S., & Sue-Chan, C. (2010). Culture and the assessment of creativity. *Management and Organization Review,* 6 (3), 415-435. DOI: 10.1111/j.1740-8784.2010.00189

Hood, S., Hopson, R., & Frierson, H. (Eds.). (2015). *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Charlotte, NC.: Information Age Publishing.

Hood, S. (1998). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *The Journal of Negro Education*, *67*(3), 187 – 196.

Hofstede, G. (2009). The moral circle in intercultural competence. In D. K. Deardorff (Ed.). *The Sage Handbook of Intercultural Competence* (pp. 85 – 94). Sage.

Kim, K. H., & Zabelina, D. (2015). Cultural bias in assessment: Can creativity assessment help? *International Journal of Critical Pedagogy*, *6*(2), 129-147. University of North Carolina.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465-491.

Lovdata. (2016). Lov om grunnskolen og den vidaregående opplæringa (opplæringslova). Retrieved from https://lovdata.no/dokument/NL/lov/1998-07-17-61 on 01 February 2018.

LaFrance, J., Kirkhart, K. E., & Nichols, R. (2015). Cultural views of validity: A conversation. In S. Hood, R. Hopson & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 49-72). Charlotte, NC: Information Age Publishing.

Menken, K. (2010). NCLB and English language learners: Challenges and consequences. *Theory Into Practice*, *49*(2), 121-128.

Mitakidou, S., Karagianni, P., & Tressou, E. (2015) The challenges and agony of researchers in a diverse marginalised context. *Researching Marginalized Groups*, 16-36.

MoNe. (2018). www.meb.gov.tr/meb_iys_dosyalar/2018_03/26191912_yonerge.pdf, Accessed at 20.10.2018.

Morris, A. (2011) *Student Standardised Testing: Current practices in OECD countries and a literature review*. Organisation for Economic Cooperation and Development (OECD), Paris.

NDET. (2017). *The education mirror* 2016. Oslo: NDET.

National Council for Curriculum and Assessment (NCCA). (2009). *Key skills framework.* http://www.ncca.ie/en/Curriculum_and_Assessment/PostPrimary_Education/Senior_Cycle/Key_Skills_Framework/KS_Framework.pdf.

National Council for Curriculum and Assessment (NCCA). (2006). *Intercultural education in the post-primary school: Guidelines for schools.* www.ncca.ie/uploadedfiles/publications/interc%20Guide_Eng.pdf

National Council for Curriculum and Assessment (NCCA). (2005). *Proposals for the Future Development of Senior Cycle Education in Ireland.* www.ncca.ie/uploadedfiles/SCycle%20Advice_0405.pdf.

Nelson-Barber, S. and Trumbull, E. (2007). Making assessment practices valid for Indigenous American students. *Journal of American Indian Education*, 132-147.

OECD (2017). *Education at a Glance. OECD indicators*. Paris: OECD.

Organisation for Economic Co-operation and Development. (2016a). *PISA 2015 results, excellence and equity in education* (Vol. I). Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/9789264266490-en

OECD (2016b). *Education at a Glance*, OECD Indicators Norway. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/eag-2016-73-en

O'Connor, M. C. (1989). Aspects of differential performance by minorities on standardized tests: Linguistic and sociocultural factors. In B. R. Gifford (Ed.), *Test policy and test performance: Education, language, and culture* (pp. 129-181). Boston: Kluwer.

Padilla, A. M. (2001). *Issues in culturally appropriate assessment.* In L.A. Suzuki, J. G. Ponterotto, & P.J. Meller (Eds.). *Handbook of multicultural assessment* (2nd ed.) (pp. 527). San Francisco: Jossey-Bass.

Poole, P., Brown, M., McNamara, G., O'Hara, J., O'Brien, S., & Burns, D. (2018). Challenges and supports towards the integration of ePortfolios in education.

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE

67

Eurasian Journal of Educational Research 79 (2019) 39-68

Lessons to be learned from Ireland. *Heliyon,* 4(11). doi: https://doi.org/10.1016/j.heliyon.2018.e00899

Qualls, A. L. (1998). Culturally responsive assessment: Development strategies and validity issues. *The Journal of Negro Education*, 67(3), 296-301. doi:10.2307/2668197

Randen, G. T. (2015). Assessment of minority language proficiency in primary school. *Norwegian as a Second Language*, 30(1-2), 350 – 372

Reynolds, M., & Trehan, K. (2000). Assessment: a critical perspective. *Studies in Higher Education*, *25*(3), 267-278.

Rothstein-Fisch, C., & Trumbull, E. (2008). *Managing diverse classrooms: How to build on students' cultural strengths*. Alexandria, US: Association for Supervision & Curriculum Development (ASCD), 2008. ProQuest ebrary. Web. 2 October 2016.

Rychen, D., & Salganik, L. (2003). *The definition and selection of key competencies: executive summary* [Online]. Available from: http://www.oecd.org/pisa/35070367.pdf [Accessed 17 September 2018]

Scheerens, J., Luyten, H., & van Ravens, J. (2011). Measuring educational quality by means of indicators. In J. Scheerens, H. Luyten, & J. van Ravens (Eds.), *Perspectives on educational quality: illustrative outcomes on primary and secondary schooling in the Netherlands* (pp. 35-50). (SpringerBriefs in education; Vol. 1, No. 1). Dordrecht: Springer. DOI: 10.1007/978-94-007-0926-3_2

Schapiro, K. A. (2009). *Migration and Educational Outcomes of Children*. Human Development Research Paper 57, United Nations Development Programme, Human Development Report Office, New York.

Shewbridge, C., Jang, E., Matthews, P., & Santiago, P. (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark*, OECD, Paris, available from *www.oecd.org/edu/evaluationpolicy*.

Slee, J. (2010). A systemic approach to culturally responsive assessment practices and evaluation. *Higher Education Quarterly*, *64*(3), 246-260.

Smith-Maddox, R. (1998). Defining Culture as a Dimension of Academic Achievement: Implications for Culturally Responsive Curriculum, Instruction, and Assessment. *Journal of Negro Education*, *67*(3), 302-17

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Oxon: Routledge.

Tabachnik, B., & Fidel, L. S. (2007). *Using multivariate statistics*. (Fifth Edition) Boston: Pearson.

Tan, J. S. (2004). Issues & observations: Cultural intelligence and the global economy. *Leadership in Action*, *24*(5), 19-21.

68

Funda NAYIR-Martin BROWN-Denise BURNS-Joe O'HARA-Gerry MCNAMARA-Guri
NORTVEDT-Guri SKEDSMO-Silje Kristin GLOPPEN-Eline F. WIESE
Eurasian Journal of Educational Research 79 (2019) 39-68

Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, *26*(6), 605-614.

Tichá, R., & Abery, B. (2016). Beyond the Large-scale Testing of Basic Skills: using formative assessment to facilitate student learning. in W.C. Smith (Ed.) *The Global Testing Culture: shaping education policy, perceptions, and practice*. Oxford: Symposium Books,

Topping, K.J. (2009). Peer assessment. *Theory into Practice*, *48*(1), 20-27.

United Nations Department of Economic and Social Affairs. (2017a). *International Migration Report 2017*. Retrieved from http://www.un.org/en/development/desa/population/migration/public ations/migrationreport/docs/MigrationReport2017.pdf

United Nations Department of Economic and Social Affairs. (2017b). *International Migration Report 2017 – Highlights.*

Villegas, A.M., & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach*. Suny Press: New York.

Wood, D., & Kurzel, F. (2008). Engaging students in reflective practice through a process of formative peer review and peer assessment. *ATN Assessment*, On 18 July 2018, Retreived from search.ror.unisa.edu.au/media/.../53109106480001831.

Young, C., McNamara, G., Brown, M., & O'Hara, J. (2018) Adopting and adapting: school leaders in the age of data-informed decision making. *Educational Assessment, Evaluation and Accountability*, *30*(2), 133-158. doi: https://doi.org/10.1007/s11092-018-9278-4

Investigation of the Orthogonality Assumption in the Bifactor Item Response Theory*

Fulya BARIS PEKMEZCI[1], H. Deniz GULLEROGLU[2]

A R T I C L E I N F O	A B S T R A C T
	Purpose: This study aims to investigate the orthogonality assumption, which restricts the use of Bifactor item response theory under different conditions. **Method:** Data of the study have been obtained in accordance with the Bifactor model. It has been produced in accordance with two different models (Model 1 and Model 2) in a simulated way. **Results:** As a result of the research, it was found out that the case that two factors were correlated (Model 1) and that all factors were correlated (Model 2) had the same effect on the accuracy of both person and item parameter estimations. While estimating the discrimination parameters, as the orthogonality violation increased, it was concluded that the bias increased, too. As the test length increased, the accuracy of estimations of discrimination and difficulty parameters, namely the reliability decreased. Increasing the number of items increased the accuracy of person parameters, which was the reliability.

Implication for Research and Practice: As test length increases, the Bifactor theory can better tolerate the orthogonality violation in estimation of person parameters. The practitioners who want to use this theory are recommended to work with large item pools. At all correlation levels, the accuracy of the parameter estimations was approximately the same. New studies can be repeated with intermediate correlation levels. Among all the parameters, the parameters whose estimation reliability is the lowest were found to be person parameters.

---

[1] Yozgat Bozok University, TURKEY. e-mail: fulyabaris@gmail.com, ORCID: https://orcid.org/0000-0001-6989-512X

[2] Ankara University, TURKEY. e-mail: denizgulleroglu@yahoo.com, ORCID: https://orcid.org/0000-0001-6995-8223

Introduction

Bifactor item response theory model was developed by Holzinger and Swineford (1937) as an extension of Spearman's Bifactor theory, as can be understood from his name. Bifactor theory assumes that there are more than one specific factor and a general factor explained by these factors, and that these specific effects also have an effect on the general factor (Spearman, 1904). As in all item response theory models, the Bifactor Model has its own assumptions. One of the assumptions of the Bifactor Model is that the data include both general and specific factors. The other assumption that the factors are orthogonal is not possible to be met in practice. In other words, test developers should write only the primary factor and also the items that measure a subdomain. The main problem is that writing such items in practice is very difficult.

According to Canivez (2016), the main advantages of the Bifactor Model are generally these: (a) the effect of the overall factor on each item and groups of items can be easily interpreted. This is not achievable with second-order models, correlated trait models, and uni-dimensional models (Chen, West & Sousa, 2006; Immekus & Imbrie, 2008); (b) the effects of both general and specific factors on the items can be estimated simultaneously (Reise, 2012; Reise, Moore & Haviland, 2010); (c) the psychometric properties that are required to score and interpret general and specific factors are obtainable through the Bifactor Model (DeMars, 2013); (d) the specific effects of general and specific traits in describing other variables are obtained more accurately; and (e) the Bifactor Model provides more accurate and reliable estimations than testlet-effect model in estimating item and person parameters.

Bifactor Model is very common in scaling the psychological properties, and differentiates the specific contributions of the facets on the general factor very well. Therefore, the Bifactor Model is quite suitable for scale development. While developing or evaluating a new multifaceted scale that aims to assess the general structure and specific facets, the power of factor loadings at general and specific factors will be a guide in choosing and evaluating items. The items will ideally have a higher loading at the general factor or at least a greater loading than the specific factor. If the items have a higher loading than the facets in the general structure, these items will be selected, however, if specific factors have larger loadings than the general factor, these items will be removed from the scale. The reason for this is that these items do not contribute significantly to the general structure. Moreover, the Bifactor Model is also used to create a uni-dimensional scale or a short uni-dimensional scale from a multidimensional scale (Stucky & Edelen, 2014; Stucky, Edelen, Vaughan, Tucker & Butler, 2014; Stucky, Thissen & Edelen, 2013). The applications of the Bifactor Model in education indicate that this model is useful in terms of scoring the subscales and assessing the reliability when subscale scores need to be used (Cucina & Byle, 2017; DeMars ,2013; Golay & Lecerf, 2011; Watkins & Beaujea, 2014).

In addition to these advantages, the Bifactor Model has also some limitations. The biggest limitation is the difficulty of meeting the orthogonality assumption of the Bifactor model (Chen, West & Sousa, 2006; Simms, Grös, Watson & O'Hara., 2008). As in the structural equation model, the Bifactor model needs a considerably larger

sample when compared to the total score and individual score approach. Additionally, the Bifactor model interpretations become quite complicated when correlations are allowed among specific factors (Rindskopf & Rose, 1988), and the model can often not be identified. In addition to these, it also gives a weak model adaptation in weak or small factor loadings as in other factorial models (Jennrich & Bentler, 2012; MacCallum, Widaman, Zhang & Hong, 1999).

When the literature about Bifactor Models is reviewed, it has been seen that the focus has always been on the determination of dimensionality and the examination of item performance in the field of education and psychology, algorithm of the Bifactor Model, and comparison of different item response theory models with Bifactor model (Brouwer, Meijer, Weekers & Baneke, 2008; Brown, Finney & France, 2011; Chen, West & Sousa, 2006; Chen, Hayes, Carver, Laurenceau & Zhang, 2012; Demars, 2006; Fukuhara, 2009; Garn, 2017; Gibbons et al., 2007; Hyland, Boduszek, Dhingra, Shevlin & Egan, 2014; Lafond, 2014; Martel, Von Eye & Nigg, 2010; Reise, Ventura et.al, 2011; Rijmen, 2009; Rodriguez, Reise & Haviland, 2016; Thomas, 2012;Yang, Song & Xu, 2002).

Although the situation that limits the use of the Bifactor Model is the orthogonality assumption, there has been only one study (Zheng, 2013) carried out in the field about testing of the orthogonality assumption under different conditions. This work (Zheng, 2013) has also been carried out under limited conditions in such a way in every simulation study. Contrary to Zheng's (2013) study, in this study, simulation conditions (test length and correlation levels) were changed. Moreover, item and person parameters were estimated according to Bayesian approach (by Quasi Monte-Carlo estimation). In addition to Zengh's (2013) study, Rindskopf and Rose (1988) found that the interpretation of model parameters gets complicated as correlations among specific factors are allowed in the Bifactor Model. Since cross loadings between factors will also allow correlations between factors, this can be considered as a kind of correlation between factors. Rindskopf and Rose (1988) could not reach any information about level of these cross loadings.

Purpose

The Bifactor Model is a theory that is limited in its use due to the orthogonality assumption that it requires. In addition to this limitation, this model is frequently used in studies of modeling psychological and educational constructs, and developing scales by ignoring the assumption. In cases where the orthogonality assumption is not met, it is not going to be possible to model psychological and educational constructs accurately for the developed scale to reach a correct factorial structure and to have correct parameter estimations. Besides, it is almost impossible to develop measurement instruments in which the correlation between factors in the fields of education and psychology is zero. Forcing the correlated factors to be orthogonal will cause loss of information regarding the measured structure, and will result in unreliable parameter estimations. The precision and the accuracy of parameter estimations, on the other hand, are important in every measurement because parameter estimations are an important element in determining item performance and

respondents' ability level. Resulting from all these reasons, it is necessary to examine the Bifactor model by allowing different correlations among specific factors, in other words, to determine if stable, precise and accurate estimations can be done despite orthogonality violation by which levels of violation are tolerated by the theory itself. It is thought that via this research, the results that are going to be obtained through examining and evaluating the orthogonality assumption that restricts the bifactor model usage under certain criteria will highly contribute to the field.

Method

Research Design

This research is based on the basic research model since it is carried out through the data obtained by Monte Carlo simulation in order to investigate the effect of the violation of the orthogonality assumption at different levels and test lengths on the item and person parameter estimation.

Simulation Study

The data for this study were generated according to two Bifactor two-parameter models with a simulation according to two models (Model 1 and Model 2). Model 1 was the model which showed the violation of orthogonality due to the cross loadings. In this model, the focus was on the effect of orthogonality violation between two specific factors on parameter estimations in all factors. On the other hand, Model 2 showed the correlations among all the specific factors.

The variables that were manipulated in specific models were the correlation levels between factors and the test lengths. The correlation acceptance levels for the models that were set up (Model 1 and Model 2) were as 0.10 (very low), 0.40 (medium), 0.70 (high) (Cohen, 1988). In the framework of this research, it was decided that the minimum test length to be 12 items with reference to the fact that a factor should have at least three items in order to be called as a factor (Kline, 1994). Different from the literature on the field, other test lengths were taken as 40 and 100, taking into consideration that the number of items in each factor was equal.

The variable to be kept constant, namely not to be manipulated, during the research was the sample size. In order to prevent bias that would arise from sample size, the largest sample size (5000) that were used in the current studies was set as a simulation sample. The summary of the research design is given in Table 1.

Table 1

The Summary of the Research Design

	Correlation Levels	Test Length	Sample Size
Model 1			
Model 1.1	$r_{3.4}$=0.10	12-40-100	5000
Model 1.2	$r_{3.4}$=0.40	12-40-100	5000
Model 1.3	$r_{3.4}$=0.70	12-40-100	5000
Model 2			
Model 2.1	$r_{2:4\leq}0.10$	12-40-100	5000
Model 2.2	$0.10 < r_{2:4} \leq 0.40$	12-40-100	5000
Model 2.3	$0.40 < r_{2:4\leq}0.70$	12-40-100	5000

As a result of the literature review, it has been seen that the replication numbers used in the Bifactor Models are generally 100 (Demars, 2006; Zhang, 2008), 200 (Zheng, 2013) and 500 (Cai, Yang & Hansen, 2011). In this study, number of replications was determined as 200 to be practical.

In order to generate the two-parameter Bifactor model data set with the determined number of replications, the distribution of the discrimination parameter (a), the difficulty parameter (b) and the person parameter (θ) should be determined. A simulation model in which the discrimination parameter (a) was uniformly distributed between a range of 0.2 to 2.0 ratios, the difficulty parameter (b) and the person parameter (θ) that were randomly distributed were set. The mathematical expression of the Bifactor two-parameter model was as follows:

$$p(y=1\,|\,\theta_g,\theta_s)=\frac{1}{1+\ exp^{\{-(d+a_g\theta_g+a_s\theta_s\ )\}}}$$

The distribution characteristics of the discrimination, difficulty and person parameters were the same for the 18 (2x3x3) condition given in Table 1. A random seed was assigned to the true parameters, which were generated for the first condition and in other conditions, and via this seed, invariance of true parameters between models was provided. The difficulty coefficient that had been produced was transformed into a multidimensional difficulty coefficient by the following formula:

$$d = -b \sqrt{a_g^2 + a_s^2}$$

This study was carried out based on Monte Carlo method using R 3.4.0 GUI software with syntax (Zheng, 2013), which was written to simulate the data according to the determined conditions and to produce Bifactor model parameters. For the accuracy of the generated syntax and the generated data files, the average bias was calculated on a model that did not contain orthogonality violation, and it was observed that the bias average was close to zero.

The Data Analysis of Simulation

Bifactor model predictions were made with the data that were generated in the simulation, and 200 for each condition with a total of 3600 (18x200) data files were obtained. Bifactor model estimations were made with "mirt" (Chalmers, 2016) package in R 3.4.0 GUI software, and descriptive statistics were generated with "psych" (Revelle, 2017) package.

The evaluation of the accuracy of parameter estimations throughout the replications was carried out via mean bias, root mean square error (RMSE), and standard error of estimates (SE).

$$\text{Average Bias } (\hat{\beta}) = \frac{\sum_{r=1}^{R} (\hat{\beta}_r - \beta)}{R}$$

$$\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_r - \beta)^2}$$

$$\text{SE}(\widehat{\beta}) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_r - \frac{\sum_{r=1}^{R} \hat{\beta}_r}{R})^2}$$

Given in the above formulas;

β: true individual parameter or item parameter

β̂: the individual and substance parameters predicted at the rth replication (Li & Rupp, 2011).

Results

Parameter Estimation Bias

The average bias values calculated from the files that were obtained from 200 replications for the models (Model 1 and Model 2), which were set up, are given in Table 2. In order to examine the recovery in the item parameter estimations, the bias was calculated by taking the average of the difference between the true parameters and the estimated parameters.

When the parameter estimation bias given in Table 2 was examined, the pattern seen for Model 1 and Model 2 was the same in all test lengths for discrimination parameters. When the test length increased from 40 items to 100 items, a decrease in the average bias was observed. Contrary to this, as the test length increaseds, the standard deviation of the bias scores got larger and the range widened. To put it in other words, the increase in the number of items led to a decrease in the reliability of the estimations. This can be explained by the increase in the amount of biased items. That is, the more correlated item was added to the model, the greater the variability got. When the models were examined among within themselves, the standard deviation values increased as the correlation between the factors increased with regard that the mean deviation did not change significantly. The average bias of the item and person parameters is given in Table 2.

Table 2

Mean Bias of Items and Person Parameters

			Mean Bias		
			12 items	*40 items*	*100 items*
			$(\overline{X}, \sigma)$	$(\overline{X}, \sigma)$	$(\overline{X}, \sigma)$
Discrimination parameter	Model 1	Model 1.1	0.010(0.050)	-0.040(0.650)	0.004(0.690)
		Model 1.2	0.002(0.120)	-0.030(0.650)	-0.009(0.690)
		Model 1.3	-0.020(0.310)	-0.030(0.650)	0.002 (0.700)
	Model 2	Model 2.1	0.020(0.070)	-0.050(0.650)	0.010(0.690)
		Model 2.2	0.000(0.180)	-0.030(0.650)	0.000(0.690)
		Model 2.3	-0.001(0.370)	-0.100(0.680)	-0.005(0.720)
Difficulty parameter	Model 1	Model 1.1.	0.260(0.330)	0.049(1.780)	-0.008(1.830)
		Model 1.2	0.240(0.340)	0.083(1.780)	-0.066(1.830)
		Model 1.3	0.260(0.330)	0.058(1.780)	-0.027(1.830)
	Model 2	Model 2.1	0.250(0.330)	0.020(1.780)	-0.040(1.830)
		Model 2.2	0.250(0.340)	0.050(1.780)	-0.060(1.830)
		Model 2.3	0.250(0.340)	0.040(1.780)	-0.060(1.830)
Person parameter	Model 1	Model 1.1	-0.010(0.670)	0.000(0.410)	0.010 (0.320)
		Model 1.2	0.000(0.690)	-0.010(0.440)	0.010 (0.340)
		Model 1.3	-0.010(0.710)	0.000(0.470)	-0.010(0.360)
	Model 2	Model 2.1	0.000(0.690)	0.010(0.410)	0.000 (0.280)
		Model 2.2	-0.001(0.720)	0.000(0.460)	0.000 (0.320)
		Model 2.3	0.000(0.800)	0.000(0.560)	0.020 (0.410)

As it can be seen in Table 2, when the estimation bias of the intercept coefficients was examined, as the test length increased for Model 1, the average bias scores decreased. The increase in the test length for Model 1 affected the parameter estimate recovery. This was not observed evidently when the test length for Model 2 was increased from 40 to 100 items. The standard deviation values increased as the test length increased, in other words the variability increased. The fact that the variability increased the reliability of estimations were reduced. At test lengths of 40 and 100 items, the greatest standard deviation values were observed on the difficulty coefficients. When the models were examined within themselves, although there was not much change in the average of bias, the standard deviations were almost the same. When all test lengths (12, 40, 100) for both Model 1 and Model 2 were examined all together, when the estimation of person parameters were examined, the distorted parameters were found to be at the test length of 12 items. Generally, the variability of bias scores was high at all test lengths. When Model 1 and Model 2 were compared, it was observed that the standard deviations were quite similar. It was observed that as the test length increased, the variability decreased for both Model 1 and Model 2. It can be said that the test length has an effect on the recovery in parameter estimations.

Increasing the test length reduced the variability. This finding is consistent with Zheng's (2013) study. Estimation accuracy was higher at the test lengths of 40 and 100 items when compared to the test length of 12 items. Directly proportional to the test length, the fact that the variability decreased indicated that the test length might have an effect on the recovery of parameters. However, the variability was high at all test lengths, and this reduced the reliability of the parameter estimation.

The Accuracy and Stability of Parameter Estimation

Estimation accuracy and stability of discrimination parameters. Table 3 shows the standard error values and average RMSE values of the discrimination parameters for Model 1 and Model 2. These values are first interpreted by model type, and then by the test length.

As it can be seen in Table 3, when the standard errors on the model basis were examined, it was observed that the table values (average value and standard deviation) were the same for both models (except the 12 item) ($\overline{X}_{SE \: \&Model \: 1}=$ 0.083 , $\overline{X}_{SE\&Model \: 2}=$ 0.086). When the RMSE averages were examined, it was observed that the table values for Model 1 and Model 2 were very close ($\overline{X}_{RMSE \: \&Model \: 1}=$ 0.410 , $\overline{X}_{RMSE \: \&Model \: 2}=$ 0.435)

Table 3

Discrimination Coefficients, Standard Error and Mean RMSE Values for Model 1 and Model 2

		Test Length		
	Model	12 items	40 items	100 items
SE	Model 1.1	0.120(0.090)	0.060(0.020)	0.050(0.010)
	Model 1.2	0.140(0.120)	0.060(0.020)	0.050(0.010)
	Model 1.3	0.160(0.150)	0.060(0.020)	0.050(0.010)
	Model 2.1	0.130(0.110)	0.060(0.020)	0.050(0.010)
	Model 2.2	0.140(0.130)	0.060(0.020)	0.050(0.010)
	Model 2.3	0.180(0.160)	0.060(0.020)	0.050(0.010)
	Model	12 items	40 items	100 items
RMSE	Model 1.1	0.140(0.120)	0.480(0.440)	0.540(0.430)
	Model 1.2	0.180(0.120)	0.480(0.440)	0.540(0.430)
	Model 1.3	0.290(0.230)	0.490(0.450)	0.550(0.430)
	Model 2.1	0.150(0.110)	0.480(0.440)	0.540(0.430)
	Model 2.2	0.220(0.140)	0.490(0.430)	0.540(0.440)
	Model 2.3	0.390(0.180)	0.540(0.430)	0.570(0.440)

According to these findings; it can be said that the fact that the two factors were correlated (Model-1) and all factors were correlated (Model-2) had almost the same effect in estimating the discrimination parameters. Consequently, there was no difference in the accuracy of parameter estimations for both models (Model 1 and Model 2). It can be concluded from this that the model parameter dis not have an influence on the accuracy of the parameter estimation.

When the RMSE values were examined according to the test length, it was considered that the test length might have an influence on the accuracy of the parameter estimation. The average RMSE values increased as the test length increased ($\overline{X}_{RMSE\&12}=$ 0.228, $\overline{X}_{RMSE\&40}=$ 0.493, $\overline{X}_{RMSE\&100}=$ 0.546). Namely, as the number of items increased, the accuracy of the discrimination parameters decreased. When the standard errors were examined, it was observed that the standard error decreased as the test length increased ($\overline{X}_{SE\&12}=$ 0.145, $\overline{X}_{SE\&40}=$ 0.060, $\overline{X}_{SE\&100}=$ 0.050). The standard error is the standard deviation of the simulation samples, in other words, a distance measure. Because of this, the standard error is actually a measure of precision (Walther &Moore, 2005). In this case, it can be said that as the test length increased, the estimations of the discrimination parameters were more reliable, that is, the test lengths might influence the estimation accuracy of the discrimination parameters.

Estimation accuracy and stability of difficulty parameters. Table 4 shows the standard error averages and the average RMSE values for Model 1 and Model 2 of the difficulty parameters in an order. When the standard errors of the models were analyzed, it was observed that there was not much difference between the table values (average and standard deviation) ($\overline{X}_{SE\&Model\ 1}=$ 0.046 , $\overline{X}_{SE\&Model\ 2}=$ 0.047). When the RMSE averages were studied, it was concluded that the condition for the standard error was also observed here. Table values were the same for Model 1 and Model 2 ($\overline{X}_{RMSE\&Model\ 1}=$ 1.056 , $\overline{X}_{RMSE\&Model\ 2}=$ 1.056).

According to these findings, it can be said that in the estimation of the difficulty parameter, the fact that two specific factors were related and that all specific factors were related had almost the same influence. To put it in other words, it can be said that model type did not affect the difficulty of parameter estimation. Table 4 presents the standard error and the average RMSE values with difficulty coefficients for Model 1 and Model 2.

As it can be seen in Table 4, it was observed that the standard error averages of Model 1.1, Model 1.2, and Model 1.3 did not differ too much when the models were examined within themselves (according to the degree of the orthogonality violation) ($\overline{X}_{SE\&Model1.1}=$ 0.043 , $\overline{X}_{SE\&Model\ 1.2}=$ 0.046, $\overline{X}_{SE\&Model1.3}=$ 0.050). The same was observed for RMSE averages, too ($\overline{X}_{RMSE\&Model1.1}=$ 1.053 , $\overline{X}_{RMSE\&Model\ 1.2}=$ 1.060, $\overline{X}_{RMSE\&Model1.3}=$ 1.056). When the model was examined for sub-models, it was observed that the standard error averages of Model 2.1, Model 2.2, Model 2.3 did not vary much ($\overline{X}_{SE\&Model\ 2.1}=$ 0.046 , $\overline{X}_{SE\&Model\ 2.2}=$ 0.046, $\overline{X}_{SE\&Model\ 2.3}=$ 0.050). The same was observed for RMSE averages, too ($\overline{X}_{RMSE\&Model\ 2.1}=$ 1.053 , $\overline{X}_{RMS\&EModel\ 2.2}=$ 1.060, $\overline{X}_{RMSE\&Model\ 2.3}=$ 1.056). The level of the orthogonality violation did not affect the estimation of the difficulty parameters. This finding overlaps with the study of Zheng (2013).

Table 4

Difficulty Coefficients Standard Error and Mean RMSE Values for Model 1 and Model 2

		Test Lengths		
	Model	12 items	40 items	100 items
	Model 1.1	0.050(0.030)	0.040(0.010)	0.040(0.010)
	Model 1.2	0.060(0.040)	0.040(0.010)	0.040(0.010)
SE	Model 1.3	0.070(0.050)	0.040(0.010)	0.040(0.010)
	Model 2.1	0.060(0.040)	0.040(0.010)	0.040(0.010)
	Model 2.2	0.060(0.050)	0.040(0.010)	0.040(0.010)
	Model 2.3	0.070(0.050)	0.040(0.010)	0.040(0.010)
	Model	12 items	40 items	100 items
	Model 1.1	0.330(0.260)	1.400(1.060)	1.430(1.130)
	Model 1.2	0.330(0.260)	1.420(1.060)	1.430(1.140)
RMSE	Model 1.3	0.330(0.270)	1.410(1.060)	1.430(1.140)
	Model 2.1	0.330(0.250)	1.400(1.070)	1.430(1.130)
	Model 2.2	0.340(0.260)	1.410(1.060)	1.430(1.130)
	Model 2.3	0.330(0.260)	1.410(1.060)	1.430(1.140)

When the RMSE values were examined according to the test length, it was observed that when the test length increased from 12 to 40, the estimation accuracy decreased, but when it increased from 40 to 100, this situation did not vary much ($\overline{X}_{RMSE\&12}$= 0.616, $\overline{X}_{RMSE\&40}$= 1.408, $\overline{X}_{RMSE\&100}$= 1.430). Contrary to study of Zheng (2013), when the accuracy of the estimations among the parameters in the framework of this study was taken into consideration, the difficulty parameters were the lowest parameters in terms of the test length and the model type. As Jennrich and Bentler (2012) pointed out in their research, when the correlation between factors was allowed, the results couldn't be interpreted. As the test length increased, the standard error values increased ($\overline{X}_{SE\&12}$= 0.310, $\overline{X}_{SE\&40}$= 0.493, $\overline{X}_{RMSE\&100}$= 0.546). The increase in the standard error indicated that the estimation accuracy decreased as the test length increased.

The estimation accuracy and stability of person parameters. Table 5 shows the standard error and RMSE averages for Model 1 and Model 2 of person parameters in an order. When the standard error values of the models were studied, it was observed that the table values (average and standard deviations) were similar ($\overline{X}_{SE\&Model\ 1}$= 0.464, $\overline{X}_{SE\&Model\ 2}$= 0.400). Table 5 shows the standard error and RMSE values for Model 1 and Model 2 of the person parameters.

Table 5

Standard Error and RMSE Values for Model 1 and Model 2 of The Person Parameters

	Model	Test Length		
		12 items	40 items	100 items
SE	Model 1.1	0.450(0.080)	0.490(0.130)	0.490(0.170)
	Model 1.2	0.450(0.080)	0.480(0.130)	0.500(0.170)
	Model 1.3	0.450(0.080)	0.480(0.130)	0.490(0.170)
	Model 2.1	0.420(0.080)	0.420(0.080)	0.360(0.070)
	Model 2.2	0.410(0.070)	0.420(0.090)	0.370(0.070)
	Model 2.3	0.380(0.070)	0.420(0.100)	0.390(0.100)
	Model	12 items	40 items	100 items
RMSE	Model 1.1	0.750(0.320)	0.620(0.210)	0.580(0.200)
	Model 1.2	0.760(0.340)	0.630(0.220)	0.590(0.200)
	Model 1.3	0.770(0.350)	0.640(0.230)	0.600(0.210)
	Model 2.1	0.740(0.340)	0.560(0.200)	0.450(0.140)
	Model 2.2	0.750(0.360)	0.590(0.220)	0.470(0.160)
	Model 2.3	0.790(0.410)	0.650(0.270)	0.540(0.210)

As it can be seen in Table 5, when the RMSE averages were to be examined, the situation for the standard error also appeared here. Table values for Model 1 and Model 2 were almost the same ($\overline{X}_{RMSE\&Model\ 1}$ = 0.660 , $\overline{X}_{RMSE\&Model\ 2}$ = 0.615). As a result, it can be said that the fact that two factors were correlated and that all specific factors were correlated had almost the same effect in parameter estimations.

When the models were analyzed within themselves (according to the degree of the orthogonality violation), it was observed that the standard error averages of Model 1.1, Model 1.2, Model 1.3 did not vary much ($\overline{X}_{SE\&Model1.1}$ = 0.477 , $\overline{X}_{SE\&Model\ 1.2}$ = 0.477, $\overline{X}_{SE\&Model1.3}$ = 0.473). The same was also observed for RMSE averages ($\overline{X}_{RMSE\&Model1.1}$ = 0.650 , $\overline{X}_{RMSE\&Model\ 1.2}$ = 0.660, $\overline{X}_{RMSE\&Model1.3}$ = 0.670). When the model 2 was examined for sub-models, it was seen that the standard error averages of Model 2.1, Model 2.2, Model 2.3 did not vary much ($\overline{X}_{SE\&Model\ 2.1}$ = 0.400 , $\overline{X}_{SE\&Model\ 2.2}$ = 0.400, $\overline{X}_{SE\&Model\ 2.3}$ = 0.397). The same was also observed for RMSE averages ($\overline{X}_{RMSE\&Model\ 2.1}$ = 0.583 , $\overline{X}_{RMSE\&Model\ 2.2}$ = 0.603, $\overline{X}_{RMSE\&Model\ 2.3}$ = 0.660).

Discussion, Conclusion and Recommendations

This research aims to analyze the effect of the Bifactor item response theory on the item and person parameter estimation under various conditions of the orthogonality assumption violation. As a result of the analyses made for this purpose, the estimation bias of the discrimination parameters for Model 1 increased as the orthogonality violation increased. The increase in test length caused a decrease in the accuracy of the

discrimination and difficulty parameters, in other words the reliability. This can be explained by the increase in the number of correlated items in specific factors. In the estimations of the discrimination parameters, an improvement in parameter estimations was observed with regard to the test length when two factors were related (Model 1), whereas this improvement was not observed when all specific factors were related (Model 2). The parameters whose estimation accuracy was the lowest were the difficulty parameters. It was observed that the model did not have an effect on the estimation accuracy of discrimination, difficulty, and person parameters. To put it in a different way, the case that two factors were correlated (Model 1) and that all specific factors were correlated (Model 2) had the same effect on the accuracy of both the person and item parameters.

Increasing the number of items increased the reliability of the estimations of person parameters. This situation observed in the person parameters was a consequence of the better explanation of the latent trait of individuals as the number of items increased. In estimations of person parameters, the least reliable parameter estimations were at the smallest test length for both models (Model 1 and Model 2). As the test length increased, the reliability of the estimations increased, too. Despite this, among the other parameters, the person parameters whose estimation reliability was the lowest at the all test lengths and the orthogonality violation levels.

The estimation of item and person parameters is an important factor in psychological and educational evaluations. The use of the Bifactor model in correlated structures will lead to biased parameter estimations, and this bias in parameter estimations will lead to bias in evaluation. The researches in the literature suggest that the Bifactor model is a very robust model that is well adapted even to the correlated structures. However, in this research when the parameter bias was examined, this robust structure could not be seen at all.

Based on the results of this study, some suggestions can be made for the practitioners or the researchers in the application of the Bifactor model. As test length increases, the Bifactor theory can better tolerate the orthogonality violation in estimation of person parameters. The practitioners who want to use this theory are recommended to work with large item pools. At all correlation levels, the accuracy of the parameter estimations was approximately the same. New studies can be repeated with intermediate correlation levels (0.25, 0.35, etc.). It is stated in the literature that there must be at least 20 items for multidimensional item response theory models. The minimum test length in this study was determined as 12 items. To obtain more unbiased results in the estimation of item parameters, determining the minimum test length as 20 items in future studies can retry the same conditions. Among all the parameters, the parameters whose estimation reliability is the lowest (highest SE averages) were found to be person parameters. Future researches can be tested with different replication numbers and different sample sizes to increase the reliability. This is only a simulation study, and is valid for the specified conditions.

References

Brouwer, D., Meijer, R. R., Weekers, A. M., & Baneke, J. J. (2008). On the dimensionality of the Dispositional Hope Scale. *Psychological Assessment*, *20*(3), 310.

Brown, A. R., Finney, S. J., & France, M. K. (2011). Using the bifactor model to assess the dimensionality of the Hong Psychological Reactance Scale. *Educational and Psychological Measurement*, *71*(1), 170-185.

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods,* 16(3), 221–248.

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. *Principles and Methods of Test Construction: Standards and Recent Advancements*. Gottingen, Germany: Hogrefe Publishers.

Chalmers, P. (2016). Mirt: Multidimensional item response theory. R package version 1.19,URL: https://cran.r-project.org/web/packages/mirt/index.html

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189-225.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219-251.

Cohen, J. (1988). *Statistical power analysis for the behavioral science (2nd ed.)*. Hillside, NJ: L. Erlbaum Associates.

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence*, *5*(3), 27.

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

DeMars, C. E. (2013). A tutorial on interpreting Bifactor model scores. *International Journal of Testing*, *13*(4), 354-378.

Fukuhara, H. (2009). *A Differential Item Functioning Model for Testlet-Based Items Using A Bi-Factor Multidimensional Item Response Theory Model: A Bayesian Approach*. The Florida State University.

Garn, A. C. (2017). Multidimensional measurement of situational interest in physical education: Application of Bifactor exploratory structural equation modeling. *Journal of Teaching in Physical Education*, *36*(3), 323-339.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information Item Bi-factor analysis. *Psychometrika*, *57*(3), 423-436.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., ... & Stover, A. (2007). Full-Information item bifactor analysis of graded response

data. *Applied Psychological Measurement*, *31*(1), 4-19.

Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the french wechsler adult intelligence scale (WAIS-III). *Psychological Assessment*, *23*(1), 143.

Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., & Egan, A. (2014). A Bifactor approach to modelling the Rosenberg Self Esteem Scale. *Personality and Individual Differences*, *66*, 188-192.

Holzinger, K. J., & Swineford, F. (1937). The Bi-Factor method. *Psychometrika*, *2*(1), 41-54.

Immekus, J. C., & Imbrie, P. K. (2008). Dimensionality assessment using the full-information item Bifactor analysis for graded response data: An illustration with The State Metacognitive Inventory. *Educational and Psychological Measurement*, *68*(4), 695-709.

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory Bi-Factor analysis: The oblique case. *Psychometrika*, *77*(3), 442-454.

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

LaFond, L. J. (2014). *Decision consistency and accuracy indices for the Bifactor and Testlet response theory models*. The University of Iowa. UMI Number: 3638391.

Li, Y., & Rupp, A. A. (2011). Performance of The S−X2 Statistic for full-information Bifactor models. *Educational and Psychological Measurement*, *71*(6), 986-1005.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, *4*(1), 84.

Martel, M. M., Von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: is there A 'G' factor?. *Journal of Child Psychology and Psychiatry*, *51*(8), 905-914.

Reise, S. P. (2012). The rediscovery of Bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data Yield Univocal Scale Scores. *Journal of personality assessment*, *92*(6), 544-559.

Reise, S. P., Ventura, J., Keefe, R. S., Baade, L. E., Gold, J. M., Green, M. F., ... & Bilder, R. (2011). Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in Schizophrenia. *Psychological Assessment*, *23*(1), 245.

Revelle, W. (2017). Psych: Procedures for psychological, psychometric, and personality research. R package version 1.7.5. URL: https://cran.rproject.org/web/packages/psych/index.html

Rijmen, F. (2009). Three multidimensional models for testlet-based tests: Formal relations and an empirical comparison. Research Report. ETS RR-09-37. *Educational Testing Service*.

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*(1), 51-67.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237.

Simms, L. J., Grös, D. F., Watson, D., & O'hara, M. W. (2008). Parsing the general and specific components of depression and anxiety with Bifactor modeling. *Depression and Anxiety*, *25*(7).

Spearman, C. E. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201–293.

Stucky, B. D., Thissen, D., & Orlando Edelen, M. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, *37*(1), 41-57.

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, 183-206.

Stucky, B. D., Edelen, M. O., Vaughan, C. A., Tucker, J. S., & Butler, J. (2014). The psychometric development and initial validation of the DCI-A short form for adolescent therapeutic community treatment process. *Journal of Substance Abuse Treatment*, *46*(4), 516-521.

Thomas, M. L. (2012). Rewards of bridging the divide between measurement and Clinical Theory: Demonstration of a Bifactor model for the brief symptom inventory. *Psychological Assessment*, *24*(1), 101.

Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, *28*(6), 815-829.

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition. *School Psychology Quarterly*, *29*(1), 52.

Yang, Y., Song, L., & Xu, T. (2002). Robust estimator for correlated observations based on Bifactor equivalent weights. *Journal of Geodesy*, *76*(6-7), 353-358.

Zhang, B. (2008). Application of unidimensional item response models to tests with items sensitive to secondary dimensions. *The Journal of Experimental Education*, *77*(2), 147-166.

Zheng, C. (2013). *Examination of the parameter estimate bias when violating the orthogonality assumption of the Bifactor model* (Doctoral dissertation). University of Kansas.

İki Faktör Madde Tepki Kuramında Diklik Varsayımının İncelenmesi

Özet

Problem Durumu: İki Faktör Modeli, çok boyutlu madde tepki kuramı (multidimensional item response theory) modellerinden biridir. İki faktör modeline göre birden fazla spesifik (özgül) faktör ve bu faktörler tarafından açıklanan bir genel faktör vardır ve ayrıca bu özgül etkilerin genel faktör üzerinde etkisinin olduğunu varsayılmaktadır. Tüm madde tepki kuramı modellerinde olduğu gibi İki Faktör modelinin de kendine özgü varsayımları vardır. İki Faktör Model'inin en önemli varsayımlarından biri verinin hem genel faktörü hem de spesifik faktörleri içermesidir. Bu varsayım karşılanması zor bir varsayım olmamakla birlikte çok boyutlu veriyi gerektirmektedir. Diğer varsayım olan faktörlerin dik (orthogonal) yani birbirinden bağımsız (ilişkisiz) olması ise pratikte karşılanması çok mümkün olmayan bir varsayımdır. İlişkili faktörleri dik olmaya zorlamak ise ölçülen yapı ile ilgili olarak bilgi kaybına neden olacak ve güvenilir olmayan parametre kestirimleri ile sonuçlanacaktır. Bu çalışma aracılığıyla İki Faktör Modelin kullanımını kısıtlayan varsayımın incelenmesi ve belirli kriterler ışığında değerlendirilmesi ile elde edilecek sonuçların alan yazına hem teorik anlamda hem de modelin daha doğru uygulanabilirliği açısından önemli katkılar sağlayacağı düşünülmektedir.

Araştırmanın Amacı: İki Faktör Kuramı, gerektirdiği varsayımdan (diklik) dolayı kullanımı sınırlanan bir kuramdır. Bu sınırlılığının yanı sıra psikolojik ve eğitsel yapıların modellenmesinde ve ölçek geliştirme çalışmalarında bu varsayım göz ardı edilerek sıklıkla kullanılmaktadır. Diklik varsayımının sağlanmadığı koşullarda psikolojik ve eğitsel yapıların doğru modellenmesi, geliştirilen ölçeğin doğru faktör yapısına ulaşması ve parametre kestirimlerinin doğru olması mümkün olmayacaktır. Bunun yanı sıra eğitim ve psikoloji alanında faktörler arası korelasyonun sıfır olduğu ölçme araçları geliştirmek neredeyse imkansızdır. İlişkili faktörleri dik olmaya zorlamak ise ölçülen yapı ile ilgili olarak bilgi kaybına neden olacak ve güvenilir olmayan parametre kestirimleri ile sonuçlanacaktır. Parametre kestirimlerinin kesinliği ve doğruluğu ise yapılan her ölçme işleminde önemli bir durumdur. Çünkü parametre kestirimleri, madde performansı ve yanıtlayıcı yetenek düzeyinin belirlenmesinde önemli bir unsurdur. Belirtilen bu gerekçelerden kaynaklı, iki faktör kuramının, spesifik faktörler arası farklı ilişki düzeylerine olanak tanıyarak incelenmesi yani hangi diklik ihlal düzeylerinin kuram tarafından tolere edilip, diklik ihlaline rağmen kararlı, kesin ve doğru kestirimler yapılabildiğinin belirlenmesi bu araştırmanın amacıdır.

Araştırmanın Yöntemi: Bu araştırma için veriler simülatif yolla iki adet (Model-1 ve Model-2) İki Faktör iki parametreli modele göre üretilmiştir. Model-1 iki spesifik faktör arasında çapraz yüklenmelerden dolayı oluşan diklik ihlalini gösteren modeldir. Burada incelenen nokta iki spesifik faktör arasındaki diklik ihlalinin tüm faktörlerdeki parametre kestirimlerine olan etkisidir. Model-2 ise, tüm spesifik faktörler arasındaki ilişkiyi göstermektedir. Spesifik modellerde manipüle edilen değişkenler faktörler arası korelasyon düzeyleri ve test uzunluklarıdır. Kurulan modeller için korelasyon kabul düzeyleri 0.10 (çok düşük), 0.40 (orta), 0.70 (yüksek) olarak ele alınmıştır. Spesifik faktörlerdeki madde sayılarına karar vermek amacıyla yapılan alan yazın incelemesi sonucunda test uzunlukları 12, 40 ve 100 madde olarak belirlenmiştir. Araştırma boyunca sabit tutulacak (manipüle edilmeyecek) değişken ise örneklem (5000) büyüklüğüdür. Replikasyon sayısı ise 200 olarak belirlenmiştir. Parametre kestirimlerinin replikasyonlar boyunca doğruluğunun değerlendirilmesi; ortalama yanlılık (mean bias), RMSE (hataların kareleri ortalamasının karekökü) ve kestirimlerin standart hatası (Standart Error) ile yapılmıştır.

Araştırmanın Bulguları: Ayırt edicilik parametreleri için tüm test uzunluklarında Model 1 ve Model 2 için görülen örüntü aynı şekildedir. Madde sayısındaki artış ayırt edicilik parametrelerinin kestirim kesinliğinde yani güvenirliğinde düşüşe neden olmuştur. Bu durum yanlı madde miktarındaki artış ile açıklanabilir. Yani modele ne kadar ilişkili madde eklenirse değişkenlik o kadar artmıştır. İki faktörün ilişkili olması durumu (Model-1) ile tüm faktörlerin ilişkili olması durumunun (Model-2), ayırt edicilik parametrelerinin kestiriminde neredeyse aynı etkiye sahip olduğu söylenebilir. Sonuç olarak her iki model için de parametre kestirim doğruluğu arasında farklılık yoktur. Buradan yola çıkarak model türünün parametre kestirim doğruluğuna etkisi olmadığı söylenebilir. Güçlük parametresinin kestiriminde, iki spesifik faktörün ilişkili olma durumu (Model 1) ile tüm spesifik faktörlerin ilişkili olma durumunun (Model 2) neredeyse aynı etkiye sahip olduğu söylenebilir. Yani model türünün güçlük parametre kestirim doğruluğuna etkisi olmadığı söylenebilir. Birey parametreleri incelendiğinde, test uzunluğu ile doğru orantılı şekilde değişkenliğin azalması test uzunluğunun parametre iyileşmesinde etkisi olabileceğine işaret etmektedir. Yine de değişkenlik tüm test uzunluklarında yüksektir. Bu durum parametre kestirim güvenirliklerini düşürmektedir. Birey parametrelerinin kestiriminde, iki spesifik faktörün ilişkili olma durumu ile tüm spesifik faktörlerin ilişkili olma durumunun neredeyse aynı etkiye sahip olduğu söylenebilir.

Araştırmanın Sonuçları ve Önerileri: Kestirim doğruluğu en düşük parametrelerin güçlük parametreleri olduğu görülmüştür. Ayırt edicilik, güçlük ve birey parametrelerinin kestirim doğruluğunda ise modelin öneminin olmadığı görülmüştür. Yani iki spesifik faktörün ilişkili olma durumu (Model 1) ile tüm faktörlerin ilişkili olma durumu (Model 2) hem birey hem de madde parametrelerinin kestirim doğruluğunda aynı etkiye sahiptir. Madde sayısını arttırmak, birey parametrelerinin kestirim kesinliğini yani güvenirliğini arttırmıştır. Birey parametrelerinde gözlenen bu durum, madde sayısı arttıkça bireyin örtük özelliğinin daha iyi açıklandığının bir sonucudur. Birey parametrelerinin kestiriminde, güvenirliği en düşük parametre kestirimleri her iki model için de (Model 1ve Model 2) en küçük test uzunluğundadır.

Test uzunluğu arttıkça kestirim güvenirliği de artmıştır. Buna rağmen tüm test uzunluklarında ve diklik ihlal düzeylerinde kestirim güvenirliği en düşük parametreler birey parametreleridir. Madde ve birey parametrelerinin kestirimi psikolojik ve eğitsel amaçlı değerlendirmelerde önemli bir unsurdur. İki faktör kuramının ilişkili yapılarda kullanılması yanlı parametre kestirimlerine, parametre kestirimlerindeki yanlılık ise değerlendirme sonuçlarında yanlılığı doğuracaktır. Literatürde varolan araştırmalar iki faktör kuramının ilişkili yapılarda bile çok iyi düzeyde uyum verdiği ve robust bir model olduğu belirtmektedir. Bu araştırmada ise parametre bazında yanlılık incelendiğinde bu robust yapı görülememiştir. İki faktör kuramı, birey parametrelerinin kestiriminde test uzunluğu arttıkça diklik varsayımı ihlalini daha iyi tolere edebilmektedir. Bu kuramı kullanmak isteyen uygulayıcıların büyük madde havuzları ile çalışmaları önerilir. Tüm korelasyon düzeylerinde parametre kestirim doğrulukları yaklaşık olarak aynı çıkmıştır. Yeni çalışmalar ara korelasyon (0.25, 0.35 vb.) düzeyleri ile tekrarlanabilir.

Anahtar Kelimeler: Çok boyutlu madde tepki kuramı, İki faktör Madde Tepki Kuramı, diklik varsayımı, parametre kestirim yanlılığı, faktör analizi.

The Effect of Teaching "Learning Strategies" on Academic Achievement: A Meta-Analysis Study*

Ibrahim YILDIRIM[1], Sevilay CIRAK-KURT[2], Sedat SEN[3]

A R T I C L E I N F O	A B S T R A C T
	Purpose: This study aimed to perform a meta-analysis of results from the experimental studies that examined the effect of learning strategies on students' academic achievement in Turkey between 2000 and 2016. **Method:** This study employed a meta-analysis method to combine the results of experimental studies on the effect of teaching learning strategies on students' academic achievement. Twenty-eight studies comprising of theses and peer-reviewed articles involving an experimental analysis of the effect of learning strategies on students' achievement were included in this study. **Results:** The random effects model was used in the

study and the mean effect size value was found at large effect level (d = 1.21). The eta-squared calculated for Cohen's d of 1.21 was found to be .268. This value obtained using 31 effect sizes from the results of 28 experimental studies conducted with 1,641 students indicated that the learning strategies had 26.8% positive effect on students' academic achievement.

Implications for Research and Practice: Sub-group analyses were also made in the sub-categories (education level, discipline, strategy teaching style, and strategy type) within the scope of the research, and it was determined that the teaching of learning strategies differed only statistically according to the discipline area. Several suggestions based on the research results were presented.

---

[1] Harran University, TURKEY. iyildirim84@gmail.com ORCID: 0000-0002-4137-2025

[2] Adiyaman University, TURKEY. sevilaycirak@hotmail.com ORCID: 0000-0001-8951-8727

[3] Harran University, TURKEY. sedatsen06@gmail.com ORCID: 0000-0001-6962-4960

Introduction

"Tomorrow's illiterate will not be the man who can't read; he will be the man who has not learned how to learn."

(Toffler, 1974)

Developments in cognitive psychology, science and technology affected educational settings together with the components included, just like they affected various other environments. In this sense, changes in curricula are observed to be in parallel with the developments. The differences between the curricula adopted in the early years of the Republic Period (in Turkey) and those adopted in the final years reflect these developments. With the developments, the point of view towards individuals has changed in educational settings. Individuals are no longer regarded as passive receivers of knowledge, but active participants in its creation. Moreover, knowledge is updated rapidly, becoming too much and too variable to be conveyed to individuals. In this regard, there is a need for training individuals who can adapt to the necessities of time; know themselves; in other words, continuously update their knowledge; take responsibility for their learning; know how to learn; and are involved in life-long learning. At this point, we can see the learning strategies that can facilitate the act of learning. According to this need, common skills (critical thinking, creative thinking, communication skills, inquiry-based skills, problem solving skills, ability to use information technologies, entrepreneurial skills, and Turkish language skills), which are included in the backbone of all programs (communication in the mother tongue, mathematical competence, digital competence, competence related to social citizenship, taking the initiative and entrepreneurship perception, cultural awareness and expression), and learning to learn is also listed among these competencies with the amendment made in 2017 (Ministry of National Education [MoNE], 2017). Within the framework of competencies in Turkey; these competencies include awareness of the individual's learning needs, ability to cope with learning difficulties, insistence on one's own learning, seeking guidance and benefiting from it (MoNE, 2018). Indeed, a good teaching involves teaching students how to learn, remember, think, and motivate themselves as much as whatever they are taught (Weinstein & MacDonald, 1986). At this point, learning strategies that facilitate learning of anticipation are emerging, and the updates that are made suggest that learning strategies need to be emphasized more. Various researchers (Arends, 1997; Demirel, 2003; Sonmez, 2007; Weinstern & Mayer; 1986) reported that learning strategies, i.e., knowing how to learn in a better and easier way (Brandt,1988/1989), constitute the basis of independent learning (Weinstein & MacDonald, 1986) and learning how to learn. Demirel (2003) briefly defined learning strategies as the set of mental tactics used by individuals in a special learning setting to facilitate acquisition of knowledge and skills. To Arends (1997), learning strategies point out to behaviour and thinking processes that include cognitive strategies such as memorizing and recalling, and the cognitive processes directing these cognitive strategies that are used by learners and affect their learning. What is common in these definitions is the fact that individuals take active role in the learning process, and know how to acquire knowledge in a better and easier way, thus going through conscious mental processes.

In this study, subject of learning strategies consists of cognitive strategies according to Pintrich's classification. Self-regulated learning strategies are grouped in four categories: cognitive, metacognitive, resource management, and motivational strategies (Pintrich, 1999). However, there is no agreement reached by researchers on the classification of learning strategies. Although there are no great differences among the classifications, the learning strategies are classified in various ways (Gagne, 1988; Ozer, 2003; Senemoglu, 2001; Weinstein & Mayer, 1986). Some researchers classify learning strategies into five major categories as rehearsal, elaboration, organization, metacognition, and motivation, which are frequently referred to in the literature (Weinstein & Mayer, 1986). Some others adapted the strategies to certain disciplines, and proposed a classification of language learning strategies (O'Malley & Chamot, 1985; Oxford, 1990) or a classification of vocabulary learning strategies (Schmitt, 1997). Hereafter, a sort of classification with regard to learning strategies in one of the most frequently cited reference books will be detailed. The references can be reviewed for further information about the other classifications.

According to the classification of Weinstein and Mayer (1986), rehearsal strategies that have an important place in selecting and acquiring knowledge involve reiterating aloud a material presented to the class as it is, underlining important parts of a text and taking notes, reciting the causes of an incident, and allowing the delivery of information to working memory for more difficult tasks. Elaboration strategies, aiming at integrating previous knowledge with newer ones, ensure the transference of information from the working memory to the long-term memory, and include summarizing, interpreting, expressing the relationship between newer knowledge with the previous ones, creating simulations and taking productive notes. The main purposes of organization strategies used for complex tasks such as defining the main idea and the outline of a text or creating diagrams to reveal important details are the selection of information to be processed in the working memory, and the establishment of relationships between the ideas in this memory. Metacognition strategies control deficiencies in learning through incorporating the determination of the learners' own learning objectives for any learning activity by themselves, the identification of the extent of goal achievement and the change of the strategy if needed, and include students' questioning themselves about comprehension of the material presented to the class and using questions to direct learning at the beginning of a chapter. Affective strategies such as helping to cope with test anxiety, to be careful and relaxed also involve the use of thinking through working in a quiet place to reduce distracting external stimuli or not focusing on failure to avoid the anxiety of failure. The studies in this area focus on the strategies used by the learner to focus attention, keep motivation, performance anxiety and effective time management.

The studies conducted on learning strategies reported that teaching of such strategies affects different variables such as academic achievement (Biyikli &Dogan, 2015; Cross & Lipson, 1984; Caliskan & Sunbul, 2011; Kaydu, 2004; Paris, Ning & Downing, 2011; Yildiz, 2003), metacognitive awareness (Bas, 2012; Yokus, 2009), and metacognitive skills (Caliskan & Sunbul, 2011) positively, indicating that learning strategies can be taught (Dansereau et al., 1979; Demirel, 1993; Demirel, Askin-Tekkol,

Cigdem & Demir, 2016). Accordingly, there are two different approaches adopted in teaching these learning strategies: integrated (subject-based) and independent (general). It is still a question of debate whether learning strategies should be taught in a "subject-based" way or in a "general" way independent of the subject (Caliskan & Sunbul, 2011). Researchers arguing that learning strategies should be integrated into subject matter (Brandt, 1988/1989; Erdem, 2005; Levin, 1986), point out that the teaching of strategies differs based on the content and requirements of the subject matter. They argue that teaching strategies cannot be independent from the subject matter and should be a part of teaching the subject. Those arguing that strategies should be taught independently of the subject (Weinstein & Mayer, 1986) point out potential disruptions that may occur due to the differences in the knowledge and skills of teachers when strategies are integrated into the subject, arguing that strategy teaching should be provided within the scope of a general program. Some researchers combine these two approaches and propose an eclectic approach in which advantages of both approaches are combined (Lenz, 1992; Somuncuoglu & Yildirim, 1998). Indeed, a strategy taught in a course might be helpful in learning another subject matter, and students may explore its advantage by themselves. On the other hand, students may assess the learning strategies taught independently of the subject based on their mental process, and find out how they can use it.

Researchers indicate that the Turkish Ministry of National Education, schools and teachers play an important role in helping individuals learn the learning (Demirel, 1993; Erdem, 2005; Ozer, 2003; Somuncuoglu & Yildirim, 1998). They also make some recommendations and a list of required qualifications of the programs for learning strategies (Lenz, 1992; Paris, 1988; as cited in Ozer, 2004). For example, Weinstein et al. (1989) indicate that critical thinking must definitely be incorporated into the process of teaching learning strategies, emphasizing that being equipped with a comprehensive "repertoire of strategies" is important for learners to reach the maturity of being selective and flexible while using these strategies. In this sense, Somuncuoglu and Yildirim (1998) state that effective teaching of strategies should be addressed in a context of critical thinking/problem solving skills, and its relationship with specific learning activities or problems must be especially emphasized. Some researchers think that learning strategies should be a part of learners' life and adopted throughout the school, in other words, these strategies should be embedded into the school's culture in order to increase the efficiency of learners (Lenz, 1988; Ozer, 2003).

Although most of the studies on learning strategies are descriptive and correlational (Akkus, İspir, Ay, & Saygi, 2011; Bulus, Duru, Balkis, & Duru, 2011; Cesur & Fer, 2011; Ghee, İsmail, & Kabilan, 2010; Ning & Downing, 2011; Saracaloglu & Karasakaloglu, 2011; Tunca & Alkin-Sahin, 2014), there are still experimental studies. Experimental studies are usually focused on examining the effect of learning strategies on academic achievement (Kaydu, 2004; Mayer, 1980; Yildiz, 2003), attitudes (Belet, 2005; Dikbas & Hasirci, 2008), metacognitive awareness (Bas, 2012; Yokus, 2009), retention of learning (Bozkurt, 2007; Meydan, 2010), metacognitive skills (Caliskan & Sunbul, 2011), and most of them pointed out positive improvements in the related variables. The redundancy of research related to learning strategies has led the

researchers to consolidate them recently (for example: Ardasheva, Wang, Adesope, & Valentine, 2017; de Boer, H., Donker, A. S., & van der Werf, 2014; Demirel, Askin-Tekkol, Cigdem, & Demir, 2016; Donker, Boer, Kostons, van Ewijk, & van der Werf, 2013; Ergen & Kanadli, 2017; Hattie & Donoghue, 2016; Keskin, 2014). The two studies have been found to compile the postgraduate theses on learning strategies (Demirel, Askin-Tekkol, Cigdem, & Demir, 2016; Keskin, 2014) in Turkey. The research conducted in Turkey reveal the thematic and methodological state of postgraduate theses in the relevant field while it is observed that a great number of meta-analysis studies have been carried out abroad. There are various studies conducted in Turkey on the effect of teaching learning strategies on students' academic achievement in various courses. The studies on different dimensions of learning strategies need to be combined, synthesized and assessed. In this regard, through a meta-analysis, this study will make a significant contribution to the literature by determining the common directions through the combination of the results of the studies conducted in Turkey. To date, only one meta-analysis study on the effect of self-regulation strategies on academic achievement (Ergen & Kanadli, 2017) has been conducted in Turkey. Ergen and Kanadli's (2017) study included both relational and experimental studies and 11 studies in the context of cognitive strategies. In this context, the present study differs from Ergen and Kanadli's (2017) study, including experimental studies involving teaching of cognitive learning strategies. The limitation of the study stems from the research objective to reveal the effects of teaching learning strategies in the context of Turkey. As a matter of fact, it would be appropriate to limit the scope of the research to Turkey in academic sense as the national culture and approaches to learning strategies would be different from other countries. This study aimed to perform a meta-analysis of results from the experimental and quasi-experimental studies that examined the effect of learning strategies on students' academic achievement in Turkey between 2000 and 2016. In this sense, answers were sought to the following research questions:

1. How effective is teaching of learning strategies on students' academic achievement according to the findings of the experimental studies conducted in Turkey between 2000 and 2016?

2. Do the findings of the experimental studies conducted in Turkey between 2000 and 2016 on the effect of teaching learning strategies on students' academic achievement differ significantly according to different variables (education level, and discipline)?

Trying to seek answers to these two research questions, this study is important in raising awareness among educational policy makers, program developers and teachers about how to teach learning strategies, and in combining and synthesizing the related studies in a systematic way, thus showing the big picture and serving as a guide to the learning processes. Moreover, the research will give instructive clues related to national teaching processes in terms of revealing the general opinion about Turkey.

Method

Increased number of primary studies on specific issues also increased attention to meta-analysis studies which are based on the synthesis of empirical studies that combine different and similar findings of such studies, and are more comprehensive, practical and resistive to the limitations of the studies (Ustun & Eryilmaz, 2014). At this point, "meta-analysis" can be seen as one of the research syntheses that was defined by Borenstein, Hedges, Higgins and Rothstein (2009) as a combination of statistical findings of quantitative studies in a systematic way. On the other hand, Cooper, Hedges and Valentine (2009) indicate that meta-analysis is different from research synthesis and has a special structure. Meta-analysis, by its very nature, provides a combination of a large amount of quantitative findings in a consistent and coherent way by taking account of the effect sizes, and aims to derive meaningful generalizations by analyzing these findings in an organized way (Cohen, Manion & Morrison, 2001). This study employed the meta-analysis method to combine the results of experimental studies on the effect of teaching learning strategies on students' academic achievement. After formulating a research question, six steps can be followed in a typical meta-analysis (Field & Gillett, 2010, p. 666): 1) doing a literature search; 2) deciding the inclusion and exclusion criteria and apply them; 3) calculation of effect sizes for each study to be included; 4) doing the basic meta-analysis; 5) considering some more advanced analyses such as publication bias analysis and exploring moderator variables; and 6) writing up the results. These aforementioned steps were followed in this study.

Publication bias emerges as an important problem in meta-analysis studies. Rothstein, Sutton and Borenstein (2006) indicate that publication bias occurs when only the findings of published studies are included in the meta-analysis or due to the tendency of academic magazines to publish the studies that are significantly different from the others. Therefore, incorporating as many and various studies as possible in the meta-analysis can be used as a precaution against publication bias and would provide the opportunity to obtain more valid and reliable results. Other factors affecting the validity of meta-analyses are; the fact that the measuring instruments used in the primary studies do not have sufficient reliability and validity, the calculation of the effect size is not explained and the study is entered into an independent research process from the area where the work is done (Greco, Zangrillo, Biondi-Zoccai, & Landoni, 2013).

Data Collection

Several online databases including Google Scholar, ULAKBİM and National Thesis Center were screened to find the published articles, full texts of papers and unpublished theses. During the search, the key word template of "ogrenme stratejileri" OR "learning strategy" AND "deneysel OR experimental" AND "basari OR achievement" for Turkish pages in Google Scholar were used and 1,060 studies were found in this way. One hundred seventy-six publications were found in the (Turkey) National Thesis Centre, and 205 publications were found in ULAKBİM database. In total, 1,436 publications in Turkey were reached either published or unpublished,

either in Turkish or English. To this end, we completed the searching process in January 2017, and included the studies conducted in Turkey throughout a period of 17 years between the years of 2000 to 2016.

The criteria used to select the studies were as follows:

1. The study should be conducted in Turkey between the years of 2000-2016;

2. The study should be in either Turkish or English, and in the form of a thesis, an article or a full text paper;

3. The study should examine the effect of learning strategies on students' achievement experimentally or quasi-experimentally;

4. The study should provide sufficient statistical data (N, $\bar{x}$, SD etc.) to allow for the calculation of effect sizes.

5. The measurement tool used in the study should have sufficient psychometric properties (i.e., validity and reliability).

As a result of the classification of the total of 1,436 publications reached by the last search on the 20th of January 2017 according to aforementioned criteria, this study concentrated on the remaining 352 studies. An e-mail was sent to the authors of the studies that met the selection criteria but were not accessible; however, the authors did not respond to e-mails. Twenty-eight studies comprising of theses and peer-reviewed articles involving an experimental analysis of the effect of learning strategies on students' achievement were included in this study. Thirty-one effect sizes were calculated in total, and analyses were conducted with these 31 effect sizes as there was more than one experimental group in three of the studies. Apart from these 28 studies, no other studies were found to provide sufficient data and meet the necessary requirements in the form of a full-text paper or poster presentation.

Figure 1. Flow Chart Showing Selection of Studies for Meta-Analysis

A power analysis was conducted using the power analysis equation (Equation 14) for the random effects model presented by Valentine, Pigott, and Rothstein (2010, p.224). Statistical power of this study was found to be .99 using the following values: medium effect size (ES = .5), average sample per group = 25, the total effect size = 31,

and the medium heterogeneity. This shows that the magnitude of 31 effects included in the study gives a high statistical power.

Descriptive Information about the Studies

This study included a meta-analysis of the experimental studies examining the effect of teaching learning strategies on students' academic achievement. Table 1 shows the descriptive information about these studies conducted in Turkey.

Table 1

Descriptive Information about the Studies

Study	Level of Education	Discipline	Type	N
Akin, 2013	High school	Music	Article	10
Aydemir, 2007	High school	Foreign Lang.	Thesis	34
Bas, 2012	Secondary school	Foreign Lang.	Article	60
Belet, 2005	Elementary sch.	Turkish	Thesis	43
Biyikli, 2015	Elementary sch.	Science	Article	135
Bolukbas, 2013	University	Foreign Lang.	Article	40
Bozkurt, 2007a	High school	Social Sciences	Thesis	52
Bozkurt, 2007b	High school	Social Sciences	Thesis	54
Caliskan, 2011	Secondary school	Turkish	Article	42
Cerci, 2005	Secondary school	Turkish	Thesis	44
Cetingoz, 2006	University	Social Sciences	Thesis	42
Derman, 2002a	Secondary school	Science	Thesis	103
Derman, 2002b	Secondary school	Science	Thesis	110
Dikbas, 2008	Elementary sch.	Social Sciences	Article	24
Evcim, 2008	University	Foreign Lang.	Thesis	46
Hasra, 2007	High school	Foreign Lang.	Thesis	60
Kaya, 2006	Elementary sch.	Turkish	Thesis	40
Kaydu, 2004	High school	Social Sciences	Thesis	52
Keban, 2010	University	Science	Thesis	39
Kuleli, 2011	University	Foreign Lang.	Thesis	50
Kurum, 2012	University	Foreign Lang.	Thesis	55
Meydan, 2010	Elementary sch.	Social Sciences	Article	68
Serce, 2013	University	Foreign Lang.	Thesis	42
Tasdemir, 2007	University	Science	Article	52
Tok, 2008a	Elementary sch.	Social Sciences	Article	82
Tok, 2008b	Elementary sch.	Social Sciences	Article	81
Tuncer, 2007	Elementary sch.	Social Sciences	Thesis	40
Uraz, 2004	University	Foreign Lang.	Thesis	30
Uysal, 2006	Secondary school	Social Sciences	Thesis	47
Yildiz, 2003	Elementary sch.	Science	Thesis	44
Yokus, 2009	University	Music	Thesis	20

It can be observed that most of the studies were conducted at the university level and in the field of social sciences between 2005 and 2009. Most of them were also found to be in the form of thesis and they were conducted with a total of 1,641 students.

Data Coding

In order to interpret the outcomes of a meta-analysis correctly, it must be ensured that the data are coded correctly (Lipsey & Wilson, 2001). Therefore, all data were coded by two researchers in the field of Educational Sciences and with meta-analysis experience separately. The encoders were fully consistent with each other (*r*=1.00).

Twenty-eight studies were included in the study. Since there were two experimental groups from different samples in three of these studies, a total of 31 effect sizes (Cohen *d*) were calculated and the analysis was performed on 31 different effect sizes. In order to control for the methodological features of the 28 studies, a coding sheet (Study Design and Implementation Assessment Device) by Valentine and Cooper (2008) was used to judge the quality of the studies.

In a meta-analysis, in addition to calculating a common effect size, it would be better to collect data for sub-groups to determine situations where effect sizes differ (Lipsey & Wilson, 2001). Therefore, we also reported some statistical information about the studies included as well as the authors' name, publication years, publication form (article, postgraduate thesis, paper), education stage at which the study was carried out (elementary, secondary or high school, and university) and the field of study (foreign language, social sciences, Turkish, and music), the type of strategy used (repetition, sense-making, organizing, monitoring the comprehension, affective, integrated, foreign language learning, and foreign language vocabulary learning).

In the literature, there is still a debate on whether learning strategies should be independent from the subject matter or should be integrated into it (Caliskan & Sunbul, 2011; Somuncuoglu & Yildirim, 1998). The studies included in the meta-analysis were also examined in the terms of this debate, and three additional sub-groups were formed (Somuncuoglu & Yildirim, 1998). These sub-groups were as follows:

1) Studies in which students were first informed of the learning strategies and how to use them, and then the regular teaching process was carried out;

2) Studies in which students were provided with activities instead of any training on learning strategies;

3) Studies in which both approaches given in the first item and were combined with the second (Eclectic: First the learning strategies are explained, and then activities based on these strategies are carried out).

In this study, we also conducted sub-group analyses in addition to calculating a common effect size of the studies to determine the situations in which effect sizes may differ.

Data Analysis

In the data analysis, it should be first decided which effect size value will be used. In correlational studies, correlations are included into the meta-analysis after being converted to the Fisher's *z* score. In other studies, based on differences, the values are

converted to Cohen's *d*, Hedges' *g* or Glass' Δ effect size values (Borenstein et al., 2009). Among these effect sizes, Cohen's *d* is more suitable for studies with a total sample size above 20 (Lipsey & Wilson, 2001). Only one of the 31 effect sizes included in this study had a sample size below 20. Therefore, the values obtained from the experimental studies were converted to Cohen's *d,* which reflected the difference between the means and analyzed through the Comprehensive Meta-Analysis (CMA v.2; Borenstein, Hedges, Higgins, & Rothstein, 2005) software. The Cohen's *d* formula used in this study was obtained by dividing the mean difference between experimental and control groups by the standard deviation of two groups (Chen & Peace, 2013). The interpretation of the effect sizes was based on the criteria set by Cohen (1988) for standardized mean differences. These rules indicated small, medium and large effect sizes with .2, .5, and .8, respectively.

In terms of the validity of the study, we tried to reveal the effect of teaching learning strategies on students' academic achievement using standardized data obtained from experimental studies as well as examining the publication bias (Rothstein, Sutton, & Borenstein, 2006). Moreover, in order to provide evidence for validity, the process of calculating and reporting the effect size was explained in detail, the measurement instruments used in the primary studies were determined to have sufficient psychometric properties, and the study was carried out by three authors who had doctoral degree in the field of Educational Sciences.

Two researchers separately carried out the article selection process, calculation of effect sizes and interpretation of the analysis results, and then they compared their findings, which ensured the reliability of the study. Full consistency was observed between the findings obtained by two researchers. Besides, the I^2 statistic was used in addition to the *Q* statistic as a measure of homogeneity/heterogeneity during the data analysis. Also, we decided whether to use fixed effects model or random effects model for the effect sizes. Sub-group analyses were performed using Analog to the ANOVA. As indicated by Lipsey and Wilson (2001), Analog to the ANOVA has similarities with ANOVA, but is basically a Chi-Square-based sub-group analysis method performed to determine whether effect sizes are different in sub-groups in a meta-analysis.

Results

This section will be presented under two headings: findings about the common effect size and findings of the sub-group analyses.

Findings about the Common Effect Size

Table 2 shows the effect sizes and the results of the homogeneity/heterogeneity tests of two different models (i.e., fixed effects and random effects) established by combining the results from the studies included in the meta-analysis.

Table 2

Effect Sizes and Homogeneity/Heterogeneity Test Results

Model	N	Mean ES	Z	SE	%95 CI		df	Q	p	I²
					Low	Up				
Fixed	31	0.933	17.15	0.05	0.83	1.04	30	189.63	.00	84.18
Random	31	1.206	8.532	0.14	0.93	1.48				

In the random effects model, the effect size was computed as 1.21 with a standard error of .14 and 95% confidence interval (a lower limit of 0.93 and an upper limit of 1.48). In the fixed effects model, the effect size was computed as 0.93 with standard error of .05 and 95% confidence interval (a lower limit of 0.83 and an upper limit of 1.04).

The data were tested for homogeneity/heterogeneity (Borenstein et al., 2009). In this sense, the $Q_{(sd=30)}$ statistic was found to be 189.63 ($p < .01$). A Q statistic value exceeding the 30 degrees of freedom and .05 confidence interval (df=30, $\chi^2_{(.05)}$ =43.77) in the chi-square distribution table showed that the data were heterogeneous. Another method for determining homogeneity/heterogeneity is the calculation of the percentage of I^2. I^2 values of 25%, 50% and 75% (and above) indicate low, moderate and high heterogeneity, respectively (Higgins & Thompson, 2002). The I^2 computed from the data was 84.18%, which indicated high heterogeneity.

The authors of the study used the random effects model as they assumed that the differences in learning strategies used in the studies involved in meta-analysis differed beyond the sampling error. Indeed, Field and Gillett (2010) also suggest the use of random effects model for meta-analysis conducted in the social sciences. The mean effect size was calculated as 1.21 based on random effects model. Based on Cohen's (1988) classification, this value indicated that the teaching of learning strategies had a high positive effect on students' achievement. Figure 2 shows the forest plot presenting the distribution of the effect sizes of the studies in the random effects model.
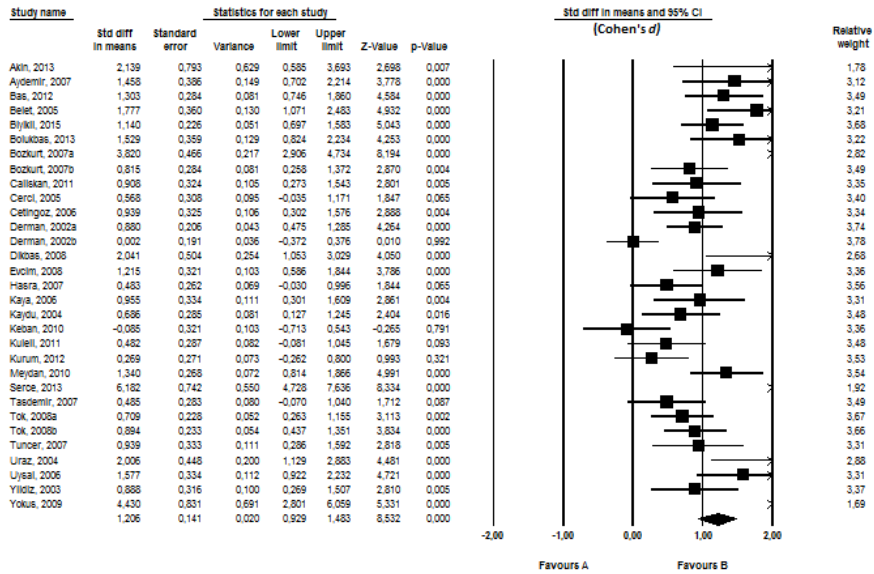
Study name	Statistics for each study							Std diff in means and 95% CI (Cohen's *d*)	Relative weight
	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value		
Akin, 2013	2,139	0,793	0,629	0,585	3,693	2,698	0,007		1,78
Aydemir, 2007	1,458	0,386	0,149	0,702	2,214	3,778	0,000		3,12
Bas, 2012	1,303	0,284	0,081	0,746	1,860	4,584	0,000		3,49
Belet, 2005	1,777	0,360	0,130	1,071	2,483	4,932	0,000		3,21
Biyikli, 2015	1,140	0,226	0,051	0,697	1,583	5,043	0,000		3,68
Bolukbas, 2013	1,529	0,359	0,129	0,824	2,234	4,253	0,000		3,22
Bozkurt, 2007a	3,820	0,466	0,217	2,906	4,734	8,194	0,000		2,82
Bozkurt, 2007b	0,815	0,284	0,081	0,258	1,372	2,870	0,004		3,49
Caliskan, 2011	0,908	0,324	0,105	0,273	1,543	2,801	0,005		3,35
Cercl, 2005	0,568	0,308	0,095	-0,035	1,171	1,847	0,065		3,40
Cetingoz, 2006	0,939	0,325	0,106	0,302	1,576	2,888	0,004		3,34
Derman, 2002a	0,880	0,206	0,043	0,475	1,285	4,264	0,000		3,74
Derman, 2002b	0,002	0,191	0,036	-0,372	0,376	0,010	0,992		3,78
Dikbas, 2008	2,041	0,504	0,254	1,053	3,029	4,050	0,000		2,68
Evclm, 2008	1,215	0,321	0,103	0,586	1,844	3,786	0,000		3,36
Hasra, 2007	0,483	0,262	0,069	-0,030	0,996	1,844	0,065		3,56
Kaya, 2006	0,955	0,334	0,111	0,301	1,609	2,861	0,004		3,31
Kaydu, 2004	0,686	0,285	0,081	0,127	1,245	2,404	0,016		3,48
Keban, 2010	-0,085	0,321	0,103	-0,713	0,543	-0,265	0,791		3,36
Kutell, 2011	0,482	0,287	0,082	-0,081	1,045	1,679	0,093		3,48
Kurum, 2012	0,269	0,271	0,073	-0,262	0,800	0,993	0,321		3,53
Meydan, 2010	1,340	0,268	0,072	0,814	1,866	4,991	0,000		3,54
Serce, 2013	6,182	0,742	0,550	4,728	7,636	8,334	0,000		1,92
Tasdemir, 2007	0,485	0,283	0,080	-0,070	1,040	1,712	0,087		3,49
Tok, 2008a	0,709	0,228	0,052	0,263	1,155	3,113	0,002		3,67
Tok, 2008b	0,894	0,233	0,054	0,437	1,351	3,834	0,000		3,66
Tuncer, 2007	0,939	0,333	0,111	0,286	1,592	2,818	0,005		3,31
Uraz, 2004	2,006	0,448	0,200	1,129	2,883	4,481	0,000		2,88
Uysal, 2006	1,577	0,334	0,112	0,922	2,232	4,721	0,000		3,31
Yildiz, 2003	0,888	0,316	0,100	0,269	1,507	2,810	0,005		3,37
Yokus, 2009	4,430	0,831	0,691	2,801	6,059	5,331	0,000		1,69
	1,206	0,141	0,020	0,929	1,483	8,532	0,000		

Figure 2. Forest Plot Showing the Distribution of the Effect Sizes of the Studies

The black squares in the forest plot show the effect sizes, while the horizontal lines on both sides of each square show the 95% confidence interval of the effect size. Besides, the relative weight in the forest plot shows the effect of the study on the result of the meta-analysis (Littell, Corcoran & Pillai, 2008; Ried, 2006). According to the forest plot shown in Figure 2, Derman's (2002) study had the largest effect on the mean effect size, while Akin's (2013) study had the smallest effect. Moreover, out of 31 effect sizes, only one was found to be negative (in favour of the control group), while 30 of them were positive (in favour of the experimental group). This indicated that the effect was in the positive direction.

Findings of the Sub-group Analyses

In addition to the analyses for determining the common effect size, groups were also compared to determine the sources of heterogeneity.

- While making the comparison, we first examined whether the level of education (elementary, secondary, high schools, and university) caused any statistically significant difference on academic achievement. Table 3 shows the Analog to the ANOVA results revealing whether the effect of teaching learning strategies on students' achievement differ depending on the level of education in the random effects model.

Table 3

Differences in Effect Sizes Depending on Educational Stages in the Random Effects Model

Variable (School)	N	Mean ES	SE	% 95 CI		df	χ^2	Q_B	p
				Low	Up				
Primary	9	1.099	0.121	0.862	1.337				
Secondary	6	0.846	0.246	0.364	1.328				
High	6	1.484	0.440	0.621	2.346				
University	10	1.529	0.375	0.793	2.265				
Total	**31**	**1.108**	**0.102**	**0.909**	**1.308**	**3**	**7.815**	**3.128**	**.37**

Note: These findings are given as mixed effects output in the CMA software.

Heterogeneity below the critical chi-square value (7.82) for the sub-group of level of Education (Q_B=3.13, $p > .05$) indicated that there was no statistically significant difference between the groups. In other words, the teaching of learning strategies had similar effects at all education levels.

- As the second sub-group, we examined whether different disciplines caused any significant difference. Table 4 shows the results from the Analog to the ANOVA performed to determine whether the effect of teaching learning strategies on students' achievement differ significantly depending on the discipline in the random effects model.

Table 4

Differences in Effect Sizes Depending on Disciplines in the Random Effects Model

Variable	N	Mean ES	SE	% 95 CI		df	χ^2	Q_B	p
				Low	Up				
Foreign language	9	1.481	0.341	0.814	2.149				
Social sciences	10	1.295	0.223	0.858	1.733				
Science	6	0.558	0.216	0.135	0.982				
Turkish	4	1.030	0.248	0.545	1.515				
Music	2	3.271	1.145	1.026	5.516				
Total	**31**	**1.042**	**0.122**	**0.803**	**1.281**	**4**	**9.488**	**11.75**	**.02**

Note: These findings are given as mixed effects output in the CMA software.

Heterogeneity above the critical chi-square value (9.49) for the sub-group of disciplines (Q_B=11.76, $p < .05$) indicated that there was a statistically significant difference between the effect sizes of the groups. In other words, the effect of teaching learning strategies on students' achievement differed significantly according to course

type. Hence, the course type was one of the sources of variance. The effect sizes showed that learning strategies mostly affected students' achievement in the field of Music Teaching (3.27) and they had the least effect in the field of Science (0.56). The effect sizes were found to be similar in Social Sciences, Foreign Language and Turkish training.

- As the third sub-group, we examined whether teaching learning strategies independently of the subject or in an integrated or eclectic way caused any significant difference. The first group (strategy teaching only) included studies in which students were provided with only strategy teaching and normal training. The second group (activity only) included studies in which activities based on learning strategies were carried out. Finally, the third group (eclectic) included studies in which both approaches were synthesized. Table 5 shows the results from the Analog to the ANOVA performed to determine whether the effects of these approaches on students' achievement differ in the random effects model.

Table 5

Differences in Effect Sizes Depending on the Ways of Teaching Strategies in the Random Effects Model

Variable	N	Mean ES	SE	% 95 CI Low	% 95 CI Up	df	χ^2	Q_B	p
Only strategy	3	1.559	0.429	0.719	2.400				
Only activity	12	1.381	0.291	0.811	1.952				
Eclectic	16	1.072	0.167	0.744	1.400				
Total	**31**	**1.191**	**0.137**	**0.922**	**1.460**	**2**	**5.991**	**1.671**	**.43**

Note: These findings are given as mixed effects output in the CMA software.

Heterogeneity below the critical chi-square value (5.99) for the sub-group of teaching style (Q_B=1.67, *p* > .05) indicated that there was no statistically significant difference between the groups. In other words, different teaching styles yielded similar results.

- As the fourth sub-group, we examined whether the type of strategy used in the studies (repetition, sense-making, organizing, monitoring the comprehension, affective, integrated, foreign language learning, and foreign language vocabulary learning) caused any significant difference. While the first five of them were the strategies in the literature, the integrated strategy expresses the teaching of at least two of the first five strategies combined. The seventh and the eighth strategies were exclusive to foreign language learning strategies. Table 6 shows the results from the Analog to the ANOVA to test the effects of the type of strategy used.

Table 6

Differences in Effect Sizes Depending on the Type of Strategy in the Random Effects Model

Variable	N	Mean ES	SE	% 95 CI		df	χ^2	Q_B	p
				Low	Up				
Repetition	1	1.14	0.23	0.70	1.58				
Sense-making	5	1.57	0.45	0.69	2.45				
Organizing	3	0.55	0.33	-0.09	1.19				
Monitoring the comprehension	1	0.71	0.23	0.26	1.16				
Integrated	15	1.15	0.18	0.80	1.50				
Foreign language learning	3	0.73	0.36	0.03	1.43				
Foreign language vocabulary learning	3	2.84	1.13	0.63	5.04				
Total	**31**	**0.998**	**0.103**	**0.79**	**1.20**	**6**	**12.59**	**9.52**	**.14**

Note: These findings are given as mixed effects output in the CMA software.

Heterogeneity below the critical chi-square value (12.59) for the sub-group of learning strategy type (Q_B=9.528, *p* > .05) indicated that there was no statistically significant difference between the groups. In other words, the use of any kind or the combination of learning strategy types yielded similar results. In detail, it could be alleged that the vocabulary learning strategies were the most effective strategies; repetition, sense-making and integrated strategies were relatively equal and low when compared to vocabulary learning strategies; and organizing, monitoring the comprehension and foreign language learning strategies were the least effective ones.

Publication Bias

The mean effect size calculations obtained in a study must reflect the reality. Publication bias is the most important threat to showing the reality, i.e., reliability of the study. Therefore, as mentioned by Rothstein, Sutton and Borenstein (2006), the funnel plot shown in Figure 3 was examined using the trim-and-fill method proposed by Duval and Tweedie (2000) in order to scrutinize the suitability of the computed effect size to the purpose.
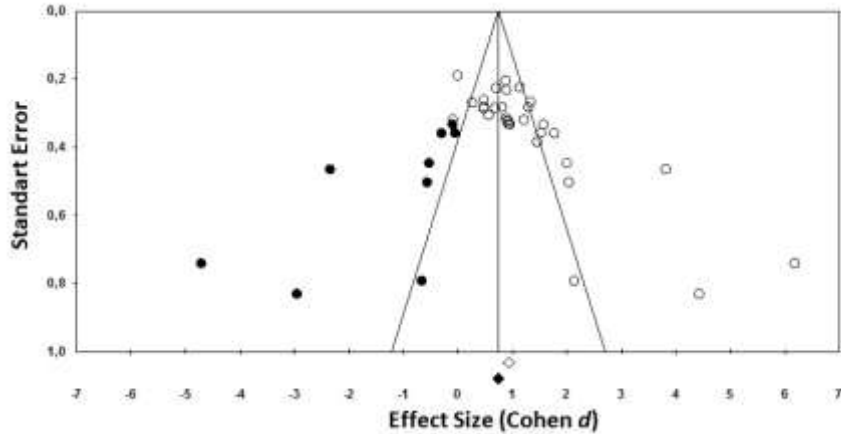
Figure 3. Funnel Plot

A funnel plot was used to examine publication bias (see Figure 3). Empty circles in Figure 3 show the studies included in the study. Filled circles show the imaginary studies that need to be included to eliminate public bias (Duval & Tweedie, 2000). According to the funnel plot, the line at the centre of empty circles should be as symmetrical as possible on both sides in order to eliminate publication bias. It can be said that the funnel plot in Figure 3 appears to be symmetric. Indeed, the funnel plot also showed that addition of only nine studies with favourable results for the control group would be enough to eliminate publication bias totally for this study. Given the 31 effect sizes were computed within the scope of this study, it can be concluded that these nine studies can be neglected. Even if it was not neglected, the standardized effect size was reported to be 0.77 in the event of including nine negatory studies according to Duval and Tweedie's (2000) trim-and-fill method. This effect size was also very close to the large level of impact according to the criteria provided by Cohen (1988). Any review of a funnel plot would be subjective; therefore, Rosenthal's (1979) "fail-safe N" (Rothstein, Sutton, & Borenstein, 2006) was also examined. The fail-safe N for effect size was found to be 2987 at .05 confidence level. As the value of 2987 was much larger than the value of 165 obtained by the formula 5k + 10 (Fragkos, Tsagris, & Frangos, 2014), it indicated that there was no publication bias in the findings.

Discussion, Conclusion and Recommendations

The purpose of this study was to reveal the effect of learning strategies on academic achievement in Turkey with the use of meta-analysis. Thus, this can be taken as a significant contribution as there has been no similar previous research on learning strategies in Turkey, and it reveals the general state on the controversial issues in the relevant literature. Moreover, it is conducted only through the studies in Turkey with the idea that it will yield more meaningful proposals for national education policies. Therefore, it can be claimed that the research results could be evaluated exclusively to

Turkey. The research is also limited by the analysis of moderator variables that are coded. We first calculated the effect size of the data obtained from the studies based on the random and fixed effects models. Since the two researchers as specialists assumed that the difference in the learning strategies used in the studies differed according to the level of education, discipline, strategy teaching style and strategy type beyond the sampling error, random effects model was preferred. As a result, the effect size was found to be 1.21. The eta-squared calculated for Cohen's *d* of 1.21 was found to be .268. This value obtained from the results of 31 effect sizes obtained from 28 studies conducted with 1,641 students indicated that the learning strategies had 26.8% positive effect on students' academic achievement.

The common effect size of Cohen's *d* found to be 1.21 from the experimental studies conducted, and the standardized effect size was reported to be 0.77 in the event of including negatory nine studies in the relevant field in Turkey indicated that teaching learning strategies had a high positive effect on students' achievement (Cohen, 1988; Ellis, 2009). This finding is in compliance with the meta-analyses carried out in the relevant field (Ardasheva, Wang, Adesope & Valentine, 2017; de Boer, H., Donker, A. S., & van der Werf, 2014; Donker, Boer, Kostons, van Ewijk, & van der Werf, 2013; Ergen & Kanadli, 2017), and individual research results conducted in Turkey revealing that teaching learning strategies improves student achievement when compared to control group (Bas, 2012; Belet, 2005; Biyikli & Dogan, 2015; Bozkurt, 2007; Cetingoz, 2006; Hasra, 2007; Kaydu, 2004; Serce, 2013; Tasdemir & Tay, 2007; Uraz, 2004; Yildiz, 2003; Yokus, 2009), while they contradict with the findings of a few studies (Derman, 2002; Keban, 2010). Indeed, this study presents a general result of all studies chosen based on certain criteria, thus revealing that positive effect of learning strategies is more common, and these strategies can be taught. On the other hand, results of the meta-analysis conducted by Ergen and Kanadli (2017) revealed that self-regulated learning strategies had a "large" effect (*d* = 0.859) on academic achievement. Their result supports the result of the positive effects of learning strategies on achievement in 33 out of 38 postgraduate theses as a part of Demirel, Askin-Tekkol, Cigdem, and Demir's (2016) study.

Ozer (2003) criticizes the teaching of course content alone in the curricula of all the levels of education in Turkey and concludes that the students' acquiring the learning strategies being largely left to themselves, they learn and use their strategies at random or by hand, the teaching learning strategies is carried out for academic studies. Likewise, Somuncuoglu and Yildirim (1998) suggested the adaptation of an education policy oriented towards learning strategies beyond their study. However, within the scope of the changes in the curricula in 2005, the common skills in the backbone of all the programs, which need to be earned by the individuals, were reflected in the curriculum as key competencies with the amendment made in 2017. Considering learning to learn as one of the focal points of all curricula among these competencies (MoNE, 2017) can be regarded as an extremely significant development. Besides, the expression that competence areas are equally important, interrelated and compatible with each other, mutually exclusive, supportive (MoNE, 2018) demonstrates that the ideas of Weinstein et. al. (1989) and Somuncuoglu and Yildirim (1998) find place in the

curricula. All those are important for learning strategies to spread over, attaching the necessary importance to learning strategies and training individuals who learn to learn in Turkey.

In addition to finding the common effect size, we also compared the groups based on the level of education, discipline, strategy teaching styles, and strategy types in order to determine the source of heterogeneity among the findings. As a result of the analyses, we found that the level of education (elementary, secondary, high schools, and university) did not cause a statistically significant difference in teaching the learning strategies. Similarly, the effect size calculated in Ergen and Kanadli (2017) showed no significant difference according to the type of school level. This indicates that teaching learning strategies is effective at all levels of education, and verifies the views of researchers (Tok, 2008) who think that the teaching of learning strategies is more effective at younger ages in terms of its importance in later learning experiences and for the acquisition of learning to learn at early ages. We also analyzed whether teaching learning strategies in different disciplines caused any significant difference, and found that teaching such strategies mostly affected students' achievement in the field of Music Teaching, while it had the least effect in the field of Science which is one of the positive sciences. Within the scope of the research, the numbers of primary studies based on the disciplines are nine in foreign languages, six in science, 10 in social sciences, four in Turkish and two in Music. It might be claimed that the difference between the numbers of the primary studies may have an impact upon the reliability of the results (Ayaz & Soylemez, 2015); and therefore, the highest achievement level was obtained in music course. On the other hand, the effect sizes were found to be similar in Social Sciences, Foreign Language, and Turkish. It can be deduced that learning strategies are more effective in verbal fields considering that the fields of foreign language, social sciences and Turkish are verbal and physical sciences are numerical. This can be explained by learning strategy activities such as summarizing, underlining, and identifying the main idea support the verbal field more. However, these results contradict with the results of Donker et al. (2013). In their meta-analysis study including 58 research related to self-regulated teaching, Donker et al. (2013) calculated the effect sizes of 0.73 in physical sciences, 0.66 in mathematics, 0.36 in reading comprehension. However, the effect size calculated in Ergen and Kanadli (2017) showed no significant difference according to the course type for self-regulated learning strategies. This result may be due to limiting this study to cognitive learning strategies, and can be explained as the learning strategies cause diverse effects for different lessons in the context of varied countries. On the other hand, the divergence between the results proves the importance of carrying out such kind of research in the context of Turkey.

As mentioned earlier, there is still a debate on whether teaching learning strategies should be independent of the subject or should be integrated into it. To reach a conclusion and find the source of heterogeneity among the findings, we examined whether teaching learning strategies independently of the subject or in an integrated or eclectic way caused any significant difference. We found that different teaching styles did not cause any significant difference in students' achievement. In other

words, different teaching styles yielded similar (influential) results. In a similar vein, the use of any or combination of the learning strategy types (repetition, sense-making, organizing, monitoring the comprehension, affective, integrated [teaching at least two of the first five strategies combined], foreign language learning, and foreign language vocabulary learning) was found to yield similar results beyond the studies. All those suggest that teaching learning strategies can be taught via either their integration into classes or independently of the lessons or the combination of both, and that the use of any kind or the combination of strategy types will result in effective consequences. In detail, it could be alleged that the vocabulary learning strategies were the most effective strategies (2.84), repetition (1.14), sense-making (1.57), and integrated strategies (1.15) were relatively equal and low when compared to vocabulary learning strategies; and organizing (0.55), monitoring the comprehension (0.71), and foreign language learning strategies (0.73) were the least effective ones. Donker et al. (2013) calculated the effect size of repetition strategies as 1.39, that of organizing strategies as 0.81 and that of sense-making strategies as 0.75, and concluded that it was not right to focus solely on what strategy to be used and how to use them in the teaching of learning strategies, and it should be emphasized when and why the strategies (situational knowledge) should be used. On the basis of a meta-analysis on language learning strategies, Ardasheva, Wang, Adesope and Valentine (2017) also reveal that the ideal teaching is carried out through a balance between strategy and knowledge and that it is much more powerful to focus less on strategy rather than a broader strategy repertoire.

Based on the findings obtained from this study, it can be asserted that the teaching of any kind of learning strategy type is effective on achievement on its own or in combination with different strategies whether independently, or integrated, or eclectically at every level of education and in every discipline - especially in verbal based courses - in Turkey. Although the concept of "learning to learn" is regarded among the competencies that constitute the backbone of the curriculum with the last amendment of MoNE, the examination of the curriculum proved that exclusively the concept of "learning strategy" appeared in a few times. This may lead to a lesser awareness in teachers. Therefore, it is recommended to incorporate learning strategy types and application examples as well as the concept of "learning strategy" which is on the basis of learning to learn into curricula at all educational stages and in all disciplines. For the subsequent studies, meta-analysis studies may be conducted in which the effects of teaching learning strategies on different variables (particularly the attitude as being one of the mostly studied ones) are compiled.

References

(References with asterisk (*) indicate studies included in the meta-analysis)

*Akin, O. (2013). The effect of elaboration strategy on academic achievement in violin training (a sample application). *Kastamonu Education Journal*, *21*(4), 1551-1560.

Akkus-Ispir, O., Ay, Z. S. P., & Saygi, E. (2011). Self-regulating learning strategies, motivation towards mathematics and thinking styles of gifted students. *Education and Science, 36*(162), 235-246.

Ardasheva, Y., Wang, Z., Adesope, O. O., & Valentine, J. C. (2017). Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes. *Review of Educational Research, 87*(3), 544–582.

Arends R. I. (1997). *Classroom instruction and management.* New-York: The McGraw-Hill Companies, Inc.

Ayaz, M. F. & Soylemez, M. (2015). The effect of the project-based learning approach on the academic achievements of the students in science classes in Turkey: A Meta-Analysis Study. *Education and Science, 178*, 255-283.

*Aydemir, U. V. (2007). *The effect of language learning strategies on the achievement of students in teaching English* (Unpublished master thesis). Uludag University, Institutes of Educational Sciences, Bursa.

*Bas, G. (2012). The effect of teaching learning strategies on students' academic achievement, attitudes and cognitive awareness levels in English courses. *Journal of Theoretical Educational Science*, *5*(1), 49-71.

*Belet, S. D. (2005). *The effect of learning strategies on reading comprehension and writing skills and attitudes towards Turkish course* (Unpublished doctoral dissertation). Anadolu University, Institutes of Educational Sciences, Eskisehir.

*Biyikli, C., & Dogan, N. (2015). The effect of using learning strategies on academic achievement. *Education and Science*, *40*(181), 311-327.

*Bolukbas, F. (2013). The effect of language learning strategies on learning vocabulary in teaching Turkish as a foreign language *Hacettepe University Journal of Education*, *28*(3), 55-68.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (version 2). Englewood, NJ: Biostat.

*Bozkurt, N. (2007). *The effect of different learning strategies applied in high school - 1 history course on students' achievement and retention*. Unpublished doctoral dissertation. Gazi University, Institutes of Educational Sciences, Ankara.

Brandt, R. (1988/1989). On learning research: A conversation with Lauren Resnick. *Educational Leadership, 46*(4), 12-16.

Bulus, M., Duru, E., Balkis, M., & Duru, S. (2011). The role of learning strategies and individual traits in predicting academic achievement of prospective teachers. *Education and Science, 36*(161), 186-198.

Caliskan, M., & Sunbul, A. M. (2011). The effects of learning strategies instruction on metacognitive knowledge, using metacognitive skills and academic achievement (Primary education sixth grade Turkish course sample). Educational Sciences: Theory and Practice, 11*(1), 148-153.

*Cerci, A. (2005). *The effect of learning strategies on students' comprehension level in teaching Turkish* (Unpublished master thesis). Selcuk University, Institute of Social Sciences, Konya.

Cesur, M. O., & Fer, S. (2011). A model explaining relationships between language learning strategies, learning styles and success in reading comprehension. *Journal of Education, 41*, 83-93.

*Cetingoz, D. (2006). *The teaching of note taking strategy, the success of history, keeping in mind and success motivation* (Unpublished doctoral dissertation). Dokuz Eylul University, Institutes of Educational Sciences, Izmir.

Chen, D. G. D., & Peace, K. E. (2013). *Applied meta-analysis with R.* Boca Raton, London, New York: CRC Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2. edition).* Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education (6th Edition).* New York: Routledge.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis (2nd edition).* New York: Russell Sage Publication.

Dansereau, D. F., Collins, K. W., McDonald, B. A., Holly, C. D., Garland, J., Diekhoff, G., & Evans, S. H. (1979). Development and evaluation of a learning strategy training program. *Journal of Educational Psychology, 71*(1), 64-73. http://dx.doi.org/10.1037/0022-0663.71.1.64

*Derman, A. (2002). *The effect of different learning strategies used in 7th grade science classrooms on students' achievement* (Unpublished master thesis). Selcuk University, Institutes of Social Sciences, Konya.

*Dikbas, Y., & Hasirci, O. K. (2008). The effect of teaching learning strategies and their use in teaching on students' academic achievement and attitudes. *Ahi Evran University, Journal of Kirsehir Education Faculty, 9*(2), 69-76.

de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research, 84*(4), 509–545.

Demirel, M. (1993). Teaching learning strategies. *Education and Science, 17*(88), 52-59.

Demirel, O. (2003). *Education dictionary (2.edition).* Ankara: Pegema Publishing.

Donker, A. S., de Boer, H., Kostons, D., van Ewijk, C. D., & van der Werf, M. P. C. (2013). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educ. Res. Rev., 11*, 1–26.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

Ellis, P. D. (2009). *Thresholds for interpreting effect sizes*. Retrieved March 13, 2017 from http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html

Erdem, A. R. (2005). Effective ways of learning: Learning strategies and teaching. *Elementary Education Online, 4*(1), 1–6.

Ergen, B. & Kanadli, S. (2017). The effect of self-regulated learning strategies on academic achievement: A meta-analysis study. *Eurasian Journal of Educational Research 69*, 55-74.

*Evcim, H. (2008). *Increasing vocabulary knowledge through language learning strategies of university preparatory class students* (Unpublished master thesis). Abant İzzet Baysal University, Institutes of Educational Sciences, Bolu.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694.

Fragkos, K. C., Tsagris, M., & Frangos, C. C. (2014). Publication bias in meta-analysis: confidence intervals for Rosenthal's fail-safe number. International scholarly research notices, 2014.

Gagne, R. M. (1988). *Essentials of learning for instruction*. 2nd Ed., New Jersey: Prentice Hall, Inc.

Ghee, T.T., Ismail, H.N., & Kabilan M.K. (2010). Language learning strategies used by MFL students based on genders and achievement groups. *US-China Foreign Language, 8*(1), 50- 58.

*Hasra, K. (2007). *The effect of the teaching of learning strategies with brain based learning approach on students' reading comprehension skills* (Unpublished master thesis). Mugla University, Institutes of Social Sciences, Mugla.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine, 21*(11), 1539-1558.

*Kaya, F. (2006). *The effect of some learning strategies on attitudes and reading comprehension in primary school fourth grade Turkish course*. Unpublished master thesis. Mustafa Kemal University, Institutes of Social Sciences, Hatay.

*Kaydu, M. (2004). *The effect of different learning strategies used in geography courses on the success of students in secondary education 1st grade* (Unpublished master thesis). Selcuk University, Institutes of Social Sciences, Konya.

*Keban, F. (2010). *Investigation of the effects of strategy teaching in cooperative learning groups in basic physics laboratories at undergraduate level* (Unpublished master thesis). Dokuz Eylul University, Institutes of Educational Sciences, İzmir.

*Kizilaslan-Tuncer, B. (2007). *The effect of the use of learning strategies in teaching on students' academic achievements, retention levels and attitudes towards the course* (Unpublished master thesis). Canakkale Onsekiz Mart University, Institutes of Social Sciences, Canakkale.

*Kuleli, M. (2011). *The role of learning strategies in learning English grammar* (Unpublished master thesis). Abant İzzet Baysal University, Institutes of Educational Sciences, Bolu.

*Kurum, E. Y. (2012). *The impact of strategy based instruction on improving the foreign language proficiency level of the 3rd grade students at the Turkish military academy*. Hacettepe University, Graduate School of Social Sciences, Ankara.

Lenz, B. K. (1992). Self-managed learning strategy systems for children and youth. *School Psychology Review, 21*(2), 211-228.

Levin, J. R. (1986). Four cognitive principles of learning-strategy instruction. *Educational Psychologist*, *21*, 3-17.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis. Applied Social Research Methods Series (Vol. 49)*. Thousand Oaks, CA: Sage publications.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Mayer, R. E. (1980). Elaboration techniques that increase the meaningfulness of technical text: An experimental test of the learning strategy hypothesis. *Journal of Educational Psychology, 72*(6), 770-784. http://dx.doi.org/10.1037/0022-0663.72.6.770

*Meydan, A. (2010). The effect of "learning to learn strategy" on the student's learning and permanency of 4th grade unit "the environment we live". *Selcuk University Journal of Social Sciences Institute, 23*, 149-157.

MoNE (2018). Primary and Secondary School Mathematics Curriculum (Grades 1-8), Ankara.

MoNE Media Briefing. (2017). On Renewal and Change in Curriculum. https://ttkb.meb.gov.tr/meb_iys_dosyalar/2017_07/18160003_basin_aciklamasi-program.pdf

Ning, K. H., & Downing, K. (2011). The interrelationship between student learning experience and study behaviour. *Higher Education Research and Development, 30*(6), 765-778.

O'Malley, J.M., & Chamot, A.U. (1990) *Learning strategies in second language acquisition.* Cambridge: University Press

Oxford, R.L. (1990) *Language learning strategies: What every teacher should know.* Boston: Heinle and Heinle.

Ozer, B. (17-18 November, 2003). Teachers' new task: learning to learn. Presented in Learning to Learn Activities. ODTU Culture and Convention Center , Ankara.

Paris, S.G., Cross, D.R., & Lipson, M.Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76*(6), 1239-1252.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining selfregulated learning. *International Journal of Educational Research, 31*, 459–470.

Ried, K. (2006). Interpreting and understanding meta-analysis graphs: a practical guide. Retrieved March 28, 2017 from https://dspaceprd01.services.adelaide.edu.au/dspace/bitstream/2440/43554/1/hdl_43554.pdf

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-664.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* West Sussex: John Wiley and Sons.

Saracaloglu, A. S., & Karasakaloglu, N. (2011). An investigation of prospective teachers' reading comprehension levels and study and learning strategies related to some variables. *Education and Science, 36*(161), 98-115

Schmitt, N. (1997). Vocabulary learning strategies. *Vocabulary: Description, acquisition and pedagogy, 199227.*

Senemoglu, N. (2001). *Development, learning and teaching (3rd Ed.).* Ankara: Gazi Bookstore.

*Serce, H. (2013). *The effects of teaching vocabulary learning strategies on vocabulary learning and strategy use* (Unpublished doctoral dissertation). Necmettin Erbakan University, Institutes of Educational Sciences, Konya.

Somuncuoglu, Y., & Yildirim, A. (1998). Learning strategies: Theoretical dimensions, research findings and results *Science and Education, 22*(110), 31-39.

Sonmez, V. (2007). *Teaching principles and methods.* Ankara: Ani Publishing.

*Tasdemir, A., & Tay, B. (2007). The effects of students' using learning strategies on science achievement in science teaching. *Uludag University Faculty of Education Journal*, *20*(1), 173-187.

Toffler, A. (1974). *Learning for tomorrow*. USA: Vintage.

*Tok, S. (2008). The effects of note-taking and KWL strategy on attitude and academic achievement. *Hacettepe University Journal of Education,* (34), 244-253.

Tunca, N., & Alkin-Sahin, S. (2014). The relationship between pre-service teachers' cognitive (metacognitive) learning strategies and academic self-efficacy beliefs. *Anadolu International Journal of Educational Sciences, 4*(1), 47-56.

*Uraz, B. (2004). *The relationship between learning strategy choice in reading comprehension and academic success.* Unpublished master thesis. Gazi University, Institutes of Educational Sciences, Ankara.

Ustun, U., & Eryilmaz, A. (2014). A research methodology to conduct effective research syntheses: Meta-analysis. *Education and Science, 39*(174).

*Uysal, I. (2006). *The effect of learning strategies in social sciences lesson on students' level of understanding and interpretation* (Unpublished master thesis). Selcuk University, Institutes of Social Sciences, Konya.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods, 13*(2), 130-149.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*(2), 215-247.

Weinstein, C.E., & MacDonald, J.D. (1986). Why does a school psychologist need to know about learning strategies?. *Journal of School Psychology, 24*(3), 257-265.

Weinstein, C.E., & Mayer, R.E. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (315-327), New York: MacMillan.

Weinstein, C.E., Ridley, D.S., Dahl, T., & Weber, E.S. (1988/1989). Helping students develop strategies for effective learning. *Educational Leadership, 46*(4), 17-19.

*Yildiz, N. (2003). *The effect of learning strategies that are provided to students in the 5th grade science lesson on their academic achievements and retention levels.* (Unpublished master thesis). Anadolu University, Institutes of Educational Sciences, Eskisehir.

*Yokus, H. (2009). *The effect of activities related to the use of learning strategies in piano education on performance achievement and metacognitive awareness* (Unpublished doctoral dissertation). Marmara University, Institutes of Educational Sciences, Istanbul.

Öğrenme Stratejileri Öğretiminin Akademik Başarı Üzerine Etkisi:
Bir Meta-Analiz Çalışması

Özet

Problem Durumu: Bireyin nasıl daha iyi ve kolay öğreneceğini bilmesinin, bağımsız öğrenmenin kısacası öğrenmeyi öğrenmenin temelini oluşturduğu ifade edilen öğrenme stratejileri, farklı araştırmacılar tarafından tanımlanmıştır. Öğrenme stratejisi öz olarak bilgi ve becerilerin elde edilmesini kolaylaştırmak için özel bir öğrenme durumunda birey tarafından kullanılan zihinsel taktiklerin tümü olarak tanımlanmaktadır. Öğrenme stratejileri ile ilgili çalışmaların büyük çoğunluğu betimsel ve ilişkisel olmakla birlikte deneysel çalışmalar da mevcuttur. Deneysel çalışmalarda genel olarak öğrenme stratejileri öğretiminin akademik başarıya, tutuma, üstbilişsel farkındalığa, kalıcılığa, yürütücü biliş becerisine vs. etkisi incelenmiş ve çalışmaların çoğunda ilgili değişkenlerde olumlu gelişmeler görüldüğü ortaya konmuştur. Türkiye'deki alan yazında öğrenme stratejileri öğretiminin öğrencilerin çeşitli derslerdeki akademik başarılarına etkisini belirlemek amacıyla yapılmış çok sayıda çalışma olmasına rağmen, bu konuda yapılmış bir meta-analiz çalışmasına rastlanmamıştır. Farklı boyutları ile ele alınan öğrenme stratejileri ile ilgili araştırmaların birleştirilmeye, sentezlenmeye ve değerlendirilmeye ihtiyacı vardır. Bu anlamda Türkiye kapsamında yapılan çalışmaların sonuçlarının birleştirilerek genel yargının ortaya konması alan yazına önemli katkılarda bulunacaktır.

Araştırmanın Amacı: Türkiye'deki alan yazında öğrenme stratejileri öğretiminin öğrencilerin çeşitli derslerdeki akademik başarılarına etkisini belirlemek amacıyla yapılmış çok sayıda çalışma olmasına rağmen, (bilişsel) öğrenme stratejilerinin etkisini deneysel olarak inceleyen araştırmaları ele alan kapsamlı bir meta analiz çalışmasına ulaşılamamıştır. Bu bağlamda bu araştırma, Türkiye'de öğrenme stratejileri öğretiminin öğrencilerin akademik başarılarına etkisini inceleyen 2000-2016 yılları arası deneysel çalışma sonuçlarını meta-analiz yolu ile bir araya getirmek amacıyla yürütülmüştür.

Araştırmanın Yöntemi: Bu araştırmada, öğrenme stratejileri öğretiminin öğrenci başarısı üzerindeki etkisini inceleyen deneysel ve yarı-deneysel araştırmaların sonuçlarını bir araya getirmek amacıyla meta-analiz yöntemi kullanılmıştır. Yapılan taramaya göre ulaşılan toplam 1436 yayın yukarıdaki kriterlere göre tasnif edildiğinde elde kalan 352 çalışma üzerinde yoğunlaşılmıştır. YÖK Ulusal Tez Merkezi'nde özetlerinden araştırma ölçütlerine uygun olduğu belirlenen ancak erişime açık olmayan tez yazarlarına e-posta gönderilmiş ancak cevap alınamamıştır. Buna göre, belirlenen

ölçütler çerçevesinde öğrenme stratejilerinin öğrenci başarısı üzerindeki etkisini deneysel olarak inceleyen lisansüstü tezler ile bilimsel makalelerden oluşan toplam 28 çalışma araştırmaya dâhil edilmiştir. Bu 28 çalışma haricinde yeterli veri sunan ve gerekli şartları sağlayan tam metin bildiri veya poster sunumu türünde çalışmaya rastlanılmamıştır.

Araştırmanın geçerliği bağlamında öğrenme stratejileri öğretiminin öğrenci başarısı üzerindeki etkisi deneysel araştırmalardan elde edilen standartlaştırılmış verilerle ortaya koyulmaya çalışıldığı açıklanmış ve yayın yanlılığı incelenmiştir. Ayrıca geçerliği sağlayabilmek adına, etki büyüklüğü hesaplama ve raporlama süreçleri ayrıntılı olarak açıklanmış, birincil çalışmalarda kullanılan ölçme araçlarının yeterli psikometrik özellikleri taşıdıkları belirlenmiş ve çalışma Eğitim Bilimleri alanında doktoralı üç akademisyen tarafından yürütülmüştür.

Araştırmanın Bulguları: Araştırmada, belirlenen kriterler çerçevesinde meta-analize dâhil edilen araştırmalardaki verilerin rastgele ve sabit etkiler modellerine göre etki büyüklüğü değerleri hesaplanmıştır. Alan uzmanı olan iki araştırmacı, çalışmalarda kullanılan öğrenme stratejilerindeki farklılığın örneklem hatasının ötesinde eğitim kademesi, disiplin, strateji öğretim tarzı ve strateji türü bağlamında farklılaştığını varsaydıkları için rastgele etkiler modeline karar vermiş ve etki büyüklüğü değeri 1.21 olarak belirlenmiştir. Araştırma kapsamında eğitim kademesi, disiplin, strateji öğretim tarzı ve strateji türüne yönelik alt grup analizleri de yapılmış ve öğrenme stratejileri öğretiminin yalnızca disiplin alanına göre istatistiksel olarak anlamlı şekilde farklılaştığı belirlenmiştir. Belirlenen etkinin geçerliğine ilişkin yayın yanlılığı olmadığı yapılan analizler ile ortaya konulmuş olup elde edilen etki değerinin geçerli olduğu sonucuna ulaşılmıştır.

Araştırmanın Sonuçları ve Önerileri: Rastgele etkiler modeline göre hesaplanan 1.21 Cohen *d* değerinin eta-kare karşılığı .268'dir. Bu değer; 1641 öğreneni kapsayan 28 çalışmadan elde edilen 31 etki büyüklüğüne göre öğrenme stratejilerinin öğrenen başarısı üzerinde % 26.8'lik olumlu bir katma değere sahip olduğunu göstermektedir. Eldeki araştırmada ortak etki değerini belirleme analizlerine ek olarak, bulgulardaki heterojenliğin nereden kaynaklandığını belirlemek amacıyla eğitim düzeyi, disiplin, strateji öğretim tarzı ve strateji türüne göre grup karşılaştırmaları da yapılmıştır. Yapılan analizler neticesinde eğitim kademesinin (ilkokul, ortaokul, lise, üniversite) öğrenme stratejileri öğretiminde anlamlı bir değişim meydana getirmediği sonucuna ulaşılmıştır. Bu sonuç öğrenme stratejileri öğretiminin eğitimin her kademesinde etkili olduğunu göstermekte; öğrenme stratejilerinin küçük yaşlarda öğretilmesinin daha etkili olduğu görüşünde olan araştırmacıların görüşlerini sonraki eğitim yaşantılarında ve öğrenmeyi öğrenmenin küçük yaşlarda kazanılmasında önemli olduğu noktalarında doğrulamaktadır. Öğrenme stratejileri öğretiminin farklı disiplinlerde anlamlı bir değişime neden olup olmadığına ilişkin analizler ise öğrenme stratejileri öğretiminin öğrenci başarısını en fazla müzik eğitimi en az pozitif bilimlerden olan Fen bilimleri alanında olumlu yönde etkilediğini göstermiştir. Araştırma kapsamında disiplin alanlarına yönelik birincil araştırma sayısı Yabancı dilde 9, Fen bilimlerinde 6, Sosyal bilimlerde 10, Türkçe'de 4 ve Müzik'te 2 olarak görülmektedir. Birincil çalışma sayıları arasındaki farkın sonuçları etkileyebileceği, bu

nedenle müzik dersindeki başarının en yüksek çıktığı düşünülebilir. Diğer yandan sözel alan olarak sınıflanabilecek Sosyal bilimler, Yabancı dil, Türkçe alanlarında etki büyüklüklerinin de birbirine yakın olduğu sonucuna ulaşılmıştır. Yabancı dil, Sosyal bilimler ve Türkçe alanlarının sözel, Fen bilimleri alanlarının sayısal olduğu düşünüldüğünde öğrenme stratejilerinin sözel alanlarda daha etkili olduğu çıkarımında bulunulabilir. Öğrenme stratejisi öğretiminin alan bağımsız, bütünleştirilmiş veya eklektik olarak gerçekleştirilmesinin öğrenci başarısında anlamlı bir değişim meydana getirmediği sonucuna ulaşılmıştır. Bir başka deyişle farklı tarzlarda öğrenme stratejisi kazandırma yaklaşımları benzer sonuçlar doğurmaktadır. Benzer şekilde çalışmalarda kullanılan öğrenme stratejisi türünün de benzer sonuçlar doğurduğu tespit edilmiştir. Tüm bunlar öğrenme stratejileri öğretiminin gerek derslere entegre edilerek, gerek derslerden bağımsız gerekse her ikisinin bütünleştirildiği şekilde öğretilebileceğini ve herhangi bir strateji türünün veya karmasının kullanılmasının tümünün etkili sonuçlar doğurduğunu göstermektedir.

Anahtar Kelimeler: öğrenme stratejileri, akademik başarı, meta-analiz.

Using Alignment Index and Polytomous Item Response Theory on Statistics Essay Test

Kana HIDAYATI[1], BUDIYONO[2], SUGIMAN[3]

ARTICLE INFO	ABSTRACT
	Purpose: Essay test in mathematics, both in the form of restricted-response and extended-response, generally consist of polytomous scored items. However, the essay test used by teachers in Indonesia has not been fully supported by sufficient quality evidence. There have been many studies focusing on the development of the essay test, but not many of them have applied the use of relevant measurement theory for the polytomous data. The evidence of content validity also has not been supported by its alignment with the curriculum. This study used alignment index to

prove the content validity and IRT polytomous GPCM to determine the characteristics of test items in order to produce an essay test that could accurately measure the achievement of students on statistical materials.
Method: Procedures of this study: (1) preparation of preliminary test, (2) trials, (3) interpretation. Trial was conducted involving 688 Junior High School students in Yogyakarta, Indonesia.
Results: The content validity of the test was good, supported by V Aiken index of 0.88–1.00 and Porter alignment index of 0.93. The test items had good construct validity. Test reliability was categorized as good with the Construct Reliability coefficient of 0.88 and the Alpha coefficient of 0.78. Judging from its characteristics, all test items were categorized as good.
Implications for Research and Practice: The use of the alignment index contribution to the verification of content validity of essay test and the use of the IRT polytomous GPCM may provide reference for the use of appropriate measurement theory to determine the item characteristics of essay test.

---

[1] Corresponding Author: Yogyakarta State University, INDONESIA, e-mail: kana@uny.ac.id, ORCID: https://orcid.org/0000-0002-9226-8500

[2] Sebelas Maret University, INDONESIA, e-mail: budiyono@staff.uns.ac.id, ORCID: https://orcid.org/0000-0001-6467-7801

[3] Yogyakarta State University, INDONESIA, e-mail: sugiman@uny.ac.id, ORCID: https://orcid.org/0000-0002-9226-8500

Introduction

Assessment is an important component for the successful achievement of learning mathematics in school. Some of the things assessed in the learning of mathematics include the understanding of concepts and problem-solving skills. Based on Curriculum 2013 in Indonesia, this is manifested in the form of an assessment of students' mathematical knowledge competence achievement. The use of the instrument in the form of essay test in the assessment of mathematical knowledge, especially on statistical material is very beneficial to be selected by teachers.

Essay tests in mathematics can be presented in various formats. In general, essay formats are usually classified into two groups: restricted response items and extended-response items (Nitko & Brookhart, 2011, p. 204). The development of mathematics learning shows that both types of essay tests can be used in mathematical assessment. The advantages of the essay test by Walstad (2006, p. 4) are: (1) it has a great potential to assess students' level of understanding at a higher level, (2) students have the freedom to prepare, choose, and present ideas in their own words while answering, (3) teachers have the opportunity to see their students answers, (4) it is suitable for achievement tests related to problem analysis, concept application, or decision evaluation. Therefore, the essay test is well suited to measure and assess the achievement of students' mathematical knowledge competence.

The use of a valid and reliable instrument that meets the criteria as a good item will provide an accurate and accountable assessment result. Validity evidence of an instrument generally includes content validity and construct validity. Evidence of content validity is done by rational analysis through expert judgment and evidence of construct validity is provided by factor analysis. Good assessment instruments, in addition to validity, should also be reliable. Reliable assessment instruments will give relatively the same results on each measurement, although measurement times are different.

Related to the validity of the content, it is usually supported by the calculation of the content validity index. One of the approaches to determine the coefficient of content validity is proposed by Aiken (1985, p. 132). Formula of V Aiken to calculate the index of content validity is based on the result of the assessment of several experts against an item in terms of the extent to how much the item represents the measured domain or contsruct. However, the development of measurement theory shows that the validity of the contents of an assessment instrument can also be obtained through alignment tests between assessments with standards in the curriculum. Ananda (2003) and Bhola, Impara, and Buckendahl (2003) stated that the alignment test results can be used as evidence of content validity. Until now, there are very little informations about the alignment test related to the assessment with the standards set by the government in the curriculum in Indonesia, especially in learning mathematics.

In Indonesia, various statistics' essay tests developed by researchers in general have been supported by evidence of the quality of the instrument related to content validity through expert judgments such as the ones developed by Buhaerah (2010), Hanjarwati and Wiyarno (2015), and Effendi and Farlina (2017). However, not many of various studies that contain statistics' essay test in Indonesia nowadays have been supported by the evidence of content validity using index alignment. Many of them also have not been supported by evidence of construct validity and item characteristics by using relevant measurement theory such as Item Respons Theory (IRT) for the polytomous data. Application of measurement theory in statistics tests have been conducted by Guler (2014) who analyzed open ended statistics questions with many facets of Rasch model. The use of IRT polytomous to date has not been widely applied, especially in the essay test. In fact, the assessment instrument in the form of essay test in mathematics learning especially related to statistical materials is generally arranged using response format in more than two categories (polytomous). One suitable model of IRT polytomous that is used in scoring the item test response is Generalized Partial Credit Model (GPCM). The assumption in GPCM usage is that the test items have different levels of difficulty and the level of difficulty of each step is not sorted out. This is quite relevant to the assessment generally done by teachers in Indonesia by providing an essay test score that is based on the number of steps answered correctly without regard to the sequence of steps.

Evidence of Content Validity with Alignment Index

Content validity aims at exploring whether the contents of a measuring instrument is representative or not in order to measure intended performance domain (Crocker & Algina, 1986, p. 218). Sireci and Bond (2014, p. 100) state that the evidence of the validity of the contents of an instrument, especially tests can be conducted through traditional and modern approaches. Traditionally, the most common method used to prove content-based validity is through expert judgment. The evidence of content validity is supported by the magnitude of the content validity coefficient of Aiken (1985, p. 132). As for the modern, new developed approach related to the validity of the content is conducted through the test of alignment between assessment and standards. Biggs (2003, p. 14) states that it is difficult to accurately obtain student achievement information in accordance with the desired objectives when the assessment is not in accordance with the standards in the curriculum. Furthermore, according to Wiggins and McTighe (2001, p. 51), without such conformity it limits the achievement of the expected outcomes because the students will not be studying what is being assessed.

Some of the current alignment methods include: (a) Webb Method, (b) La Marca Method, (c) Survey of Enacted Curriculum (SEC) Method, (d) Bloom's revised taxonomy Method, and (e) Method of alignment Project 2061 from the American Association for the Advancement of Science (AAAS). Empirical studies show that Bloom's revised taxonomy can be used as a tool for aligning test results in a higher level of inter-rater reliability than

some other taxonomies (Nasstrom & Henriksson, 2008). Developments in alignment studies indicate that Bloom's revised taxonomy method for testing alignment between assessment and standards in the curriculum can be modified with the Porter model in terms of calculating its alignment index. Alignment index ranges from 0 (no alignment) to 1 (perfect alignment). Nasstrom and Henriksson (2008) conducted a study of alignment between assessment and standards in the curriculum by using Bloom's revised taxonomy of the associated cognitive complexity. The formula of alignment index of Porter (*P*) is as follows.

$$P = 1 - \frac{\sum_{k=1}^{k} \sum_{j=1}^{j} |a_{jk} - b_{jk}|}{2}$$

Where: *j* is the number of rows, *k* is the number of columns in each matrix *X* and *Y*, a_{jk} and b$_{jk}$ is the ratio in cells in row *j* and column *k* for each *x* and *y* ratio matrix. Research on alignment studies by Nasstrom and Henriksson (2008) in Sweden shows that Bloom's revised taxonomy is the best model to prove harmony especially in mathematics subjects.

Item Response Theory Polytomous GPCM

Several models of the proposed item response theory polytomous are: Nominal Response Model (NRM), response model for multiple-choice items, Rating Scale Model (RSM), Partial Credit Model (PCM), Graded Response Model (GRM), sequential model for ordered response, and the Generalized Partial Credit Model (GPCM) (Van der Linden & Hambleton, 1997, p. 30). Thorpe and Favia (2012) stated that the assumptions that must be met in the analysis of test items based on IRT polytomous are sample size and unidimensionality of data. The sample size in IRT polytomous according to Reeve and Fayers (2005) is at least 250, but a sample size of about 500 is preferable for the accuracy of parameter estimation.

GPCM is one of the suitable models used to learn the characteristics of test items used in Indonesia. This is because the math test items in Indonesia are generally scored using a partial credit system that is the answer to each settlement step to the right answer is appreciated and the level of difficulty of each step is not sequenced. The general form of GPCM is stated as follows (Muraki, 1993, p. 351-352).

$$P_{jk}(\theta) = \frac{exp\left[\sum_{v=1}^{k} Z_{jv}(\theta)\right]}{\sum_{e=1}^{m_j} exp\left[\sum_{v=1}^{e} Z_{jv}(\theta)\right]}$$

$$\text{and} \quad Z_{jv}(\theta) = Da_j(\theta - b_{jv}) = Da_j(\theta - b_j + d_v)$$

Where: $P_{jk}(\theta)$ is the probability of a participant with ability θ who obtains k score category on item j, D is a scaling constant that puts the trait (θ) scale in the same metric as the normal ogive model (D=1.7), a_j is a slope parameter, b_{jh} is an item-category parameter, b_j is an item location parameter, d_v is a category parameter, m_j+1 is number of item in j, and *D* is the scale

factor (D=1.7). Estimation of item parameters and ability on IRT politomus can be done with the help of Parscale software from SSi (Muraki & Bock, 1997).

Based on the item response theory, the function of the item information states the strength or contribution of the item in uncovering the latent trait measured by the test. The function of the item information on the item response of polytomous is given by Samejima (Muraki, 1993) as follows.

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{m_j} \left|T_c - \bar{T}_j(\theta)\right|^2 P_{jc}(\theta)$$

Where: $\bar{T}_j(\theta) = \sum_{c=1}^{m_j} T_c P_{jc}(\theta)$, $I_j(\theta)$ is the information function of item j, D is a constanta which can have value of 1 or 1.7, a_j is a slope parameter of item j. $\bar{T}_j(\theta)$ item response function for a polytomous-scored item. Based on the value of the function, the item information can be determined by the function value of the test information ($I(\theta)$) and the estimated value of Standard Error Measurement ($SEM(\hat{\theta})$) with the following formula (Hambleton, Swaminathan, & Rogers, 1991, p. 94).

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta) \quad \text{and} \quad SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

This study is using alignment index and IRT polytomous GPCM in order to make sure that essay test produced in this study can accurately measure the achievement of students' knowledge competencies on Junior High School statistical materials. Theoretically, the test of quality includes content validity through expert judgment supplemented by V Aiken's index calculation and alignment index. The empirical evidence includes: (a) testing of construct validity using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), (b) determining item characteristics, information function, and standard error measurement using IRT polytomous GPCM approach; and (c) determining Construct Reliability coefficient and reliability the test by its internal consistency with Cronbach's Alpha.

Method

Research Design

The present study used the development research and descriptive survey. In this study, the alignment index and polytomous item response theory were used in order to analyze the answers given to nine items statistics' essay test.

Research Sample

Participants in the trials were 688 Junior High School students of class VII in Yogyakarta city of Indonesia that applied Curriculum 2013. Selection of school was done by purposive cluster sampling technique which was based on the category of school and certain considerations. The tests in this study were general and aimed at all students with low, moderate or high ability. Therefore, trials were conducted on three junior high schools representing high, medium, and growing qualities as well as representing public and private schools. Clasification of school quality was done based on the results of the National Examination of 2015/2016 academic year lessons. The selected school were SMP N 5 Yogyakarta which represented high quality, SMP IT Abu Bakar Yogyakarta which represented medium quality, and SMP Muhammadiyah 2 Yogyakarta which represented growing quality.

Procedures

The arrangement of statistics' essay test in this study used modification of the instrument development model from Wilson (2005, p. 18-19) and the development of test instruments from Oriondo and Antonio (1998, p. 34). Procedures of this study involved three steps namely (1) preparation of preliminary test, (2) trials, and (3) interpretation of trial results. Preliminary tests were designed by referring to the basic competencies, indicators, and contents of Bloom's revised taxonomy to each item as presented in Table 1.

Table 1

Basic Competencies, Indicators, and Bloom's Revised Taxonomy

Basic Competencies	Indicators	Item	Bloom's Revised Taxonomy
Analyze the relationship between data and the way of presentation (table, line graph, bar chart, and pie chart).	1. Describes various ways of collecting data.	1a	B3
	2. Describes various ways of presenting data.	1b	B3
	3. Presents data using tables.	2a	C3
	4. Analyzes the relationship between the data presented in tabular form.	2b	C4
	5. Presents data using bar charts.	3a	C3
	6. Presents data using pie charts.	3b	C3
	7. Analyzes the relationship between the data presented in the form of bar and circle diagrams.	3c	C4
	8. Presents data using line graphs.	4a	C3
	9. Analyzes the relationship between the data presented in the form of a line graph.	4b	C5

Description: B3: Conceptual knowledge and cognitive processes "apply", C3: Procedural knowledge and cognitive processes "apply", C4: Procedural knowledge and cognitive processes "analyze", C5: Procedural knowledge and cognitive processes "evaluate".

Data Analysis
Content Validity

Test validation was performed through expert judgment. Experts conducted a qualitative test review and provided an assessment of the suitability between the item with the indicator in the form of a Likert scale with five answer options. In addition, to strengthen the evidence of content validity for the purpose of alignment tests, the review sheet also featured a format for the assessment of the expert on the revised content of Bloom's taxonomy on each test item. Based on the assessment of experts, in addition to qualitative assessment related to feasibility of the test, V Aiken's index was also calculated. The formula of V Aiken is as follow:

$$V = \frac{s}{[n\,(c-1)]}\,, \quad s = \Sigma\, n_i\,(r_i\text{-}l_o)$$

Where: V is Aiken validity index, n_i is the number of experts who choose the criteria of i, r_i *is* the criteria of i, l_o is the lowest rating, n is the number of expert, and c is the number of rating.

The valid criterion of an item is to compare the value of V calculated with V value, that is the minimum value of the content validity index based on the number of rater in V table Aiken (1985). The number of raters in this study was six and the number of ratings was five then the minimum index of content validity based on table V Aiken was 0.79. The result of the expert judgment was a proof of theoretical quality of the instrument. The empirical evidence was obtained from the trial and interpretation of test results.

Construct Validity

To obtain evidence of construct validity, factor analysis using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were used. Using the EFA approach, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO MSA) and Barlett Sphericity tests were used to determine whether the items' test matched the factor analysis or not. The match criteria were a minimum KMO MSA score of 0.50 and statistically significant Barlett Sphericity test results (Hair et al., 2010). This study extracted factors using Principal Component Analysis (PCA) conducted using SPSS version 20.0. Using CFA approach, data analysis began with the requirements analysis test in order to determine whether the data already met the requirements to be analyzed by CFA technique, which required to test the model by using the joint multivariate normal distribution. In this study, CFA was conducted with Structural Equation Modeling (SEM) by using Lisrel program version 8.51. The criteria used were if p value

was >0.05; then, the distribution is normal, and if p value was ≤0.05; then, the distribution is not normal (Yamin & Kurniawan, 2009).

After the requirements analysis test, data analysis was performed to verify the validity of scale constructs through first order CFA. The criteria for a valid indicator in representing the construct were if t value was>1.96 and the value of Standarized Loading Factor (SLF) was at least 0.3 (Hair et al., 2010; Igbaria et al., 1997). The criteria of sample size in SEM according to Comrey and Lee (1992) cited by MacCallum et al. (1999, p. 840) are: 100 (poor), 200 (fair), 300 (good), 500 (very good), and ≥1000 (excellent). This study involved 688 participants; thus, the sample size of the study belonged to the very good category.

Based on the data of the test results, the validity of the construct was conducted through the first order Confirmatory Factor Analysis (CFA) with the help of Lisrel software. The criterion of the validity of an item in representing the construct was the value of t value>1.96 and the value of Standarized Loading Factor (SLF) of at least 0.3 (Hair et al. 2009, p. 119; Igbaria et al., 1997, p. 290). The model fit criteria used were Root Mean Square Error of Approcimation (RMSEA) between 0.03 to 0.08, p value>0.05, Goodness of Fit Index (GFI)≥0.90, Adjusted Goodness of Fit Index (AGFI)≥0.90 (Hair et al, 2010, p. 641-644).

Those criteria are based on Garson (2009) which state that support for the fit of the model developed through empirical data can be seen at least from three compatibility measures representing three different fit model categories. The three categories of fit model are absolute fit measures, incremental fit measures, and parsimonious fit measures. If two of the three categories meet the criteria, then the developed model matches the data. Therefore, the criteria for model fit used in this study were RMSEA, p value, GFI which represented absolute fit measures, and AGFI which represented incremental fit measures. The fulfillment of unidimensionality assumptions was also seen from the plot of eigenvalue (Hambleton, Swaminathan, & Rogers, 1991, p. 56). Naga (1992, p. 297) stated if eigenvalue of the first factor several times the eigenvalue of the second factor, while the eigenvalue of the second factor and above are almost the same, it can be said that the unidimensional requirement has been fulfilled.

Characteristics of Test Items

Characteristics of test items were obtained through item analysis based on IRT polytomous GPCM with the help of Parscale version 4.1. Based on the results of Parscale analysis, three characteristics items were obtained; namely, the estimation of discrimination parameters (*a*), location parameters (*b-global*), and a set m-1 parameter of difficulty level (*b*). The b_{jk} parameter was obtained by subtracting the parameter value b-global$_j$ with the parameter value d_{jk}. The resulting graphics included Item Characteristic Curve (ICC), Item Information Curve (IIC), and Test Information Curve (TIC). The good

item criteria were: (1) the parameter values $a>0.25$ on the logit scale, and (2) *b-global* parameter values ranged from -3 to 3 on the logit scale (Wells, Hambleton, & Purwono, 2008).

Reliability

One popular method in psychometry which is often used to determine reliability based on internal consistency is the alpha coefficient from Cronbach. Coefficient alpha can provide a reliability estimate for a measure composed of items scored with values other than 0 and 1 (Cronbach, 1951). The formula for estimating the reliability of essay test scores uses the basic coefficient alpha is as follows (Ebel & Frisbie, 1991, p. 85).

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right)$$

Where: k is the number of separately scored essay test questions, s_i^2 is the variance of students'scores on a particular item, $\sum s_i^2$ is the sum of the item variances for all test items, and s_t^2 is the variance of the total essay scores.

A good reliability criterion is a minimum of 0.7 (Nunnally, 1981, p. 245). As according to Ebel & Frisbie (1991, p. 86) if the test is used as a standard test, the reliability coefficient should be between 0.85-0.95 while for the minimum class, it is not lower than 0.65. Kayapınar (2014, p. 114) stated that reliability coefficient value might be more accurate and reliable if the accepted interpretation of a meaningful correlation coefficient for this kind of measurements can be considered as .90 minimum for giving evidence of reliable ratings. In addition to Alpha's Cronbach, on the use of SEM can also be obtained Construct Reliability (CR). The formula for calculating CR is as follows (Wijanto, 2008).

$$Construct\ Reliability = \frac{(\sum std.loading)^2}{(\sum std.loading)^2 + \sum e_j}$$

Where: std. Loading is Standarized Loading Factor (SLF), e is error variances. Hair, et al (2010) suggests that the estimation of CR>0.7 is good, while CR between 0.6 and 0.7 is acceptable, provided that the construct validity indicator is good.

Results

The validity of test content was conducted through qualitative and quantitative analyses. Qualitative results were statements by experts who claimed that the test was feasible and ready for use. Quantitative results included the assessment of experts on the suitability of the item with the indicator and the revised charge of Bloom's taxonomy. Based on the result of calculation of V Aiken's index as presented in Table 2, it was found

that V of test items were 0.88-1.00. This meant that all test items had good content validity in terms of their conformity with the indicator.

Table 2

V Aiken's Index Calculation Result

Item	s	n	c-1	V_{table}	V
1a	21	6	4	0.79	0.88
1b	21	6	4	0.79	0.88
2a	24	6	4	0.79	1.00
2b	24	6	4	0.79	1.00
3a	22	6	4	0.79	0.92
3b	22	6	4	0.79	0.92
3c	22	6	4	0.79	0.92
4a	23	6	4	0.79	0.96
4b	23	6	4	0.79	0.96

Evidence of the validity of this content was also supported by the calculation of Porter alignment index that was equal to 0.93. The magnitude of the alignment index was included in the category of excellent so it can be said that the test had a very good alignment with the standards set out in Curriculum 2013 in Indonesia. The test that was declared as eligible by the experts was then tested on the trial with 688 students.

Based on trial, EFA was conducted in order to identify the factors that made up the test. Based on SPSS version 20.0, the value of Barlett Test of Sphericity was 3105.039 with 0.000 significance level. This showed a significant correlation between variables. The calculation result of KMO MSA was 0.729 which indicated that the adequacy of the sample was good. Factor extraction was done using PCA method. Based on the extraction results, three factors forming the test constructs were obtained. The findings related to the factors are given in Table 3.

Table 3

Findings Related to Factors Obtained as a Result of the Principal Component Analysis

Factor	Eigenvalue	Variance Percentage	Total Variance Percentage
1	3.745	41.606	41.606
2	1.684	18.713	60.319
3	1.062	11.801	72.120

Based on Table 3, it can be seen that there were three factors with an eigenvalue bigger than 1.00. All three factors explained around 72.120% of the total variance. The first factor described 41.606% of the total variance. The contribution of the factors to the total variance percentage decreased after the first factor. Eigenvalue of the first factor was 3.745, more than twice from eigenvalue of the second factor. That means the test only measured one ability. This situation can be seen in eigenvalue graph on Figure 1.
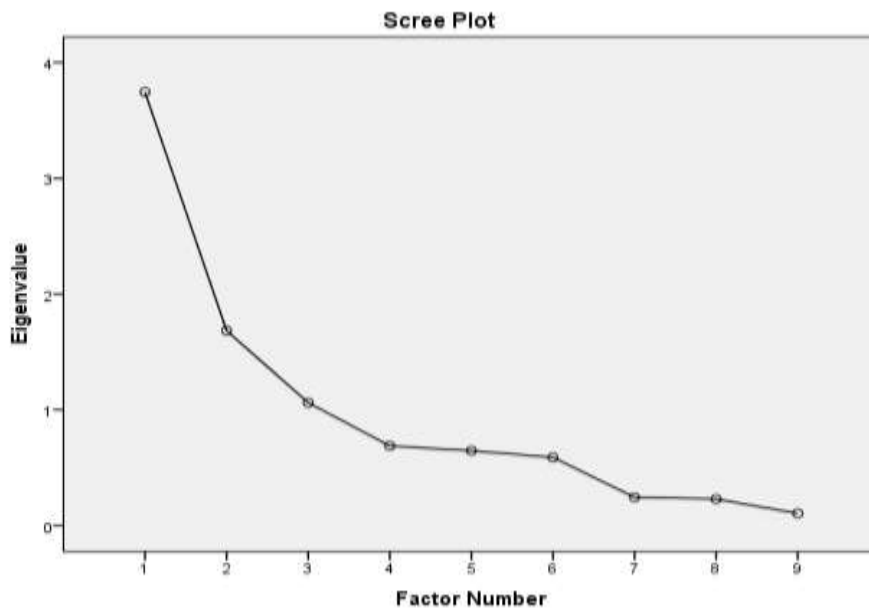
Figure 1. Eigenvalue Graph

For the next step, CFA was conducted in order to confirm that items really explained the test. The result of preliminary analysis showed that the data did not have multivariate normal distribution, so CFA second order analysis was done with Weighted Least Square (WLS) estimation model. Based on Lisrel analysis results, RMSEA value of 0.031, Chi Square of 20.11 with p value 0.06, GFI of 0.99, and AGFI of 0.98 were obtained. These results indicated that the suitability of the model was met. Here were CFA's fisrt order results for t value and Standarized Loading Factor (SLF) values as well as CR and Alpha coefficients.

As shown on Table 4, all of the items were significant in supporting the test constructs with the lowest support by item 1a and the highest by item 4a. It showed that the

construct validity of the test was good. Test reliability was good with a CR coefficient of 0.88 and an Alpha coefficient of 0.78. This meant that in terms of construct, the test consisted of items that could accurately measure students' statistical skills. Based on the reliability scores, it can be said that the test was reliable which meant that the measurement results obtained through this test were consistent.

The characteristics of the test items were determined using the IRT polytomous GPCM approach. For the analysis of the data using Parscale software version 4.1., the following parameters were obtained.

Table 4

Result of First Order CFA

Item	First order CFA				Validity	Reliability	
	t value	Explanation	SLF	Error		CR	Alpha
1a	5.80	Significant	0.24	0.18	Sufficient	0.88	0.78
1b	7.46	Significant	0.48	0.19	Good		
2a	22.8	Significant	0.89	0.099	Good		
2b	7.78	Significant	0.33	0.54	Good		
3a	6.26	Significant	0.48	0.12	Good		
3b	14.74	Significant	0.77	0.88	Good		
3c	9.34	Significant	0.37	0.53	Good		
4a	24.38	Significant	0.99	0.0067	Good		
4b	16.58	Significant	0.44	0.84	Good		

As shown in Table 5, all the test items were categorized as good with the discrimination index (a_j) as a whole located at 0.310-2.008 and with the item difficulty index (b_j) of -2.329 to 0.475 on the logit scale. The analysis also obtained the function of test information and standard error measurement as presented in Figure 2.

Table 5

Results of Parameter Test Estimation on Trial

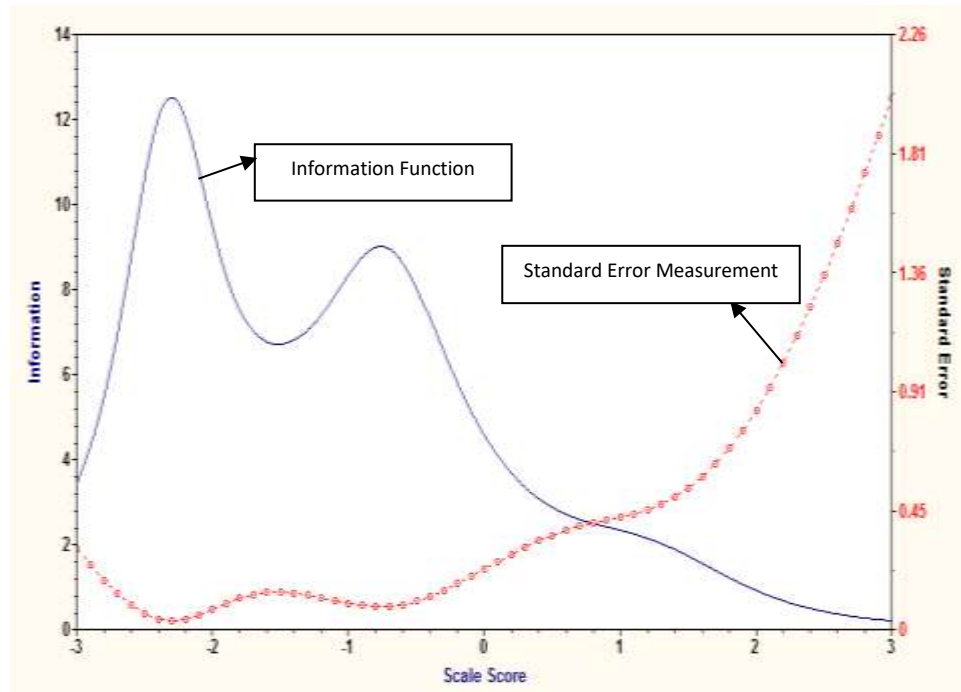| Item	a_j	b_j	d_k	b_{jk}	Explanation
1a	1.216	-1.849	0.000	-1.849	Good
1b	0.905	-2.179	0.183	-2.362	Good
			-0.183	-1.996	
2a	1.371	0.475	0.716	-0.241	Good
			-0.716	1.191	
2b	0.865	-1.086	0.171	-1.257	Good
			-0.171	-0.915	
3a	2.008	-2.329	-0.059	-2.270	Good
			0.059	-2.388	
3b	0.359	-1.691	-3.745	2.054	Good
			3.662	-5.353	
			1.981	-3.672	
			-3.214	1.523	
			1.316	-3.007	
3c	0.813	-0.977	0.008	-0.985	Good
			-0.008	-0.969	
4a	1.314	-0.715	-0.353	-0.362	Good
			0.353	-1.068	
4b	0.310	0.010	0.575	-0.565	Good
			0.545	-0.535	
			-1.121	1.131	

Figure 2. Information Function and Standard Error Measurement

Figure 2 showed that the test provided accurate information for students with ability estimate (theta) of -3.0 to 0.8. Even tests also provided accurate information for students with theta less than -3.0. The test provided the highest information for students with tetha around -2.4.

Discussion, Conclusion and Recommendations

Discussion

Instrument of assessment of knowledge aspect in the form of essay test produced by this study was proved empirically and theoretically to be in a good quality. Based on the validity of the contents and the validity of the construct, the entire test items were of good quality. The content validity index of V Aiken on the overall test item was 0.88-1.00 and the Porter alignment index was 0.93. This level of alignment belonged to very high category. This is because the test has been compiled based on Bloom's revised taxonomy

as a reference to the applicable curriculum. The results of the alignment test strengthened the evidence of the validity of the content because it showed the suitability of the instrument items with the standard in the curriculum used.

The use of alignment index in this study will serve as a reference related to the validity of the contents of an assessment instrument. This study is an important one as revealed by Tindal (2005) who states that the results of alignment studies can be used to determine whether restructuring of the assessment is necessary or not. The use of alignment index on this study has provided more in-depth information about the quality of the test, which means that the test has alignment with the standards in the curriculum. It would be beneficial to use alignment index in other assessment instruments, especially in the form of tests, because information about alignment between assessment and standards is certainly very useful in policy making related to assessment and education. In addition, further development regarding alignment testing in assessment of mathematics activities can also be achieved in alignment testing between standards in curriculum and handbooks used by students. This is also suggested by Hasmi, Hussain, and Shoaib (2018) who reviewed the alignment between curriculum of mathematics and textbook using the SEC method.

All test items supported the test constructs. From the point of reliability coefficient, the test reliability also belonged to the good category. The test had good reliability with a CR coefficient of 0.88 and an Alpha coefficient of 0.78. It showed that the test had a high consistency of measurement results. However, one should pay attention to various factors affecting reliability in the implementation of this instrument for a large scale. As stated by Ebel and Frisbie (1991), if the test is used as a standard test, then, the reliability coefficient should be between 0.85-0.95 when used for the minimum class of 0.65. If test is used on a large scale and is intended for crucial decision, it should be standardized and has a reliability coefficient between 0.85-0.95. The tests produced in this study were more suitable for assessment within the scope of the class.

The quality of the test instrument in terms of its item characteristics was good. All test items were categorized as good with a discrimination index of 0.310-2.008 and item difficulty index of -2.329 to -0.475 in logit scale. From the points of the test information function and standard error measurement, the test provided accurate information on theta -3.0 to 0.8 and the highest information on the theta of about -2.4. It meant that the tests produced in this study were appropriate to be used for students with moderate and lesser abilities.

Conclusion and Recommendations

The essay test produced by this study was of good quality proven by several theoretical evidence supports and accurate empirical evidence. The result of this study is very useful, especially for Junior High School teachers because the result of the test can

reveal achievement of students' statistical knowledge competence appropriately. In addition, for researchers and other educational practitioners, the result of this study is very important as it supports references regarding the quality of an assessment instrument. Further studies need to be conducted in other areas with larger and more diverse participants so as to further generalize and provide evidence to the quality of the instruments produced in this study. The alignment study also needs to be conducted with other methods such as Webb, SEC, or with other approaches such as online systems. Further studies can also be carried out by examining the use of the proposed item response of polytomous theory such as NRM, the response model for multiple-choice items, RSM, PCM, GRM, and sequential models for ordered responses in determining the characteristics of a test.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142.

Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd.

Anderson, D., Irvin, S., Alonzo, J., & Tindal, G. A. (2015). Gauging item alignment through online systems while controlling for rater effects. *Educational Measurement: Issues and Practice*, 34, 22-33.

Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.

Bhola, D. S., Impara, J. C., & Buchendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice,* 22(3), 21–29.

Biggs, J. (2003). *Teaching for quality learning at university.* Glasgow: The Society for Research into Higher Education & Open University Press.

Buhaerah. (2010). Pengembangan perangkat pembelajaran berdasarkan masalah pada materi statistika di kelas IX SMP. Gamatika. Nomor 1. Nopember.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont: Wadsworth Group.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement.* USA: Prentice-Hall Inc.

Effendi, K. N. S. & Farlina, E. (2017). Kemampuan berpikir kreatif siswa SMP kelas VII dalam penyelesaian masalah statistika. Jurnal Analisa. 3(2). 130-138.

Garson, G. D. (2009). *Overview structural equation modeling*, http://faculty.chass.ncsu.edu/garson/PA765/structur.htm.

Guler, N. (2014). Analysis of open-ended statistics questions with many facet Rasch model. *Eurasian Journal of Educational Research*, 55, 73-90. http://dx.doi.org/10.14689/ejer.2014.55.5.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7thed). Prentice Hall [versi elektronik].

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.

Hanjarwati, R. & Wiyarno, Y. (2015). Pengembangan bahan ajar matematika (materi statistik) dengan menggunakan model active learning sistem 5 M untuk siswa kelas VII. *Jurnal Teknologi Pembelajaran Devosi*. 5(2).

Hasmi A., Hussain T., & Shoaib A. (2018). Alignment between Mathematics Curriculum and Textbook of Grade VIII in Punjab. *Bulletin of Education and Research,* April 2018, 40(1), 57-76.

Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A. L. M. (1997). Personal computing acceptable factors in small firms: A structural equation model. *MIS Quarterly*, September, 279-299.

Kayapinar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136, http://dx.doi.org/10.14689/ejer.2014.57.2.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement,* 17(4), 351-393.

Muraki, E. & Bock, D. (2002) PARSCALE 4.1 Computer program. Chicago: Scientific Software International, Inc.

Naga, D. S. (1992). Pengantar teori skor pada pengukuran pendidikan. Jakarta: Gunadarma

Nasstrom, G. & Henriksson, W. (2008). Alignment of standards and assessment: A theoretical and empirical study of methods for alignment. *Eletronic Journal of Research in Educational Psychology,* 6(3), 667-690.

Nitko, A. J. & Brookhart,S. M. (2011). *Educational assessment of students* (6th ed.). Boston, MA: Pearson Education Inc.

Nunnally, J. C. (1981). *Psychometric theory* (2nd ed). New Delhi: McGraw-Hill Publishing Company Limited.

Oriondo, L. L. & Antonio, D. E. M. (1998). *Evaluation educational outcomes.* Manila: Rex Printing Compagny.

Reeve, B. B., & Fayers, P. (2005). Appliying item response theory modeling for evaluating questionnaire item and scale properties. Dalam Fayers, P. & Hays, E.D. (Eds), *Assessing quality of life in clinical trials: Methods of practice* (2nd ed). New York: Oxford University Press.

Sireci, S. & Bond, M. F. (2014). Validity evidence based on test content. *Psicothema,* (26)1, 100-107.

Thorpe, G. L. & Favia A. (2012). *Data analysis using item response theory methodology: An introduction to selected programs and applications.* Psycology Faculty Scholarship. Paper 20. http://digitalcommons.library.umaine.edu/psy_facpub/20.

Tindal, G. (2005). *Alignment of alternate assessments using the webb system.* Washington, DC: Council of Chief State Officers.

Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Walstad, W. B. (2006). Testing for depth of understanding in economics using essay questions. *Journal of Economic Education.* Washington: Winter.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education (Research monograph No. 6).* Washington, DC: Council of Chief State School Officers.

Wells, C. S., Hambleton, R. K. & Purwono, U. (Juni 2008). Item response theory. Polytomous response IRT models and aplicatios. *Handout* delivered on the Training of Educational Assessment and Psychology (Psychometry), at Yogyakarta State University.

Wiggins, G. & McTighe, J. (2001). *Understanding by Design (2nd Ed.).* Alexandria, VA: Association for Supervision and Curriculum Development.

Wijanto, S. H. (2008). *Structural equation modeling dengan Lisrel 8.8.* Yogyakarta: Graha Ilmu.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah: Lawrence Erlbaum Associates, Inc.

Eurasian Journal of Educational Research
www.ejer.com.tr

A Scale Development Study to Measure Secondary School Teachers' Opinions On Coaching Behaviours[1]

Ozgun OZMEN[*]

A R T I C L E I N F O	A B S T R A C T
	Purpose: Coaching is an approach that enables individuals to realize their own potential and strengths, to increase their awareness, to find themselves, and that encourages and supports the process of fulfilling goals by balancing different spheres of life. Adoption and implementation of coaching philosophy also have great importance in terms of teachers' relationship with students. In this study, it was aimed to identify the opinions of high school teachers on adoption and implementation of coaching behaviors according to the variables of gender, age, education level, branch, status of participating

in coaching training, conferences or seminars, and status of considering the display of coaching behavior significantly.

Method: The study was designed as a descriptive survey model study and employed the quantitative research method. The sample of the study composed of 362 teachers in total who worked in high schools located in districts of Ankara, in 2017-2018 academic year.

Findings: Teachers adopted all of the sub-dimensions of coaching behaviors, yet they were incompetent in terms of implementation. Female participants demonstrated a significant difference in terms of adoption and implementation of coaching behaviors. Significant differences occurred in the adoption and implementation status of the ones who participated in coaching seminars or conferences and considered the display of coaching behavior significant. In addition, there was a significant difference in vocational education high school teachers' adoption and implementation levels of the dimension "enabling coachees to make choices according to their core values" in comparison to teachers of other branches.

Implications for Research and Practice: The research findings revealed that teachers who participated in coaching training, conferences and seminars provided benefits to their students by implementing coaching skills. Therefore, it is suggested for teachers to get training from leading experts of the field. It is possible to say that new legislative regulations that will be made by the Ministry of National Education by including coaching education to their in-service trainings, and inclusion of coaching implementations in educational institutions will make a positive contribution to teachers' personal and professional, and students' academic and individual development.

Introduction

Nowadays, there is a rapid change in every sphere of both daily and professional lives of people. While some of these changes have a positive impact on people, some may have a negative impact as well. In order to cope with this rapid change occurring in their lives, people search for opportunities. In this context, one of the methods that people apply is getting coaching support. Coaching can be expressed as an approach that enables individuals to realize their own potential and strengths to increase their awareness, to find themselves, and that encourages and supports the process of fulfilling goals by balancing different spheres of life. Provision of coaching for teachers will make a positive contribution both to teachers' professional development and students' educational and instructional success. The study aims to identify the impacts of teachers' coaching behavior both on themselves and students in education-instruction environments.

In the literature, there are plenty of definitions for the term 'coaching'. Coaching is a method that is used in different environments and fields such as the health sector, military, workplace and personal life for a different variety of purposes (Engel, 2011; Frisch, 2013; Jarrett, 2013; Newnham-Kanas, Irwin & Morrow 2008). The processes of questioning and giving feedback developed by Socrates 2400 years ago are among the first efforts in the field of coaching (Nielsen & Norreklit, 2009). Socrates can be considered as the first coach of the history and the expression of Socrates "I cannot teach anybody anything. I can only make them think" can be provided as its proof (Wilson, 2008).

Coaching is a term, which has been defined by a number of theorists and researchers differently. According to the definition of Poussard (2003), coaching is teaching someone to learn by himself/herself instead of providing information. In a certain sense, coaching is providing guidance. Individuals have enough resources to perform, create or realize a task. In this case, what is important is to be the one who is a 'guide', not an 'enforcer' (Uckun & Kilinc, 2007). The coachee who receive coaching service improves effective communication skills, adapts to different situations more easily, and acquires more flexible and moderate behavioral patterns in various spheres of life. The time and management skills of an individual also develop. In addition to that, it becomes a tool that helps to learn from the mistakes. The contributions of an individual to the workplace increase and the individual takes responsibility for his/her decisions (Barutcugil, 2006). Coaches recognize the obstacles that a coachee confronts by making observations in this process, and by reflecting the expressions of a coachee. In this way, coaches are able to determine the sources and obstacles that a coachee has, and conditions that a coachee already has resistance to. If required, these resistance points are reviewed by the coach to determine the direction of a coachee (Griffiths, 2005).

Coaching encourages individuals to think and question, supports them to raise awareness in different spheres of life about themselves, and supports them to find unique solutions and improve these solutions. As Griffiths (2005) states, coaches are the individuals who focus on helping their coachee in different aspects of life, and who

help to eliminate obstacles without adding new ones during the coaching process. Coaches help their coachee to reach their objectives by monitoring their questioning, thinking, action planning, and effort and assessment strategies. Furthermore, they aim to enhance individuals' potential and learning skills by increasing their motivation and asking strong questions (Gonzalez, 2008; Lawler, 2011; Sue - Chan & Latham, 2004). As published by ICF [International Coach Federation] (2018) it can be said that the explanation which highlights that coaching is a type of relationship which a coach and coachee are in cooperation within the process encouraging coachee to think more is an integrative expression for all definitions.

Purpose of Coaching

Coaching meetings help a coachee to clarify an objective and determine action steps towards reaching this aim. According to Çetin (2015), the main purpose of coaching is to enable a coachee to improve the skill of realizing an objective, and to be able to manage this process. The mental condition and action steps are exposed by a coach by asking strong questions. In the advancing process, plans are made to realize actions. Coaching is a medium that lightens the unknown points for the purpose of providing preferences which are conscious, creative, imaginative, and related to behaviors. The main purpose of coaching is to increase performance and maintain the cognitive, affective and behavioral changes (Douglas & McCauley,1999). By means of the coaching process, the emotional obstacles that prevent a coachee from reaching an objective can be converted into motivation (Curtis & Kelly, 2013). In this way, coachees can reach their aim, increase their self-efficacy, and enhance their self-confidence (Grant, 2006). It is important to carry out a systematic, scheduled process so as to reach the aim of coaching.

In coaching practices, models such as Single Loop Coaching Model, Double Loop Coaching Model (O'Connor & Lages, 2004), SMART Goals (Doran,1981), Self-Regulatory Loop Model (Grant,2003), RE-GROW Model (Greene & Grant, 2003) are used. The most common model is the GROW model which consists of the first letters of the words goal, reality, options, and wrap up. This model which was developed by Graham Alexander in 1980s, and afterwards popularized by Sir John Whitmore (Whitmore, 1992), separates the coaching session into four interrelated phases as presented in Table 1 (Grant, 2011).

Table 1

GROW Model

Abbreviation	Definition	Example Questions
G – Goal	Coachee is asked to clarify what s/he wants to achieve from each session. Determines the focus of coaching.	What do you want to achieve this session? How would you like to feel afterwards? How would be the best use of this time?
R – Reality	Raise awareness about present realities. Examine how current situation is impacting coachee's goals.	How have things gone in the past week? How have you handled any problems? What worked? What didn't work?
O -Options	Identify and assess available options. Encourage solution-focused thinking and brainstorming.	What possible options do you have? What has worked for you in the past? What haven't you tried yet, what might work?
W - Wrap-Up	Helps the coachee to determine next steps. Develops an action plan and generates motivation.	What is the most important thing to do next? What might get in the way? Who might be able to support you? How will you feel when this is done?

Sources: (Grant & Greene, 2004; Landsberg, 1999; Spence & Grant, 2007; Whitmore, 1992).

Each step of the Grow Model was defined in Table 1, and the sample coaching questions on these steps were presented in the article of Grant which was published in 2011 in ''The Coaching Psychologist'' journal under the title of ''Is it time to REGROW the GROW model? Issues related to teaching coaching session structures'' (Grant & Greene, 2004; Landsberg, 1997; Spence & Grant, 2007; Whitmore, 1992). It can be argued that GROW Model is an effective session in terms of a coachee's process of determining and structuring a goal.

Coaching Skills

There are a number of important coaching skills that a coach should have during the coaching meetings. According to the Coaching Skills and Strategies Model that was developed by Creswell (1998), five main skills of coaching are effective communication, questioning, raising awareness, planning, and supporting. In the framework of the emotional environment established during the coaching meetings, a coach should listen to a coachee carefully by considering the principle of confidentiality, be able to ask effective questions to provide feedback, help to determine the concrete steps, and cope with the behaviors and emotions of a coachee such as being resistant and self-sabotaging (Curtis & Kelly, 2013; Griffiths, 2005). According to Wilkins (2000), coaches should possess effective communication skills and strategies. Effective communication skills include a coach's skills of listening, questioning, using intuitions, establishing dialogues, and having a mutual interaction with a coachee. Moreover, effective communication skills include strategies such as that a coach should possess awareness, focus, support, and challenges. International

Coaching Federation (ICF) grouped the core competencies as setting the foundation, establishing a relationship simultaneously and co-creating the relationship, communicating effectively, facilitating learning and results (ICF, 2018).

Education Coaching

Education coaching is commonly used particularly in developed countries (USA, England, Australia), however, it can be asserted that there is a limited number of studies conducted on education coaching in Turkey; and therefore, the issue has been underestimated in educational institutions. The explanation of learning and its measurement varies according to the different philosophical theories and psychological studies. However, the common point of these theories on learning process is the eventual change which will be reached as a result of this process. This change can be achieved in terms of people's behaviors, attitudes, skills or knowledge (Jones, 2017). According to Griffiths and Campbell (2009), learning occurs through experience, and result in a permanent or long-term change in knowledge, skills and behaviors of both the coach and coachee. According to Gordon (2008), the responsibility of a teacher in this journey is to maintain this relationship by establishing a special bond with a student for ensuring a productive process. While teaching is carried out by a single individual in this process, the responsibility of learning belongs to another individual.

In the case of coaching implementation in educational environments, coaching will have positive contributions to teachers and students as a result of the learning outcomes. Knight (2006) describes coaching practices in education as a method that might enable eliminating ineffective vocational education methods. Because, there is a cooperation between a teacher and a professional development specialist which is not based on an assessment in coaching. In this cooperation, the primary purpose of both sides is to learn together, improve lecturing style, and enhance student development. In this sense, Gynnild (2007) states that education coaching aims at improving students' creativity and research related skills, making them active in the learning process, and preventing rote learning by filling the gaps in the education system.

From the perspective of teachers, coaching is a professional training strategy that focuses on enhancing the quality of education and student success by putting an emphasis on practices of oneself or other teachers, and provides one-to-one learning opportunities (Loucks-Horsley, Love, Stiles, Mundry & Hewson, 2009). According to Druckman and Bjork (1991), coaching is based on observing students, providing tips, feedback and new tasks to students; reminding them something or drawing their attention to an important issue. They indicated that the purpose of all these actions is to approximate student performance level to an expert performance level as much as possible. Neufeld and Roper (2003) stated that coaching field knowledge is more effective than implementing a curriculum.

A successful coaching implementation has a positive impact on students thinking skills and academic success. In this direction, it can be argued that including coaching practices in educational environments are fundamental for the development of students and teachers. When the relevant literature is reviewed in the framework of

Turkey, no study that addresses the coaching behaviors of teachers comprehensively as in the present study can be found in the field of educational sciences. In this sense, it is believed that this study will be beneficial for the studies conducted in MoNE, teachers, school administrators, and academicians. The purpose of this study is identifying the opinions of teachers who work in public high schools within the borders of Ankara province on the status of adoption and implementation of coaching behaviors. In line with the general purpose of the study answers were sought for the questions given below:

1) What are the opinions of teachers who work in public high schools within the borders of Ankara province on the status of adoption and implementation of coaching behaviors concerning the four dimensions of coaching as a) Effective listening b) Asking strong questions c) Raising awareness d) Enabling coachees to make choices according to their core values?

2) Do the opinions of teachers who work in public high schools within the borders of Ankara province on the status of adoption and implementation of coaching behaviors show a significant difference according to the variables of gender, age, education level, branch, status of participating in coaching training, conferences or seminars, and status of considering the display of coaching behavior significant?

Method

Research Model

The study which was conducted with the purpose of identifying the opinions of teachers who work in public high schools within the borders of Ankara province on the status of adoption and implementation of coaching behaviors was a quantitative study and designed as a general survey model. General survey models are surveying organizations that are conducted on a population consisting of a large number of components to make a judgment on the entire population or on a sample or group that will be taken from the population (Karasar, 2005). In the study, a quantitative research model was employed.

Population and Sample

The population of the study consisted of teachers who work in high schools located in the central districts of Ankara (Altındağ, Çankaya, Etimesgut, Gölbaşı, Keçiören, Mamak, Pursaklar, Sincan, Yenimahalle), in the 2017-2018 academic year. The study was conducted with the participation of 665 teachers in total, including 303 teachers in the pre-implementation phase, and 362 teachers in the post-implementation phase. The stratified sampling method was employed in sample selection. In the selection of schools, in which research instruments were implemented, random sampling method was used. The number of high school teachers who worked in high schools located in the central districts of Ankara was provided by Ankara Provincial Directorate of National Education, Statistics Department for the years of 2016-2017, and the number of teachers taken for sampling (according to districts) are provided in Table 2.

Table 2

The Number of Teachers Working in Public Schools According to the Districts

	Districts	Number of Teachers	Number of Teachers Included in the Sample
1	Altındağ	456	26
2	Çankaya	1936	112
3	Etimesgut	609	35
4	Gölbaşı	197	11
5	Keçiören	847	49
6	Mamak	615	36
7	Pursaklar	95	5
8	Sincan	452	26
9	Yenimahalle	1057	61
	Total	6264	362

Source: Ankara Provincial Directorate of National Education, Statistics Department, 2017 (www.ankara.meb.gov.tr)

Table 3

Demographic Information on the Teachers

Variables		n
Gender	Female	277
	Male	135
Age	25 years old and below	22
	26-34 years old	59
	35-44 years old	133
	45-54 years old	166
	54 years old and above	32
Education Level	Associate Degree	5
	Bachelor Degree	334
	Master's Degree	67
	PhD	6
Branch	Social Sciences	54
	Science	52
	Turkish Language - Literature	62
	Mathematics	50
	Foreign Language	73
		59
	Other branches	50
Status of participating in coaching training, conferences or seminars etc.	Yes	87
	No	325
Status of considering the display of coaching behavior significant.	Yes	97
	No	3

As presented in Table 3, 67.2% of the teachers were female and 40.3% of them were between 45-54 years old. 81,1% of the teachers had bachelor degree, 1,2% of them had an associate degree, 16.3% of them had a master's degree and 1.5% had a PhD. When teachers' branches were examined it was seen that 13,5% of them were in social sciences branch, 13% of them were in science, 15.5% of them were in Turkish language-literature, 12.5% of them were in mathematics, 18.3% of them were in in foreign language, 14.8% of them were in vocational education and 12.5% of them were in other branches (e.g. art, music, physical education etc.). 78.9% of the teachers did not participate in a coaching training, conference or seminar, and 56.3% of them participated in a coaching training, conference or seminar for 2-3 times. 97% of the teachers considered the display of coaching behavior significant.

Development of the Data Collection Tool and Implementation

For the purpose of the study, "High School Teachers' Opinions on Coaching Behaviors Scale" developed by the researcher was used. In the development phase of the scale, first of all, a literature review was conducted, and national and international studies were examined. Afterwards, an item pool was created for the scale. The scale draft was presented to experts, and re-organized according to the suggestions. Then, a pre-implementation was carried out with 303 high school teachers in order to test the suitability for comprehensibility and clarity principle of the items included in the scale.

Kaiser-Meyer-Olkin (KMO) and Bartlett Sphericity Test results were examined to determine the suitability of the collected data to the factor analysis. The results such as the KMO value of .949, and Bartlett Test result which was found significant ($p<0.01$) proved that the data were suitable for factor analysis. In order to demonstrate the factor structure of the scale, first of all, an exploratory factor analysis was performed.

As a result of the analysis, it was found out that "Coaching Behaviors Scale" was consisted of four dimensions as Effective Listening, Asking Strong Question, Raising Awareness, and Enabling Coachees to Make Choices According to Their Core Values. The total variance that the four dimensions explained together was 52.46%.

The total Cronbach Alpha reliability coefficient and Cronbach Alpha reliability coefficient concerning the factors were calculated respectively as .829, .929,.897 and .941. According to these values, it was possible to argue that the construct validity of the scale was ensured, and the data collected through the scale were reliable.

In order to verify the structure which consisted of four factors through factor analysis, confirmatory factor analysis (CFA) was performed (Brown, 2006). According to the analysis results, fit indexes were calculated as $\chi 2 = 2204.88$, p = 0.0000, sd=1115, $\chi 2 / sd = 1.97$, CFI=0.98, NFI = .96, NNFI=.98, GFI=.77, AGFI=.74, IFI= .98, RFI= .95 ve RMSEA= .057. When all the findings were examined together, it was possible to state that the structure concerning the measurement tool was verified according to the sample size, and the scale was structurally and statistically suitable. The factor model and standardized values are presented in Figure 1.

EL: Effective listening

ASQ: Asking strong questions

RA: Raising awareness

ECMCCV: Enabling Coachees to make choices according to their core values?

Chi-Square=2204.88, df=1115, P-value=0.00000, RMSEA=0.057
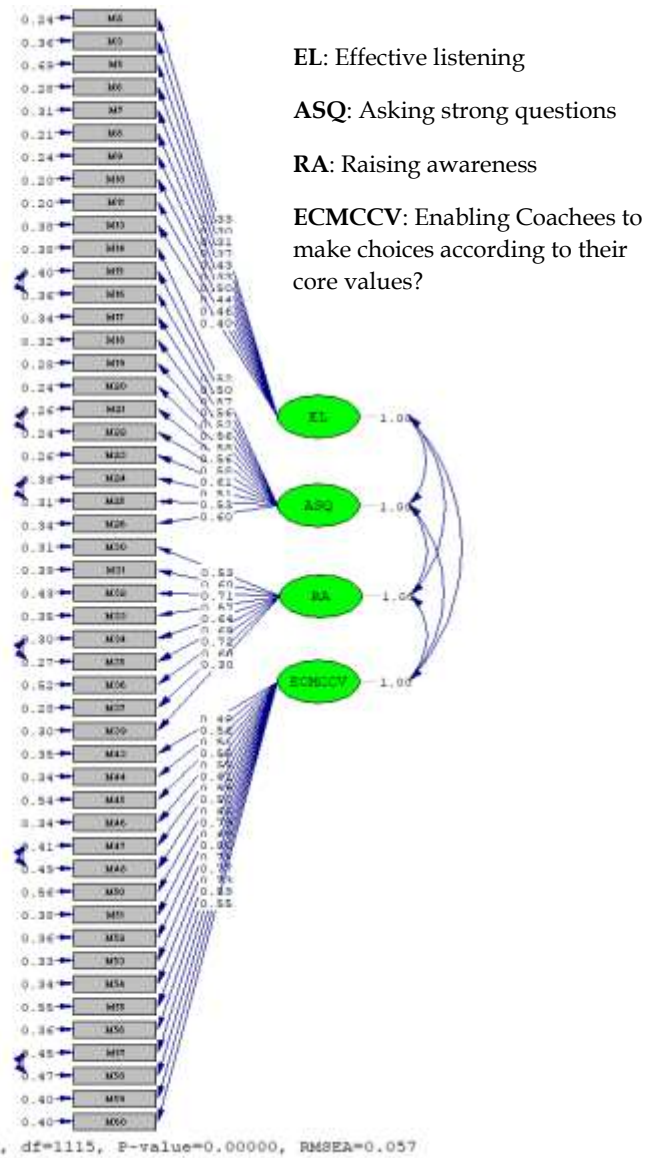
Figure 1. Confirmatory Factor Analysis Diagram for High School Teachers' Coaching Behaviors and Standardized Values

Analysis of the Data

The data analysis of the study was carried out through 412 surveys which were collected in the main implementation. In the data analysis, SPSS 24.0 (Statistical Package for Social Sciences) package program was used. In the data analysis, the

normality of the distribution was examined, and it was seen that there was a normal distribution in all dimensions concerning the scale. In the study, arithmetic mean, standard deviation, frequency, t-test, ANOVA and Kruskal Wallis Analysis were employed.

The significance level was regarded as .05. In the case that the difference was significant in ANOVA analyses, LSD test was used to determine which groups had significance. In the analysis, non-parametric Kruskal-Wallis Test was used for the variables when the group size was below 30. In the analysis of the education status variable, the category of associate degree was integrated with a bachelor degree, and the master's degree category was integrated with PhD category due to the fact that the group sizes were below 30. The age variable was considered as a constant variable in the scale, and in the analysis period, it was separated into five categories as below 25 years old, 26-34 years old, 35-44 years old, 45-54 years old, and 54 and above.

Results

The descriptive results of the opinions of teachers who worked in public high schools within the borders of Ankara province on the status of adoption and implementation of coaching behaviors concerning the four dimensions of coaching as a) Effective listening b) Asking strong questions c) Raising awareness d) Enabling Coachees to make choices according to their core values are presented in Table 4.

Table 4

The Descriptive Analyses Results on the Opinions of Teachers Who Work in Public High Schools within the Borders of Ankara Province on the Status of Adoption and Implementation of Coaching Behaviors

Adoption					Implementation	
$\bar{X}$	SD	Dimension			$\bar{X}$	SD
4.68	.570		1.	Talking to student clearly	4.43	.656
4.62	.712		2.	Listening students without interrupting	4.38	.659
4.67	.630		3.	Not doing something else while listening students	4.44	.773
4.59	.616		4.	Listening students by not being emotional	4.40	.672
4.70	.524	*Effective Listening*	5.	Listening students without judgement	4.37	.721
4.67	.605		6.	Making an eye contact while talking to students	4.50	.724
4.49	.702		7.	Enabling students to show their emotions while talking to them	4.34	.706
4.50	.645		8.	Valuing students' emotions	4.50	.645
4.63	.603		9.	Trying to understand students thoughts	4.43	.660
4.59	.636		10.	Asking students confirmation questions to be certain that they understand accurately	4.29	.784

Table 4 Continue

Adoption					Implementation	
$\bar{X}$	SD	Dimension			$\bar{X}$	SD
4.48	.718		11. Asking students questions that will make them realize their inner world		4.14	.823
4.47	.738		12. Asking open-ended questions to students to know them better		4.23	.805
4.47	.733		13. Asking open-ended questions to students to provide them insights		4.11	.855
4.62	.615		14. Asking open-ended questions to students to clarify the situation that students want to talk about		4.30	.733
4.52	.713		15. Asking open-ended questions to students that will put them in action		4.11	.882
4.62	.623		16. Enabling students to think more by asking them the question "How?"		4.41	.710
4.61	.628	*Asking Strong Questions*	17. Enabling students to think more about the sources of their opinions by asking the question "Why you think in this way?"		4,30	.732
4.55	.685		18. Enabling students to examine their opinions by asking the question "Why you think in this way?"		4.34	.742
4.58	.708		19. Enabling students to express the reasons of their opinions by asking the question "What makes you think in this way?"		4.32	.792
4.54	.731		20. Providing students a new perspective by asking the question "How else I could behave?"		4.28	.842
4.62	.642		21. Confirming the accuracy of students' information by asking them the question "How do you know this"		4.28	.774
4.58	.652		Enabling students to enhance their thoughts that support their opinions by asking the question "Can you explain this more?"		4.25	.816
4.49	.750		22. Encouraging students to ask more questions by asking the question "What kind of questions do you have in your mind?"		4.14	.881
4.59	.627		23. Realizing students' strengths		4.30	.788
4.65	.609	*Raising Awareness*	24. Enabling students to see their aspects that are not known by them yet by others		4.11	.818
4.51	.712		25. Directing students to various activities (sports/art etc.) to discover their strengths		4.02	.968
4.46	.721		26. Observing students carefully in the classroom to determine their strengths		4.18	.816
4.71	2.616		27. Talking to students about their strengths		4.17	.815
4.59	.676		28. Enabling students to realize their own strengths		4.16	.792

Table 4 Continue

Adoption				Implementation	
$\bar{X}$	SD	*Dimension*		$\bar{X}$	SD
4.41	.818		29. Sharing students' strengths with parents	3.77	.991
4.37	.772	*Raising Awareness*	30. Making suggestions for their aspects that can be improved	4.01	.807
4.48	.724		31. Appreciating positive behavior of students immediately	4.27	.762
4.41	.784		32. Allocating responsibilities to students in classroom that are appropriate for their strengths	4.10	.868
4.45	.728		33. Supporting students to legitimate their goals	4.03	.861
4.30	.850		34. Encouraging students to determine challenging goals	3.97	.872
4.44	.763	*Enabling Coaches to make choices according to their core values*	35. Supporting students to adopt the goals they determine	4.07	.835
4.41	.854		36. Asking students to imagine that they reach their goals	4.02	.883
4.30	.895		37. Asking students to think about tasks that they enjoy to accomplish	4.03	.869
4.35	.842		38. Reminding students their previous success	3.91	.928
4.27	.962		39. Enabling students to think about professions appropriate to their strengths	4.04	.952
4.09	.887		40. Asking students to make an explanation about their purpose of life	4,09	.887
3.80	1.013		41. Asking students to define their philosophy of life	3.80	1.013
3.97	.901		42. Asking students to think about things that they value the most	3.97	.901
3.57	1.121		43. Asking students to write the things they value at most according to order of importance	3.57	1.121
3.92	.941		44. Asking students to express what the concept of success means to them	3.92	.941
3.80	1.021		45. Asking students to express what the concept of happiness means to them	3.80	1.021
3.76	.967		46. Asking students to express themselves about the relationship between their goals and values	3.76	.967
3.81	.919		47. Enabling students to think about the reasons behind their behaviors	3.81	.919
4.14	.888		48. Enabling students to think about how their decisions will impact their lives.	4.14	.888

The findings showed that arithmetic mean of participants' opinions regarding the adoption level of the effective listening dimension varied between $\bar{X}$ =4.70 and $\bar{X}$ =4.49. In this dimension, the most adopted item was "Listening students without

judgement" ($\overline{X}$ =4.70), and the least adopted item was "Enabling students to show their emotions while talking to them" ($\overline{X}$ =4.49). The weighted average concerning this dimension was calculated as $\overline{X}$ =4.61. The arithmetic mean of students opinions regarding the implementation level of the effective listening dimension varied between $\overline{X}$ =4.50 and $\overline{X}$ =4.29. In this dimension, the most implemented item was "Listening students without judgement" ($\overline{X}$ =4.50), and the least implemented item was "Asking students confirmation questions to be certain that they understand accurately". The weighted average concerning this dimension was calculated as $\overline{X}$ =4.41.

The findings showed that arithmetic mean of participants' opinions regarding the adoption level of the asking strong questions dimension varied between $\overline{X}$ =4.62 and $\overline{X}$ =4.47. In this dimension, the most adopted item by the participants was "Asking open-ended questions to students to clarify the situation that students want to talk about" ($\overline{X}$ =4.62), and the least adopted item was "Asking open-ended questions to students to know them better ($\overline{X}$ =4.47). The weighted average concerning this dimension was calculated as $\overline{X}$ =4.55. The arithmetic mean of students opinions regarding the implementation level of the asking strong questions dimension varied between $\overline{X}$ =4.41 and $\overline{X}$ =4.11. In this dimension, the most implemented item was "Enabling students to think more by asking them the question "How?" ($\overline{X}$ =4.41), and the least implemented item was "Asking open-ended questions to students to provide them insights". The weighted average concerning this dimension was calculated as $\overline{X}$ =4.25.

The findings showed that arithmetic mean of participants' opinions regarding the adoption level of the raising awareness dimension varied between $\overline{X}$ =4.71 and $\overline{X}$ =4.37. In this dimension, the most adopted item by the participants was "Talking to students about their strengths" ($\overline{X}$ =4.71), and the least adopted item was"Making suggestions for their aspects that can be improved" ($\overline{X}$ =4.37). The weighted average concerning this dimension was calculated as $\overline{X}$ =4.53. The arithmetic mean of students' opinions regarding the implementation level of the raising awareness dimension varied between $\overline{X}$ =4.30 and $\overline{X}$ =3.77. In this dimension, the most implemented item was "Realizing students' strengths" ($\overline{X}$ =4.30), and the least implemented item was "Sharing students' strengths with parents" ($\overline{X}$ =3.77). The weighted average concerning this dimension was calculated as $\overline{X}$ =4.11.

The findings showed that arithmetic mean of participants' opinions regarding the adoption level of the enabling coachees to make choices according to their core values dimension varied between $\overline{X}$ =4.45 and $\overline{X}$ =3.57. In this dimension, the most adopted item by the participants was "Supporting students to legitimate their goals" ($\overline{X}$ =4.45), and the least adopted item was "Asking students to write the things they value at most according to order of importance" ($\overline{X}$ =3.57). The weighted average concerning this dimension was calculated as $\overline{X}$ =4.10. The arithmetic mean of students' opinions regarding the implementation level of the raising awareness dimension varied between $\overline{X}$ =4.14 and $\overline{X}$ =3.57. In this dimension, the most implemented item was "Enabling students to think about how their decisions will impact their lives." ($\overline{X}$ =4.14), and the least implemented item was "Asking students to write the things

they value at most according to order of importance" ($\overline{X}$ =3.57). The weighted average concerning this dimension was calculated as $\overline{X}$ =3.94.

The analysis results which revealed that teachers' opinions on the adoption and implementation status of coaching behaviors did not demonstrate a significant difference according to certain independent variables provided in Table 5.

Table 5

T-test Analysis Results on Teachers' Adoption Levels of Sub-Dimensions of Coaching Behaviors Scale According to the Variables of Gender, Status of Participating in Coaching Training, Conferences or Seminars, Status of Considering the Display of Coaching Behavior Significant and Education Level

Dimensions	Variables	Categories	n	$\overline{X}$	SD	df	t	P
Effective listening	Gender	Female	277	41.72	4.03	410	2.95	.003*
		Male	135	40.26	5.01			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	41.75	4.09	410	1.20	.228
		No	325	41.11	4.50			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	41.48	4.16	410	2.45	.018*
		No	43	39.20	5.91			
	Education Level	Associate Degree Bachelor Degree	339	41.25	4.24	410	.095	.925
		Post Graduate Degree	73	41.20	5.20			
Asking strong questions	Gender	Female	277	63.78	7.01	410	2.59	.01*
		Male	135	61.53	8.82			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	63.95	7.48	410	1.23	.218
		No	325	62.80	7.76			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	63.48	7,15	410	2.47	.017*
		No	43	59.27	10.86			
	Education Level	Associate Degree Bachelor Degree	339	63.17	7.32	410	.622	.536
		Post Graduate Degree	73	62.45	9.36			

Table 5 Continue

Dimensions	Variables	Categories	n	$\overline{\text{X}}$	SD	df	t	P
Raising awareness	Gender	Female	277	40.51	4.76	410	2.32	.025*
		Male	135	39.31	5.21			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	40.78	4.58	410	1.40	.162
		No	325	39.94	5.02			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	40.39	4.68	410	2.57	.013*
		No	43	37.81	6.36			
	Education Level	Associate Degree Bachelor Degree	339	40.31	4.79	410	1.70	.089
		Post Graduate Degree	73	39.23	5.52			
Enabling Coachees to make choices according to their core values	Gender	Female	277	73.67	11.34	410	2.06	.04*
		Male	135	71.15	12.20			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	74.08	10.83	410	1.11	.268
		No	325	72.51	11.89			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	73.42	11.11	410	2.32	.024*
		No	43	67.93	15.01			
	Education Level	Associate Degree Bachelor Degree	339	73.00	11.55	410	.583	.560
		Post Graduate Degree	73	72.12	12.32			

Table 5 showed that there was a significant difference between the participants' opinions on the adoption level of the effective listening dimension according to the variables of gender $[t_{(410)}= 2.95; p<.05]$ and status of considering the display of coaching behavior significant$[t_{(410)}= 2.45; p<.05]$. The female participants and the participants who considered the display of coaching behavior significant had more positive opinions on the adoption level of the effective listening dimension. There was no significant difference between participants' opinions on the adoption level of the effective listening dimension according to the variables of the status of participating

in coaching training, conferences or seminars *[t$_{(410)}$= 1.20; p>.05]* and education level*[t$_{(410)}$= .095; p>.05]*.

There was a significant difference between the participants' opinions on the adoption level of the asking strong questions dimension according to the variables of gender *[t$_{(410)}$= 2.59; p<.05]* and status of considering the display of coaching behavior significant*[t$_{(410)}$= 2.47; p<.05]*. The female participants and the participants who considered the display of coaching behavior significant had more positive opinions on the adoption level of the asking strong questions dimension. There was no significant difference between participants' opinions on the adoption level of the asking strong questions dimension according to the variables of the status of participating in coaching training, conferences or seminars*[t$_{(410)}$= 1.23; p>.05]* and education level *[t$_{(410)}$= .622; p>.05]*.

There was a significant difference between the participants' opinions on the adoption level of the raising awareness dimension according to the variables of gender *[t$_{(410)}$= 2.32; p<.05]* and status of considering the display of coaching behavior significant *[t$_{(410)}$= 2.57; p<.05]*. The female participants and the participants who considered the display of coaching behavior significant had more positive opinions on the adoption level of the raising awareness dimension. There was no significant difference between participants' opinions on the adoption level of the raising awareness dimension according to the variables of the status of participating in coaching training, conferences or seminars*[t$_{(410)}$= 1.40; p>.05]* and education level *[t$_{(410)}$= 1.70; p>.05]*.

There was a significant difference between the participants' opinions on the adoption level of the enabling coachees to make choices according to their core values dimension according to the variables of gender *[t$_{(410)}$= 2.06; p<.05]* and status of considering the display of coaching behavior significant*[t$_{(410)}$= 2.32; p<.05]*. The female participants and the participants who considered enact of coaching behavior significant had more positive opinions on the adoption level of the enabling coachees to make choices according to their core values dimension. There was no significant difference between participants' opinions on the adoption level of the enabling coachees to make choices according to their core values dimension according to the variables of the status of participating in coaching training, conferences or seminars*[t$_{(410)}$= 1.11; p>.05]* and education level*[t$_{(410)}$= .583; p>.05]*.

The variables of gender, status of considering the display of coaching behavior significant, and education level concerning the teachers' implementation level sub-dimensions of the coaching behavior scale were analyzed through t test. The analysis results are provided in Table 6.

Table 6

T-test Analysis Results on Teachers' Implementation Levels of Sub-Dimensions of Coaching Behaviors Scale According to the Variables of Gender, Status of Participating in Coaching Training, Conferences or Seminars, Status of Considering the Display of Coaching Behavior Significant and Education Level,

Dimensions	*Variables*	*Categories*	*n*	$\overline{\mathrm{X}}$	*SD*	*df*	*t*	*P*
Effective Listening	Gender	Female	277	39.7	4.37	410	2.05	.046*
		Male	135	38.7	4.99			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	40.1	4.69	410	1.75	.080
		No	325	39.2	4.56			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	39.5	4.46	410	1.13	.258
		No	43	38.6	5.71			
	Education Level	Associate Degree, Bachelor Degree	339	39.2	4.64	410	1.76	.079
		Post Graduate Degree	73	40.2	4.36			
Asking strong questions	Gender	Female	277	59.5	7.87	410	1.85	.064
		Male	135	57.9	9.10			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	60.5	8.51	410	1.87	.062
		No	325	58.6	8.23			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	59.3171	8.06157	410	1.68	.098
		No	43	56.6279	10.09013			
	Education Level	Associate Degree, Bachelor Degree	339	58.8	8.36	410	1.15	.250
		Post Graduate Degree	73	60.0	8.12			

Table 6 Continue

Dimensions	Variables	Categories	n	$\overline{\text{X}}$	SD	df	t	P
Raising awareness	Gender	Female	277	36.9	5.08	410	1.22	.222
		Male	135	36.2	5.44			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	38.2	4.96	410	3.10	.002*
		No	325	36.3	5.20			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	36.9	5.00	410	1.89	.065
		No	43	34.9	6.51			
	Education Level	Associate Degree, Bachelor Degree	339	36.6	5.26	410	.785	.433
		Post Graduate Degree	73	37.1	4.98			
Enabling Coaches to make choices according to their core values	Gender	Female	277	67.3	10.88	410	1.68	.093
		Male	135	65.3	11.62			
	Status of Participating in Coaching Training, Conferences or Seminars	Yes	87	69.4	10.35	410	2.58	.010*
		No	325	65.9	11.26			
	Status of Considering the Display of Coaching Behavior Significant	Yes	369	67.1	10.715	410	2.13	.038*
		No	43	62.5	13.83			
	Education Level	Associate Degree, Bachelor Degree	339	66.2	11.27	410	1.78	.075
		Post Graduate Degree	73	68.8	10.40			

Table 6 demonstrated that there was a significant difference between the participants' opinions on the implementation level of the effective listening dimension according to the variable of gender *[t$_{(410)}$= 2.05; p<.05]*. The female participants had more positive opinions on the implementation level of the effective listening dimension. There was no significant difference between participants' opinions on the implementation level of the effective listening dimension according to the variables of

the status of participating in coaching training, conferences or seminars*[t$_{(410)}$= 1.75; p>.05]*, status of considering the display of coaching behavior significant*[t$_{(410)}$= 1.13; p>.05]* and education level *[t$_{(410)}$= .1.76; p>.05]*.

There was no significant difference between participants' opinions on the implementation level of the asking strong questions dimension according to the variables of gender *[t$_{(410)}$= 1.85; p>.05]*, the status of participating in coaching training, conferences or seminars*[t$_{(410)}$= 1.87; p>.05]* , status of considering the display of coaching behavior significant*[t$_{(410)}$= 1.68; p>.05]* and education level *[t$_{(410)}$= 1.15; p>.05]*.

There was a significant difference between the participants' opinions on the implementation level of the raising awareness dimension according to the variable of the status of participating in coaching training, conferences or seminars *[t$_{(410)}$= 3.10; p<.05]*. The participants who participated in coaching training, conferences or seminars had more positive opinions on the adoption level of the raising awareness dimension. There was no significant difference between participants' opinions on the implementation level of the raising awareness dimension according to the variables of gender *[t$_{(410)}$= 1.22; p>.05]*, the status of considering the display of coaching behavior significant*[t$_{(410)}$= 1.89; p>.05]* and education level *[t$_{(410)}$= .785; p>.05]*.

There was a significant difference between the participants' opinions on the adoption level of the enabling coachees to make choices according to their core values dimension according to the variables the status of participating in coaching training, conferences and status *[t$_{(410)}$= 2.58; p<.05]* and considering the display of coaching behavior significant*[t$_{(410)}$= 2.13; p<.05]*. The participants who participated in coaching training, conferences or seminars had more positive opinions on the implementation level of the enabling coachees to make choices according to their core values dimension. There was no significant difference between participants' opinions on the implementation level of the enabling coachees to make choices according to their core values dimension according to the variables of gender *[t$_{(410)}$= 1.68; p<.05]* and education level *[t$_{(410)}$= 1.78; p>.05]*.

The ANOVA results on teachers' adoption level of the sub-dimensions of the Coaching Behaviors Scale according to the 'branch' variable are presented in Table 7.

Table 7

ANOVA Results on Teachers' Adoption Level of the Sub-Dimensions of the Coaching Behaviors Scale According to the 'Branch' Variable (N=400)

Dimensions	Variables	Categories	n	$\overline{X}$	SD	df	F	p	Sig.Dif (LSD)
Effective	Branch	1-SS	54	41.0	3.96	6;393	.721	.633	-
		2-Sc	52	41.2	4.72				
		3-TLL	62	41.03	5.20				
		4-Math	50	41.44	4.39				
		5-FL	73	41.1	4.11				
		6-VE	59	42.1	3.94				
		7-Others	50	40.52	4.59				
Asking strong questions	Branch	1. Social Sciences	54	62.0	8.12	6;393	.564	.759	-
		2.Science	52	63.5	8.36				
		3.Turkish Language - Literature	62	62.6	8.63				
		4.Mathematics	50	62.6	8.07				
		5.Foreign Language	73	62.8	6.95				
		6.Vocational Education	59	64.4	7.19				
		7.Others	50	62.6	7.12				
Raising awareness	Branch	1. Social Sciences	54	39.4	5.16	6;393	1.49	.178	-
		2.Science	52	39.0	5.35				
		3.Turkish Language - Literature	62	40.2	4.94				
		4.Mathematics	50	40.1	4.58				
		5.Foreign Language	73	40.0	4.8				
		6.Vocational Education	59	41.6	4.5				
		7.Others	50	40.2	4.9				

Table 7 Continue

Dimensions	Variables	Categories	n	$\overline{\mathrm{X}}$	SD	df	F	p	Sig.Dif (LSD)
Enabling Coachees to make choices according to their core values	Branch	1. Social Sciences	54	70.4	12.2				
		2.Science	52	69.8	12.9				
		3.Turkish Language - Literature	62	73.7	11.4				
		4.Mathematics	50	73.3	9.62	3;393	2.83	.010*	1-6 2-6 5-6
		5.Foreign Language	73	71.6	11.4				
		6.Vocational Education	59	77.5	10.1				
		7.Others	50	73.2	12.8				

Table 7 presented that there was a significant difference between the participants' opinions on the teachers' adoption level of the sub-dimensions of the Coaching Behaviors Scale according to the 'branch' variable according to the dimensions of effective listening [$F_{(3-393)}$= .721; p>.05], asking strong questions [$F_{(3-393)}$= .564; p>.05] and raising awareness [$F_{(3-393)}$= 1.49; p>.05]. In addition, there was a significant difference between participants' opinions concerning 'enabling coachees to make choices according to their core values' according to the 'branch' variable [$F_{(3-393)}$= 2.83; p<.05]. According to the LSD analysis results –which was among the Post Hoc tests conducted to determine which groups had differences between them- the opinions of the vocational education high school teachers ($\overline{\mathrm{X}}$=77.55) were more positive in comparison to the teachers of branches of social sciences ($\overline{\mathrm{X}}$=70.40), science ($\overline{\mathrm{X}}$=69.80) and foreign language ($\overline{\mathrm{X}}$=71.69) at the adoption level.

The ANOVA results on teachers' implementation level of the sub-dimensions of the Coaching Behaviors Scale according to the 'branch' variable are presented in Table 8.

Table 8

ANOVA Results on Teachers' Implementation Level of the Sub-Dimensions of the Coaching Behaviors Scale According to the 'Branch' Variable (N=400)

Dimensions	Variable	Categories	n	$\overline{X}$	SD	df	F	p	Sig.Dif (LSD)
Effective listening	Branch	1. Social Sciences	54	39.88	3.57	6;393	.590	.739	
		2.Science	52	39.71	4.90				
		3.Turkish Language - Literature	62	39.14	5.08				
		4.Mathematics	50	39.98	4.66				
		5.Foreign Language	73	38.75	4.92				
		6.Vocational Education	59	39.30	4.53				
		7.Others	50	39.04	4.36				
Asking strong questions	Branch	1. Social Sciences	54	58.68	7.88	6;393	.549	.771	
		2.Science	52	59.40	8.30				
		3.Turkish Language - Literature	62	58.33	8.79				
		4.Mathematics	50	59.52	8.47				
		5.Foreign Language	73	57.71	9.13				
		6.Vocational Education	59	58.79	7.98				
		7.Others	50	60.10	7.18				
Raising awareness	Branch	1. Social Sciences	54	36.25	5.14	6;393	1.46	.191	
		2.Science	52	35.17	5.14				
		3.Turkish Language - Literature	62	36.70	4.68				
		4.Mathematics	50	37.12	5.23				
		5.Foreign Language	73	36.42	5.13				
		6.Vocational Education	59	37.20	5.92				
		7.Others	50	37.92	4.88				
Enabling Coaches to make choices according to their core values	Branch	1. Social Sciences	54	64.25	11.89	3;393	3.23	.004*	1-6 1-7 2-3 2-4 2-6 2-7 5-6 5-7
		2.Science	52	62.71	10.69				
		3.Turkish Language - Literature	62	67.17	9.23				
		4.Mathematics	50	67.02	11.26				
		5.Foreign Language	73	65.23	10.73				
		6.Vocational Education	59	70.08	10.78				
		7.Others	50	69.30	12.24				

Table 8 revealed that there was no significant difference between the participants' opinions on the teachers' implementation level of the sub-dimensions of the Coaching Behaviors Scale according to the 'branch' variable according to the dimensions of effective listening [$F_{(3-393)}$= .590; p>.05], asking strong questions *[$F_{(3-393)}$= .549; p>.05]* and raising awareness [$F_{(3-393)}$= 1.46; p>.05]. In addition, there was a significant difference between participants' opinions concerning 'enabling coachees to make choices according to their core values' according to the 'branch' variable *[$F_{(3-393)}$= 3.23; p<.05].* According to the analysis results performed to determine which groups had significant difference between them; it was found out that the vocational education high school teachers' opinions ($\overline{X}$=70.08) were more positive in comparison to the teachers of social sciences($\overline{X}$=64.25), science($\overline{X}$=62.71) and foreign language ($\overline{X}$=65.23); and the opinions of teachers of other branches (art, music, physical education etc.) ($\overline{X}$=69.30) were more positive in comparison to the teachers of social sciences ($\overline{X}$=64.25), science ($\overline{X}$=62.71) and foreign language ($\overline{X}$=65.23) at the implementation level. Furthermore, the opinions of teachers of Turkish language –literature ($\overline{X}$=67.17) and mathematics ($\overline{X}$=67.02) were more positive in comparison to the teachers of science ($\overline{X}$=62.71). The Kruskal-Wallis test results on teachers' adoption level of the sub-dimensions of the Coaching Behaviors Scale according to the 'age' variable are provided in Table 9.

Table 9

Kruskal-Wallis Results on Teachers' Adoption Level of the Sub-Dimensions of the Coaching Behaviors Scale According to the 'Age' Variable (N=400)

Dimensions	Variable	Age Group	n	Mean Rank	df	χ^2	P	Sig.Dif (LSD)
Effective listening	Age	**1**-25 years old and below	22	207.70	4	7.68	.104	-
		2-26-34 years old	59	211.90				
		3-35-44 years old	133	221.02				
		4-45-54 years old	166	201.91				
		5-54 years old and above	32	159.19				
Asking strong questions	Age	**1**-25 years old and below	22	204.16	4	6.60	.158	-
		2-26-34 years old	59	196.66				
		3-35-44 years old	133	217.55				
		4-45-54 years old	166	210.23				
		5-54 years old and above	32	160.98				

Table 9 Continue

Dimensions	Variable	Age Group	n	Mean Rank	df	χ2	P	Sig.Dif (LSD)
Raising awareness	Age	**1**-25 years old and below	22	227.86	4	12.16	.016*	1-5 2-5 3-5 4-5
		2-26-34 years old	59	201.96				
		3-35-44 years old	133	216.98				
		4-45-54 years old	166	209.68				
		5-54 years old and above	32	140.13				
Enabling coachees to make choices according to their core values	Age	**1**-25 years old and below	22	221.75	4	3.54	.471	-
		2-26-34 years old	59	201.48				
		3-35-44 years old	133	208.18				
		4-45-54 years old	166	211.59				
		5-54 years old and above	32	171.88				

Table 9 showed that there was no significant difference between the participants' opinions on the teachers' adoption level of the sub-dimensions of the Coaching Behaviors Scale according to the 'age variable according to the dimensions of effective listening[$\chi^2_{(4)}$= 7.68; p>.05], asking strong questions [$\chi^2_{(4)}$= 6.60; p>.05], and enabling coachees to make choices according to their core values [$\chi^2_{(4)}$= 3.54; p>.05].

On the other hand, there was a significant difference in the teachers' adoption level of the sub-dimensions of the Coaching Behaviors Scale according to the 'age' variable in the 'raising awareness' dimension [$\chi^2_{(4)}$= 12.16; p>.05]. According to Mann Whitney U test results performed to determine the groups which had a difference between them, the participants who were 54 years old and above had more negative opinions in comparison to the participants who were 25 years old or below (U=201.500, p<0.05), 35-44 years old (U=1379.00, p<0.05), and 45-54 years old (U=1741.00, p<0.05).

The Kruskal-Wallis test results on teachers' implementation level of the sub-dimensions of the Coaching Behaviors Scale according to the 'age' variable are provided in Table 10.

Table 10

Kruskal-Wallis Results on Teachers' Implementation Level of the Sub-Dimensions of the Coaching Behaviors Scale According to the 'Age' Variable (N=400)

Dimension	Age Group	n	Mean Rank	df	χ^2	P	Sig.Dif (LSD)
Effective listening	Age						
	1.25 years old and below	22	185.36				
	2.26-34 years old	59	188.31				
	3.35-44 years old	133	218.83	4	4.68	3.22	-
	4.45-54 years old	166	209.95				
	5.54 years old and above	32	185.41				
Asking strong questions	Age						
	1.25 years old and below	22	198.84				
	2.26-34 years old	59	204.20				
	3.35-44 years old	133	205.50	4	.542	.969	-
	4.45-54 years old	166	210.90				
	5.54 years old and above	32	197.31				
Raising awareness	Age						
	1.25 years old and below	22	223.45				
	2.26-34 years old	59	210.40				
	3.35-44 years old	133	196.39	4	3.68	.451	-
	4.45-54 years old	166	215.51				
	5.54 years old and above	32	182.97				
Enabling coachees to make choices according to their core values	Age						
	1.25 years old and below	22	183.23				
	2.26-34 years old	59	205.09				
	3.35-44 years old	133	203.02	4	1.45	.835	-
	4.45-54 years old	166	212.98				
	5.54 years old and above	32	205.95				

Table 10 revealed that there was no significant difference between the participants' opinions on the teachers' implementation level of the sub-dimensions of the Coaching Behaviors Scale according to the 'age variable in terms of the dimensions of effective listening [$\chi^2_{(4)}$= 4.68; p>.05], asking strong questions [$\chi^2_{(4)}$= .542; p>.05], raising awareness [$\chi^2_{(4)}$= 3.68; p>.05], and enabling coachees to make choices according to their core values [$\chi^2_{(4)}$= 1.45; p>.05].

Discussion, Conclusion and Recommendations

The purpose of the study was to identify the opinions of teachers on the coaching behaviors according to the level of adoption and implementation levels of the four main sub-dimensions (effective listening, asking strong questions, raising awareness, enabling coachees to make choices according to their core values) of the developed scale.

The research findings revealed that the adoption levels of female participants and the participants who considered the display of coaching behavior meaningful showed a significant difference in all sub-dimensions of the teachers' coaching behaviors scale. Bozkurt (2004) obtained similar results in the study that examined primary school teachers' perceptions of communication skills. The findings of the aforementioned study showed that female teachers' communication skills had a higher average in "effectiveness" and "competence" sub-dimensions in comparison to male teachers. Furthermore, the finding of the proposed study which demonstrated that the participants who considered the display of coaching behavior significant adopted all of the sub-dimensions of coaching behaviors proved the importance of this behavior. In the present study, the opinions of female participants on the implementation level of the effective listening dimension were more positive. Similarly, the study conducted by Güven and Akyüz (2001) that addressed the opinions of prospective teachers' communication and problem-solving skills found out that female teachers had more positive opinions on transparency, equality, and competency sub-dimensions. Öztürk's (2007) study which compared the coaching skills of public and private school managers' coaching skills, Akçil's (2012) study which addressed the coaching roles of school managers, and Arslan's (2012) study which assessed teachers' coaching skills from a demographic perspective revealed that there was no significant difference between the two genders.

The proposed study showed that educators who participated in coaching education had certain coaching skills they implemented. In a similar way, the study carried out by Arthur-Kelly et al. (2017) examined the outcomes of the distance and direct coaching education provided for the Australian educators working on early childhood period with children who displayed coercive behavior and tended to display such kind of behavior. The research results showed that by means of the education provided for educators, children made a progress in terms of improving their skills, covering up their skill-related deficiencies, and reducing their behavioral problems. In the scope of the aforementioned study, the educators who received distance and direct coaching education made a progress in terms of understanding children's behavior better, supporting their positive behavior, assessing communication-related needs accurately, and developing open hypotheses about the motives of their behavior. In the study carried out by Gregory, Allen, Mikami, Hafen and Pianta (2017), the impact of vocational training on the behaviors of secondary and high school students was examined. The study included teachers from 87 different schools, and coaching and feedback education was provided with the aim of improving the interaction between teachers and students. According to the research findings, significant increases were determined in the positive behaviors of students.

Both in the proposed study and in the study conducted by Arslan (2012), teachers' education levels and acquisition of undergraduate degree, bachelor degree or postgraduate degree did not have any impact on their coaching skills. The study conducted by Öztürk (2007) revealed teachers with a master's degree thought that their administrators had higher levels of coaching skills in comparison to the teachers with a bachelor degree. The difference between these studies might have stemmed from the fact that Öztürk assessed the opinions of teachers on the coaching skills of administrators.

In the study conducted by Çetin (2015) which examined the coaching skill levels of senior high school-counsellors, similarly to the findings of the present study, when the findings were assessed according to the school type, it was determined that coaching levels of teachers in Anatolian Teacher High School in terms of communication sub-dimension were higher than coaching skills of teachers in other types of schools. Arslan (2012) shared in the findings of the study examining the coaching skills of teachers that there was no significant relationship between branches of teachers and coaching skills. The reason for the different findings obtained in the proposed study might result from conducting studies on different samples.

Conclusion

The proposed study and the similar studies revealed that female participants frequently show significant differences in terms of the levels of 'effective listening' and 'asking strong questions' which are among the communication skills included in coaching behaviors. The international experimental studies demonstrate that there are regular increases in students behavioral, cognitive and developmental levels by means of distance or direct coaching education provided to teachers. The participants who considered the display of coaching behavior significant showed a significant difference in terms of all dimensions of the scale concerning the adoption levels and in terms of raising awareness and enabling coachees to make choices according to their core values concerning the implementation level. The participants who participated in coaching training, conferences or seminars showed a significant difference in terms of raising awareness and enabling coachees to make choices according to their core values dimensions. A significant difference was found in the adoption and implementation levels of vocational education high school teachers in comparison to the teachers of other branches concerning the enabling coachees to make choices according to their core values dimension. In addition, a significant difference was found in the implementation levels of the art, music and physical education, and Turkish language –literature teachers concerning enabling coachees to make choices according to their core values dimension. Despite the fact that it was not included in the findings of the proposed study, various studies in the literature found out that there were significant differences regarding the display of coaching behavior by teachers with a master's degree.

Recommendations

Participants adopt the coaching behaviors yet fall short in implementing them. Therefore, coaching training should be included in in-service trainings provided by

the Ministry of National Education. The systematic and planned organization of coaching trainings is important as it is for all training programs. Trainings should be provided by leading experts in the form of activity-based trainings. Furthermore, school administrators should have the responsibility of supporting these trainings. According to the research findings, it was revealed that vocational education high school teachers and teachers of other branches (art, music, physical education etc.) enabled students to make choices according to their core values. For this reason, MoNE may ensure the implementation of activities within the curriculum that will enable to establish an effective communication to support teachers to have a one-to-one communication in schools besides vocational education high schools. Future studies may employ mixed research method in order to reveal the display of coaching behavior levels of general and vocational education high school teachers, and differences among them. Studies that examine the gender variable (for example by addressing only the male participants) can be designed considering the findings of the proposed study and of the similar studies in the literature, in terms of the positive significant differences of woman participants regarding the display of coaching behaviors. Furthermore, studies that compare the education levels and branches of teachers in the implementation of coaching behaviors can be conducted. The findings showed that teachers who participated in coaching training, conferences or seminars were more effective in terms of implementing coaching skills. In this direction, studies can be conducted by setting up two different teacher groups by providing a coaching training only to one group, and experimental studies that will examine the differences in terms of the fields of skills between the groups and the impact of this situation on students. According to the research results, it was found that teachers had inaccurate beliefs about the outcomes of coaching training. A large number of teachers thought that these behaviors should be implemented only by counsellors. In this sense, coaching trainings, seminars or conferences that will be planned by the MoNE can make a major contribution to an increase of teachers' awareness on this aspect. In-service trainings can be provided regarding the needs-analyses of teachers.

The implementation levels of the sub-dimensions of the scale demonstrated that female teachers' implementation levels concerning the effective listening dimension were higher. The participants who participated in coaching training, conferences or seminars showed a significant difference in terms of implementation level of the raising awareness dimension. The implementation level of the participants who participated in coaching training, conferences or seminars or considered the display of coaching behavior significant were higher in terms of enabling coachees to make choices according to their core values dimension. In addition, significant differences were found in the implementation levels of the vocational education high school teachers concerning enabling coachees to make choices according to their core values dimension in comparison to teachers of social sciences, science and foreign languages; the teachers of other branches (art, music, physical education etc.) in comparison to teachers of social sciences, science and foreign languages; and teachers of Turkish language-literature and mathematics in comparison to teachers of science. 97% of the teachers stated that they found teachers' display of coaching behavior significant.

Results of the Study and Suggestions: The research results revealed that teachers who participated in coaching training, conferences or seminars provided benefits to their students by implementing coaching skills. Therefore, it suggested to provide coaching training to teachers given by leading experts. It is possible to argue that preparation of new legislation by the Ministry of National Education that will include coaching trainings in the in-service trainings framework of teachers, and inclusion of coaching implementations in educational institutions can make positive contributions to the professional and personal development of teachers, and academic and personal development of students.

References

Barutcugil, İ. (2006). *Yöneticinin yönetimi* [*Manager's management*]. İstanbul: Kariyer Yayınları.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: GuilfordPress.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publishing, Inc.

Curtis, D., & Kelly, L. (2013). Effects of a quality of life coaching intervention on psychological courage and self-determination, *International Journal of Evidence Based Coaching and Mentoring, 11*(1), p. 20.

Cetin, E. (2015). *Öğrenci görüşlerine göre lise son sınıf rehber öğretmenlerinin koçluk beceri düzeyleri* [*Investigation of 12th grade students' class counselors' coaching talents by student perspective*]. Sakarya Üniversitesi, Sakarya.

Doran, G.T. (1981). There's a SMART way to write management's goals and objectives. *Management Review*, 70, 35-36.

Douglas, C. A., & McCauley, C. D. (1999). Formal developmental relationships: A Survey of organizational practices. *Human Resources Development Quarterly*,10(3), 203-220.

Druckman, D., & Bjork, R. A. (1991). *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press.

Engel, R. J. (2011). *An examination of wellness coaches and their impact on client behavioral outcomes* (Order No. 3479465). Available from ProQuest Dissertations & Theses Global. (904395557).

Frisch, M. (2013). Evidence-based well-being/positive psychology assessment and intervention with quality of life inventory. *Social Indicators Research*, doi: 10.1007/s11205-012-0140-7.

Gonzalez, D. (2008). *Executive coaching effectiveness: The coachee's experience* (Order No. 3339018). Available from ProQuest Dissertations & Theses Global. (304814863). Retrieved April, 12, 2018 from

http://search.proquest.com/docview/304814863?accountid=39364.

Gordon, T. (2008). *Etkili öğretmenlik eğitimi*. [*Effective teacher training*]. Çeviren: K. Sermin. İstanbul: Profil.

Grant, A. (2003). The impact of life coaching on goal attainment, metacognition and mental health. *Social Behavior and Personality*, *31*(3), 253-264.

Grant, A. (2006). A personal perspective on professional coaching and the development of coaching psychology. *International Coaching Psychology Review*, *1*(1), 12-22.

Grant, A. (2011). Developing an agenda for teaching psychology. *International Coaching Psychology Review, 6(2), 84.*

Grant, A. (2012). Making positive change: A randomized study comparing solution-focused vs. problem-focused coaching questions. *Journal of Systemic Therapies, 31*(2).

Grant, A.M., & Greene, J. (2004). *Coach yourself: Make real change in your life* (2nd ed.). London: Momentum Press.

Greene, J., & Grant, A.M. (2003). *Solution-focused coaching: Managing people in a complex world.* London: Momentum Press.

Gregory, A., Allen, P.J., Mikami, A.Y., Hafen, C.A., & Pianta, R.C.(2017). Effects of a professional development program on behavioral engagement of students in middle and high school. *HHS Public Access*. Psychol Sch.

Griffiths, K. (2005). Personal coaching: A model for effective learning. *Journal of Learning Design, 1*(2), 55.

Griffiths, K., & Campbell, M. (2009). Discovering, applying and integrating: The process of learning in coaching. *International Journal of Evidence Based Coaching and Mentoring, 7(*2), 16.

Gynnild, V. Vd. (2007). Teaching as coaching: a case study of awareness and learning in engineering education. *International Journal Of Science Education*, 29(1), 1-17.

ICF Türkiye. (2018). Why coaching is important?. Retrieved April, 4, 2018 from http://www.icfturkey.org/neden-kocluk/kocluk-nedir?

Jarrett, T. A. (2013). Warrior resilience and thriving (WRT): Rational emotive behavior therapy (REBT) as a resiliency and thriving foundation to prepare warriors and their families for combat deployment and posttraumatic growth in operation Iraqi freedom, 2005-2009. *Journal of Rational--Emotive & Cognitive--Behavior Therapy, 31*(2), 93-107. doi: 10.1007/s10942-013-0163-2.

Jones, H.B. (2017). *How former students attain goals through a process of self-coaching when training has ended*: A case study. Unpublished MA dissertation submitted. University of The Rockies.

Karasar, N. (2005). *Bilimsel araştırma yöntemi [Scientific research method]* (19.Baskı). Ankara: Nobel Yayın.

Knight, J. (2006). Instructional coaching. *The School Administrator, 63*(4), 36–40.

Landsberg, M. (1999). *Koçluğun taosu* [*Taos of coaching]* (Translate: H. B. Çelik). İstanbul: Sistem Yayıncılık.

Lawler, S. (2011). *Coaching for sustained change: How is it accomplished?* (Unpublished doctoral dissertation) Benedictine University, Lisle, Illinois.

Loucks-Horsley, S., Love, N., Stiles, K., Mundry, S., & Hewson, P. (2009). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks, CA: Corwin Press.

Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity. Promises and practicalities*. Washington, D.C. The Aspen Institute.

Newnham-Kanas, C., Irwin, J., & Morrow, D. (2008). Co-active life coaching as a treatment for adults with obesity. *International Journal of Evidence Based Coaching and Mentoring, 6(*2), p.1.

Nielsen, A., &Norreklit, H. A. (20009). Discourse analysis of the disdisciplinary power of management coaching. *Society and Business Reviews*, 2009: 202-14.

O'Connor, J., & Lages, A. (2004). *NLP ile koçluk* [*Coaching with NLP: How to be a Master Coach, 2004*] (Translate: Zaman, M.) İstanbul: Hayat Yayınları.

Poussard, M. J. (2003). *Yönetimde yeni bir stil coaching* [*A new management style coaching*]. Morpa Kültür Yayınları: İstanbul.

Spence, G.B., & Grant, A.M. (2007). Professional and peer life coaching and the enhancement of goal striving and well-being: An exploratory study. *Journal of Positive Psychology, 2*(3), 185–194.

Sue-Chan, C., & Latham, G. (2004). The relative effectiveness of external, peer, and self-coaches. *Applied Psychology: An International Review, 53*(2), 260-278. Retrived April, 12, 2018 from www.coachfederation.org,

Uckun, C.G., & Kilinc, Ġ.(2007). *Koçluk ve mentorluk tarihsel geliğimi, yöntemleri ve uygulamaları* [*Coaching and mentoring, historical development, methods and applications*]. Ürün yayınları: Ankara

Whitmore, J. (1992). *Coaching for performance.* London: Nicholas Brealey.

Wilkins, B. M. (2000). *A grounded theory study of personal coaching*. Missoula, MT: The University of Montana.

Wilson, C. (2008). Performance coaching and training in the workplace. Retrived April, 12, 2018 from http://www.theocm.co.uk/sites/default/files/documents/resources/Performance%20Coaching%20and%20Training%20in%20the%20Workplace.pdf.

Lise Öğretmenlerinin Koçluk Davranışlarına İlişkin Ölçek Geliştirme Çalışması

Özet

Problem Durumu: Öğrenciler kendi becerilerinin, güçlü ve geliştirmesi gereken yanlarının, duygusal ve zihinsel durumlarının farkında olduklarında eğitim-öğretim ortamlarına daha istemli bir şekilde dahil olurlar. Bu anlamda öğretmenlerin koçluk davranışlarına sahip olmasının ve uygulamasının, öğrencilerin farkındalığını geliştirmek, onların tutum ve davranışlarında olumlu değişiklik yaratmak adına oldukça uygun bir yöntem olduğu söylenebilir. Öğretmenler öğrencilerine güçlü sorular sorarak ve geri bildirimler vererek onların eleştirel düşünme ve problem çözme becerilerini geliştirirler. Böylece öğrenciler süreç boyunca yaşadıkları deneyimleri faydalı birer öğrenme aracına dönüştürebilirler. Kendilerinin farkında olan ve kendilerini tanıyan öğrenciler nasıl öğrendiklerini ve ne şekilde ders çalışmaları gerektiği konusunda bilgi sahibi olur. Bu durumun onların akademik performanslarını arttırmasına da katkı sağladığı söylenebilir. Öğretmenlerin eğitim-öğretim ortamlarında koçluk desteğinden yararlanmaları, öğrencilerin gerek bireysel ihtiyaçlarını tespit etmesine gerekse toplumsal ihtiyaçlara yönelik davranış ve beceri örüntüleri kazanmasına bununla birlikte, bilinç düzeyi yüksek bireyler olarak yaşantılarını sürdürmelerine yardımcı olur. Kendisini sorgulamayı öğrenen bireyler olarak gelişim göstermeleri, öğrencilerin özlerine ait değerler sitemini keşfetmelerine ve kendine uygun değerlere göre de seçimler yapabilmesine imkan sağlar. Koçluk sürecinin aynı zamanda öğretmenler için de kazanımları olduğunu söylemek mümkündür. Öğrenciyle sürekli etkileşimin olduğu, karşılıklı duyuşsal ve zihinsel süreçlerin devreye girdiği koçluk uygulamaları yoluyla öğretmenler, kendi farkındalıklarını daha da artırabilir ve iç görülerini geliştirebilirler. Aslında bu sürecin onlar içinde yeni bir içsel yolculuk, keşif ve deneyim süreci olduğu söylenebilir. Sınıf içindeki öğretim uygulamaları üzerine daha derinlemesine düşünmelerinin ise mesleki gelişimlerine katkı sağlayabileceği ifade edilebilir. Tüm bu değerlendirmeler ışığında koçluğun, eğitimin ve öğretimin niteliğinin artmasına, öğretmenlerin bireysel ve mesleki gelişimlerine, öğrencilerin akademik performanslarına ve bireysel gelişimlerine olumlu katkı sağlayan uygulamalardan biri olduğu düşünülmektedir.

Araştırmanın Amacı: Bu araştırmada, Ankara ili kamu liselerinde görev yapan öğretmenlerin koçluk davranışlarını benimseme ve uygulama durumlarına ilişkin görüşlerinin belirlenmesi amaçlanmıştır. Araştırmanın genel amacı doğrultusunda aşağıdaki sorulara yanıt aranmıştır: Ankara ili kamu liselerinde görev yapan öğretmenlerin koçluk davranışlarını benimseme ve uygulama durumlarına ilişkin görüşleri, koçluğun dört boyutu olan, a) Etkin dinleme, b) Güçlü sorular sorma, c)

Farkındalık yaratma, d) Öz değerlerine uygun seçimler yapmalarını sağlama boyutlarında nasıldır?

Ankara ili kamu liselerinde görev yapan öğretmenlerin koçluk davranışlarını benimseme ve uygulama durumlarına ilişkin görüşleri, cinsiyet, yaş, eğitim seviyesi, branş, daha önce koçluk eğitim/konferansı ya da seminerine katılma durumu ve koçluk davranışını önemli görme durumu değişkenlerine göre anlamlı bir farklılık göstermekte midir?

Araştırma genel tarama modelinde olup, nicel araştırma yöntemi kullanılmıştır. Örneklem seçiminde tabakalı örnekleme tekniğinden yararlanılmıştır. Bu araştırmanın örneklemi Ankara Büyükşehir Belediye Sınırları içerisinde yer alan dokuz ilçeden toplam 362 öğretmenden oluşmaktadır.

Araştırmanın Yöntemi: Araştırmanın nicel verileri, araştırmacı tarafından geliştirilen ''Lise Öğretmenlerinin Koçluk Davranışlarına İlişkin Görüşleri'' ölçeği ile toplanmıştır. Araştırmada nicel verilerin analizinde SPSS Paket Programı kullanılmıştır. Araştırmanın birinci amacına ilişkin analizler aritmetik ortalama, standart sapma ve frekans hesaplamalarıyla yapılmıştır. Araştırmanın ikinci alt amacına ilişkin olarak katılımcıların görüşlerinin cinsiyet, koçluk eğitimi, konferansı ya da seminerine katılma, koçluk davranışı göstermenin önemli görülme durumu ve eğitim durumu değişkenlerine göre farklılaşıp farklılaşmadığı t-testi ile analiz edilmiştir. Katılımcıların görüşlerinin yaş değişkenine göre farklılaşıp farklılaşmadığı Kruskal-Wallis analizi, branş değişkenine göre farklılaşıp farklılaşmadığı ANOVA ile analiz edilmiştir. Manidarlık düzeyi .05 alınmıştır. Araştırmanın üçüncü alt amacına ilişkin olarak katılımcıların görüşleri için frekans ve yüzde hesaplamaları yapılmıştır.

Araştırmanın Bulguları: Araştırmadan elde edilen sonuçlara göre, ölçeğin alt boyutlarını benimseme düzeylerine bakıldığında, kadın katılımcıların ve koçluk davranışı göstermenin önemli olduğunu düşünenlerin ölçeğin tüm alt boyutlarını benimseme düzeyleri daha yüksektir. 54 yaş üstü katılımcıların ise diğer yaş aralığındakilere göre farkındalık yaratma boyutunu benimseme düzeyi daha düşüktür. Bununla birlikte, meslek lisesi öğretmenleri sosyal bilimler, fen bilimleri ve yabancı dil branşındaki öğretmenlere göre öz değerlerine göre seçimler yapmalarını sağlama boyutunu benimseme düzeyinde daha olumlu düşüncelere sahiptirler.

Ölçeğin alt boyutlarını uygulama düzeylerine bakıldığında, kadın öğretmenlerin etkin dinleme boyutunu uygulama düzeyi daha fazladır. Koçluk eğitimi, konferansı ya da seminerine katılan katılımcılar farkındalık yaratma boyutunu uygulama düzeyinde anlamlı bir farklılık göstermektedir. Koçluk eğitimi, konferansı ya da seminerine katılanların ve koçluk davranışı göstermenin önemli olduğunu düşünenlerin öz değerlerine uygun seçimler yapmalarını sağlama boyutunu uygulama düzeyi daha yüksektir. Bununla birlikte meslek lisesi öğretmenlerinin sosyal bilimler, fen bilimleri ve yabancı dil branşındakilere göre, diğer branştaki öğretmenlerin (resim, müzik, beden eğt.vb.), sosyal bilimler, fen bilimleri ve yabancı dil branşındakilere göre, Türkçe-edebiyat ve matematik branşındaki öğretmenlerin ise fen bilimlerindeki öğretmenlere göre öz değerlerine uygun seçimler yapmalarını sağlama boyutunu

uygulama düzeyinde anlamlı şekilde farklılık görülmüştür. Katılımcıların %97'si öğretmenlerin koçluk davranışı göstermelerinin önemli olduğunu belirtmişlerdir.

Araştırmanın Sonuçları ve Öneriler: Araştırma bulguları gösteriyor ki koçluk eğitimi, konferansı ya da seminerine katılmış olan öğretmenler koçluk becerilerini uygulayarak öğrencilerine fayda sağlamaktadırlar. Bu nedenle öğretmenlerin, alanında uzman kişilerce koçluk eğitimi almaları önerilmektedir. Milli Eğitim Bakanlığı'nın öğretmenlerin hizmet içi eğitimlerine koçluk eğitimlerini ekleyecek şekilde yeni yasal düzenlemeler yapmasının ve eğitim kurumlarında koçluk uygulamalarına yer verilmesinin, öğretmenlerin bireysel ve mesleki, öğrencilerin de akademik ve bireysel gelişimlerine olumlu katkı sağlayacağını söylemek mümkündür.

Anahtar Kelimeler: Koçluk, koçluk eğitimi, eğitim koçluğu, öğretmen koçluğu, öğretmenlerin koçluk davranışları.

School Happiness: A Scale Development and Implementation Study*

Senol SEZER[1], Ertug CAN[2]

ARTICLE INFO	ABSTRACT
	Purpose: This research aimed to determine school happiness level of the teachers based on their opinions. For this purpose, 'School Happiness Scale' was developed by the researchers. The validity and reliability of the School Happiness Scale were tested. And then, the scale was applied to 484 teachers to determine their school happiness level. **Research Method:** In the study, a quantitative research design was employed during the data collection and the analysis phases. **Results:** The School Happiness Scale consists of 26 items and five sub-factors. The item-total correlations range from .54 to .86. The five factors explain 65.09% of the total variance. The results showed that teachers

perceived moderate level of school happiness in some sub-dimensions such as 'physical equipment', and 'activities'. They also perceived high level of school happiness in 'collaboration', and 'school management' sub-themes. The scores varied according to the school type and teaching field variables.

 Implications for Research and Practice: The results revealed that the school principals should encourage teachers, and display democratic and fair school management attitudes to increase school happiness. The top managers can elaborate to enhance the physical conditions of the schools. In order to increase school happiness of teachers, the collaboration can be improved among the members of school community. School Happiness Scale can be applied to different participants, and so the school happiness levels of the teachers as well as the validity and reliability scores of scale can be measured.

Introduction

Throughout the 20th Century, education discourse commonly had a tendency towards preparing learners as productive individuals for the world of work, especially in the context of growing economic development and competition. In the 21st Century, it is expected from education systems to educate students as they can cope with the rapid change along with increasing mobility, life stress, inequalities, global warming and environmental concerns. As the symptoms of 'unhappiness' begin to emerge, these expectations reflect the need to reposition the school more than to serve as an educational institution, but rather an environment which allows the social and emotional development of students. It can be said that happiness is not just a goal of education, but also a factor in school effectiveness.

There are essentially two ways to approach the study of happiness one of which is to think about what is meant by the term, and the other is to ask people what makes them happy (Thoilliez, 2011). Happiness is defined as the satisfaction level of an individual as a whole (Selim, 2008). Happiness can also be defined as the feelings of an individual such as joy, gladness, hope and physical and spiritual well-being (Koknel, 1992). Actually, happiness was defined by the researchers in different ways. Huebner (1991) defined happiness as life satisfaction; according to Seligman, Parks and Steen (2004), happiness is the meaning attached to life; Lyubomirsky, Sheldon and Schkade (2005) defined happiness as the positive feelings about life. Veenhoven (2008), defined happiness as 'the evaluation of life as a whole'. Happiness is described by Diener (1984), as more positive emotions than the negative ones and generally as satisfaction taken from life. Similarly, happiness is identified by Seligman (2011) as a multidimensional structure that includes meaning of life, positive feelings, responsibility, positive relationships, and success. Results from research reveal that happiness of individuals is highly functional for the success, because happy people perceive the world as a safer place and feel more confident (Boehm & Lyubomirsky, 2008; Fredrickson, 2013). They also make decisions easily, cooperate more easily, and are more tolerant (Lyubomirsky & King, 2005; Pan & Zhou, 2013; Schnittker, 2008). While the happiness of adults depends on many variables such as health, family life, social relations, security, freedom, moral values, income level, and working conditions; children's happiness can be related to the variables such as meeting level of their basic needs, love, trust, communication, health, and play (Ahn, Garcia, & Jimeno, 2004; Clair, 2012; Thoilliez, 2011). Growing efforts to measure happiness have also coincided with increased efforts to measure the quality of education, for instance through global indices and international student assessments.

School happiness is expressed as the emotional well-being, which is the result of harmony between school's expectation and personal needs of students, teachers, school managers and other employees depending on certain environmental factors (Engels, Aelterman, Petegem, & Schepens, 2004). The school happiness is associated with physical factors, individual factors, social-emotional factors, and instructional factors (Talebzadeh & Samkan, 2011). Yildirim (2014) asserts that cooperation among staff, fair and helpful assessment and feedback, positive school climate, student-oriented teaching practices, classroom climate and personal development increase

teachers' well-being in school. Aelterman, Engels, Van Petegem and Verhaeghe (2007) argue that well-being of teachers is associated with teacher-parent relations, support from colleagues, self-efficacy, workload, positive attitudes towards innovations, and support from school principal. Positioning teachers' well-being within wider social and professional contexts, which teachers function in, is necessary to gain an understanding of the complex mutual interaction between individual, relational and external factors that affect, constrain and mediate the happiness of the teachers.

It is clear that a happy school environment is very important in terms of effective learning and bringing out students' talents (Boehm & Lyubomirsky, 2008). Similarly, Bird and Markle (2012) indicate that happy school environment not only contributes to student's academic success but also improves other life skills, such as healthy communication, lifelong success, and self-fulfillment. Moreover, decrease in school happiness can lead to lower school success, loneliness, stress, depression, and drug addiction by causing weak relationships for students (Yucel & Vogt-Yuan, 2016). In other words, students need to have a strong sense of happiness, especially during school life, in terms of having positive beliefs about life and being hopeful for the future. Aydin (2016) specifies that children should be educated with great care, by teaching them to love nature and all living things. To promote learner happiness and well-being in schools do no imply that learning be made easier or require less effort, but rather that such approaches could help to contribute a distinctive love of learning. Consequently, teachers play a crucial role in the learning process of students. Happy teachers provide a non-threatening environment for students in learning process as a facilitator, planner, instructor, mediator and explainer.

In recent years, there has been an increasing number of studies related to the student learning in a happy school environment, in Turkey. These studies generally focus on subjective well-being in school (e.g. Asici & Ikiz, 2018; Certel, Bahadir, Saracaloglu, & Varol, 2015; Gundogdu & Yavuzer, 2012; Ozturk & Cetinkaya, 2015; Turkdogan & Duru, 2012; Turkmen, 2012; Ucan & Kiran-Esen, 2015; Yaliz-Solmaz, 2014). In some studies, school happiness is associated with different variables (e.g. Buyuksahin-Cevik & Yildiz, 2016; Demir-Celebi & Sezgin, 2015; Dogan, Sapmaz, & Akinci-Cotok, 2013; Ozdemir & Koruklu, 2011; Ozturk, Meral, & Yilmaz, 2017; Terzi, 2017; Sarıcam, 2014). It is seen that limited studies has been conducted on school happiness (e.g. Demiriz & Ulutas, 2016; Telef, 2014; Unuvar, Calısandemir, Tagay, & Amini, 2015). International studies conducted by Chaplin (2009), Holder and Klassen (2010), López-Pérez, Sánchez and Gummerum (2015), McCabe, Bray, Kehle, Theodore and Gelbar, (2011), Mahon and Yarcheski (2002), Park and Peterson (2006), Schnittker (2008), Talebzadeh and Samkan (2011), Uusitalo-Malmivaara (2012), Uusitalo-Malmivaara and Lehto (2013), Van Hal, et al. (2014), Weaver and Habibov (2010) focused on the relationships among subjective well-being, happiness and success in a similar way. However, number of the studies on teachers' school happiness (e.g. Acton & Glasgow, 2015; Aelterman, Engels, Van Petegem, & Verhaeghe, 2007; Yildirim, 2014) is limited. When the related literature is reviewed, it is seen that a comprehensive measurement tool which aims to determine teachers' happiness in school has not been developed yet. On the other hand, it is mentioned that the number of studies on

teachers' school happiness is limited. The purpose of this study was to develop a comprehensive measurement tool to determine the teachers' school happiness. In addition, it was also aimed to determine the school happiness level of teachers.

Method

Development of the School Happiness Scale

1. Creation of item pool: At this stage, firstly views of 20 teachers about the standards for school happiness were asked. Then, the data were analyzed by using content analysis technique, and the basic standards for school happiness were determined. Consequently, 34 most prominent views were taken into consideration, the items pool was created, and the items of the school happiness scale emerged.

2. Application of the expert opinion: In the second stage, two field experts from the Department of Educational Sciences were consulted to determine the suitability level of each item to measure teachers' school happiness.

3. Rewriting of scale items: In the third stage, the scale items were rewritten in line with the changes foreseen by the field experts.

4. Item analysis, explanatory factor analysis: In the fourth stage, the raw state of the scale was applied to 430 teachers working in different schools. Consequently, item analysis and explanatory factor analysis were carried out in the context of validity and reliability studies based on the data.

Item analysis is used to determine the points given by participants to each item and it is actually the determination process of what level the items are sufficient to measure the participants' attitudes (Everitt, 2006). At this stage, the psychometric properties of the scale were examined. Firstly, a Z-test analysis was performed on the total scores given to each item of School Happiness Scale. Z-test results showed that all data changed between -3 and +3. Consequently, Cronbach's analysis was performed for all items, and the results were evaluated (Cokluk, Sekercioglu, & Buyukozturk, 2012).

Explanatory Factor Analysis (EFA)

Kaiser-Meyer-Olkin (KMO) and Barlett Test were applied to determine whether the scales were suitable for factor analysis (Field, 2005). KMO value was .93, Barlett's test results ($\chi^2_{(430)}$ = 6978.09, p< .001) were significant. The results showed that the number of participants were sufficient, and that the data were appropriate for factor analysis (Leech, Barrett, & Morgan, 2005). A factor load of .50 was considered as a measure of substance permanence during EFA (Tabachnick & Fidell, 2013). Based on this criterion, 8 items were removed from the scale due to insufficient correlation. After the elimination process a five-factor 'School Happiness Scale' emerged that explained 65.09% of the total variance. The scale consisted of 26 items and the item-total correlations ranged from .54 to .86. The factor loads related to sub-factors were given in Table 1.

Table 1

Factor Loads Related to the Sub-factors (η=430)

Item	F1	F2	F3	F4	F5
SH01. The school's physical environment is healthy	.696				
SH02. School is sufficient for healthy eating opportunities	.633				
SH03. School's physical equipment is sufficient for social activities	.739				
SH04. Students have enough play areas in the school	.662				
SH06. Students love school		.706			
SH07. Students feel safe at school		.665			
SH08. Rewarding is fair in school		.602			
SH09. Students feel themselves valued at school		.733			
SH10. The responsibilities are shared in school		.621			
SH23. Students do the activities in love		.690			
SH25. Learning activities support students' development		.562			
SH11. A cooperation and solidarity culture is prevalent in school			.537		
SH12. Teachers love their profession			568		
SH14. Teachers motivate students to succeed			.629		
SH15. Teachers are sufficient in their profession			.680		
SH17. Open communication is prevalent in school			.652		
SH18. Mutual respect is prevalent in school			.767		
SH19. Mutual sensibility and tolerance are prevalent in school			.752		
SH26. Learning environment is more cooperative than competition			.566		
SH20. Guidance activities are sufficient				.674	
SH29. Social activities are sufficient				.642	
SH30. Special attention is paid to sports activities in school				.782	
SH31. School principal exhibits educational leadership					.786
SH32. School principal appreciates employees					.829
SH33. School principal exhibits fair attitude					.855
SH34. School principal exhibits a democratic attitude					.865
Eigenvalue	2.46	4.26	4.35	2.11	3.74
Variance (%)	%9.45	%16.40	%16.73	%8.13	%14.38
Total: 65.09					

Based on the content of factors, the first sub-factor was designated as the 'Physical Equipment' and consisted of four items. The factor loadings of items ranged from .633 to .739. The eigenvalue of the factor was 2.46 which corresponded to 9.45% of the total variance. The second sub-factor was designated as the 'Learning Environment' and

consisted of seven items. The factor loadings ranged from .562 to .733. The eigenvalue of the factor was 4.26 which corresponded to 16.40% of total variance. The third sub-factor was designated as the 'Collaboration' and consisted of eight items. The factor loadings ranged from .537 to .767. The eigenvalue of the factor was 4.35 which corresponded to 16.73% of total variance. The fourth sub-factor was designated as the 'Activities' and consisted of three items. The factor loadings ranged from .642 to .782. The eigenvalue of the factor was 2.11 which corresponded to 8.13% of total variance. The fifth sub-factor was designated as the 'School Management' and consisted of four items. The factor loadings ranged from .786 to .865. The eigenvalue of the factor was 3.74 which corresponded to 14.38% of total variance. The five factors explained 65.09% of total variance.

Confirmatory Factor Analysis

In Confirmatory Factor Analysis process, the sample group was determined as 449 teachers working in different schools. The factor structure of five-factor model was tested by Confirmatory Factor Analysis (CFA) using LISREL 8.51. In general, the values agreed on are values of $\chi^2/df = 2$ or less. In this model, χ^2/df was calculated as 3.77. In large samples, the ratio of χ^2/df below 3 shows that the fit is excellent, and below 5 shows a moderate compliance (Kline, 2005; Sumer, 2000). Items with .05 or lower RMSEA and SRMR values, show excellent fit (Cokluk, Sekercioglu, & Buyukozturk, 2012; Kline, 2005). In this model, SRMR value was calculated as .057 and RMSEA as .079. Byrne (2011), indicated that the items between .05 and .08 RMSEA and SRMR values show acceptable fit. Modifications have been applied between items 3rd and 4th, 15th and 16th. These modifications were made to improve the model. In addition, item modifications were practiced in the same sub-factors that they were correlated. Schreiber, Nora, Stage, Barlow and King (2006) assert that the modifications are made to improve the model. In addition, the item modifications give more practical results if the items are in same sub-factors because they are correlated. The results showed that the model was suitable for first-level Confirmatory Factor Analysis. Based on the Confirmatory Factor Analysis the fit index values were given in Table 2.

Table 2

School Happiness Scale Fit Index Values (N=449)

Model	χ^2	χ^2/df	NFI	NNFI	CFI	GFI	SRMR	RMSEA
Five-Factor Structure	1082.69	3.77	.86	.88	.89	.84	.057	.079

In Table 2, it was seen that $\chi^2_{(287)} = 1082.39$, $p < 0.001$, $\chi^2/df = 3.77$, NFI= .86, NNFI = .88, CFI = .89, GFI =.84, SRMR = .057, RMSEA = .079.

Cronbach Alpha values related to the sub-factors were given in Table 3.

Table 3

Cronbach Alpha Values Related to the Sub-factors (N=449)

Sub-factor	Cronbach Alpha (a)
1. Physical Equipment	.735
2. Learning Environment	.872
3. Collaboration	.828
4. Activities	.704
5. School Management	.940
6. Total (School Happiness)	.932

In Table 3, it was seen that Cronbach Alpha (α) value for Physical Equipment sub-factor was α= .735, Learning Environment sub-factor was α= .872, Collaboration sub-factor was α= .828, Activities sub-factor was α= .704, and school management sub-factor was α= .932. In addition, Cronbach Alpha (α) value for School Happiness Scale was α= .932.

In Table 4, the standardized solution values, T-values and R^2 values were given.

Table 4

Standardized Solution, T-values, and r^2 Values of School Happiness Model (N=449)

Item	Standardized Solution	T-Values	R^2
SH1	0.73	15.23	0.53*
SH2	0.61	12.40	0.37*
SH3	0.63	12.79	0.39*
SH4	0.50	9.60	0.25*
SH5	0.66	15.13	0.43*
SH6	0.70	16.31	0.48*
SH7	0.64	14.61	0.41*
SH8	0.72	17.21	0.52*
SH9	0.79	19.37	0.62*
SH10	0.76	18.37	0.58*
SH11	0.69	16.16	0.48*
SH12	0.61	13.59	0.37*
SH13	0.56	12.29	0.31*
SH14	0.74	17.42	0.54*
SH15	0.74	17.60	0.55*
SH16	0.74	17.36	0.54*
SH17	0.32	6.59	0.10*
SH18	0.63	14.19	0.40*
SH19	0.66	15.07	0.44*
SH20	0.75	16.70	0.57*
SH21	0.65	14.12	0.43*
SH22	0.52	10.72	0.27*
SH23	0.85	22.43	0.73*
SH24	0.86	22.86	0.75*
SH25	0.95	27.03	0.91*
SH26	0.95	26.67	0.89*

*p< .01

In Table 4, it was seen that standardized solution values were between 0.32 and 0.95. T-values were between 6.59 and 27.03. In addition, item total correlation values were between 0.10 and 0.91.

In this section, the application process of the School Happiness Scale and the results were given.

Research Design

This study, aiming to determine the school happiness level of teachers, was designed in a survey model. Survey model is used to determine the certain characteristics of a group (Buyukozturk, Kılıç-Cakmak, Akgun, Karadeniz, & Demirel, 2016). In a survey model, it is intended to depict a situation that existed in the past or is still continuing its existence. In this process, the researchers aim to determine the participants' opinions on a topic or case as well as their interests, skills, abilities, and attitudes, etc. (Fraenkel & Wallen, 2012; Karasar, 2017). Since it was aimed to develop a School Happiness Scale and implement, it was envisaged that the survey model was more suitable for this study.

Research Sample

The population was 3.418 teachers in Kırklareli province. The participants were determined by using cluster sampling method. In cluster sampling method, the universe is divided into groups called clusters, and each cluster is defined as a sampling unit. The randomly selected clusters are brought together and the sample is formed (Comlekci, 2001). Consequently, the teachers were divided in four groups considering the school type, and 121 teachers were determined for each clusters. The participants were 484 teachers working in different schools. The demographic characteristics of the participants were given in Table 5.

Table 5

The Demographic Characteristics of Teachers (N=484)

Variable		f	%
Gender	Female	309	63.8
	Male	175	36.2
Professional Seniority	1-6 Years	107	22.1
	7-12 Years	82	16.9
	13-18 Years	88	18.2
	19-24 Years	96	19.8
	Over 25 Years	111	22.9
School Type	Primary School	121	25.0
	Secondary School	121	25.0
	Vocational High School	121	25.0
	Academic High School	121	25.0
Teaching Field	Classroom Teacher	116	24.0
	Science Teacher	92	19.0
	Social Sciences Teacher	153	31.6
	Arts, and Sports Teacher	63	13.0
	Vocational Education Teacher	60	12.4

Data Collection Tool

Data were collected through the School Happiness Scale. The School Happiness Scale (SHS) was developed by the researchers.

Data Analysis

The data were analyzed by using the IBM SPSS 22 program. To determine the school happiness, descriptive analysis was performed and the average scores and standard deviations were calculated. Independent t-test was used to determine whether the mean scores differed in terms of gender variable. In addition, ANOVA was used to determine whether the mean scores differed in terms of professional seniority, school type and teaching field variables. Moreover, the origin of differences was determined by using the Scheffe test which was used as a post hoc test.

Results

The Results Related to the Teachers' School Happiness Level

The mean and standard deviation scores related to the school happiness levels were given in Table 6.

Table 6

The Results Related to the School Happiness (N=484)

Sub-dimensions	$\bar{x}$	SD
Physical equipment	3.57	.75
Learning environment	3.95	.58
Collaboration	4.09	.60
Activities	3.66	.81
School management	4.05	.94
School happiness (Total)	3.64	.50

In Table 6, it was seen that the highest scores related to the teachers' school happiness were in Collaboration sub-dimension ($\bar{x}$ =4.09, S= .60). In School Management sub-dimension, the school happiness scores were ($\bar{x}$ =4.05, S= .94), in Learning Environment sub-dimension were ($\bar{x}$ =3.95, S= .58), in Activities sub-dimension were ($\bar{x}$ =3.66, S= .81) and the Physical Equipment sub-dimension were ($\bar{x}$ =3.57, S= .75). The school happiness scores in general were ($\bar{x}$ =3.64, S= .50).

The frequencies of teacher opinions related to the sub-themes were given in histograms.
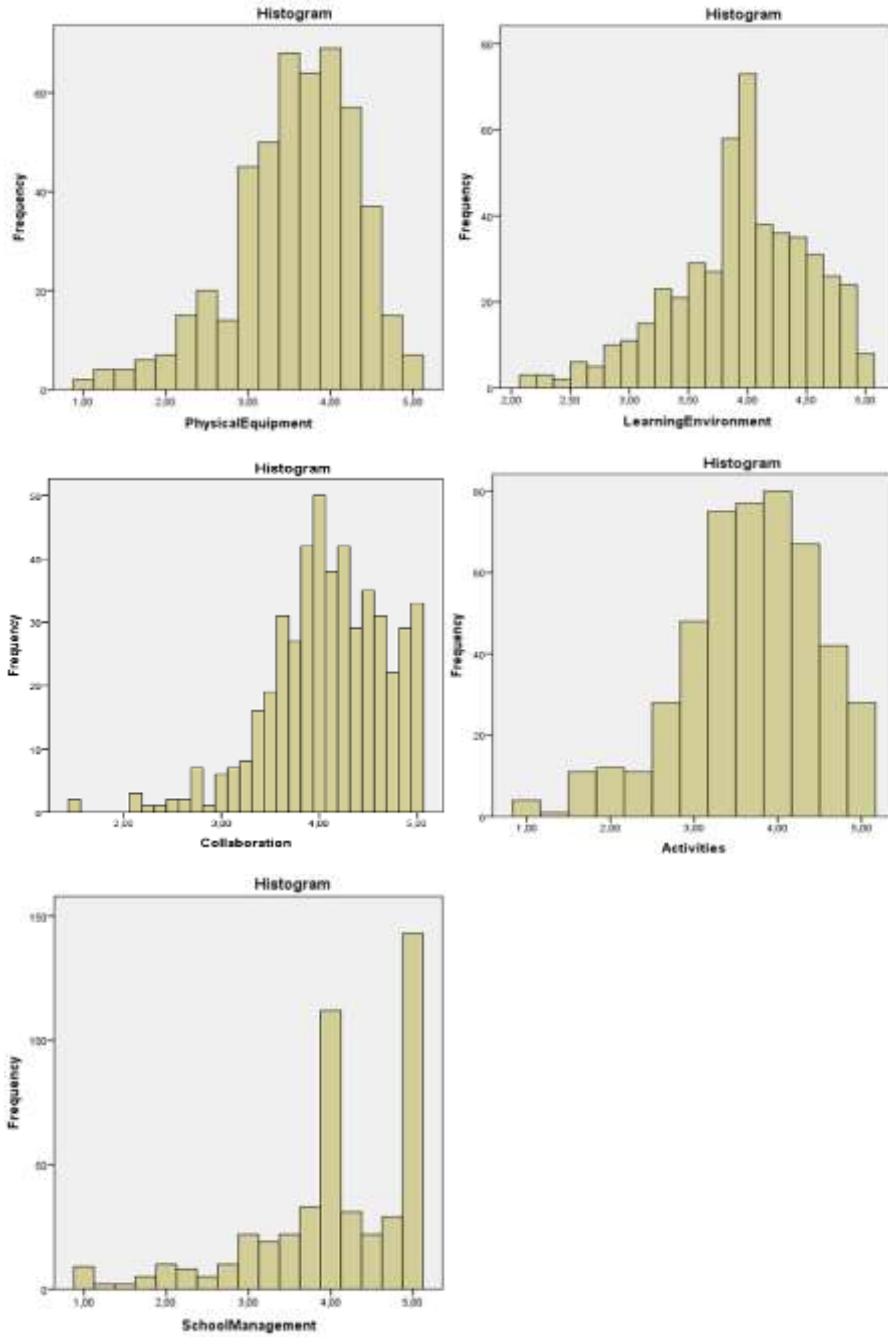
Figure 1. The Histograms Related to the Frequencies of Teacher Opinions in Sub-themes.

The School Happiness Scores in terms of Gender, Professional Seniority, School Type and Teaching Field Variables

In Table 7, independent t-test results were given in terms of gender variable.

Table 7

School Happiness Scores in terms of Gender Variable (N=484)

School Happiness	Gender	n	$\bar{x}$	SD	t	df	p
Physical equipment	Female	309	3.45	.76	4.345	482	.211
	Male	175	3.76	.69			
Learning environment	Female	309	3.92	.58	1.258	482	.740
	Male	175	3.99	.58			
Collaboration	Female	309	4.06	.59	1.540	482	.289
	Male	175	4.15	.61			
Activities	Female	309	3.64	.79	.711	482	.721
	Male	175	3.69	.83			
School management	Female	309	3.98	.93	2.047	482	.583
	Male	175	4.16	.94			

p< .05

In Table 7, it was seen that the school happiness scores did not differ statistically in terms of gender variable. These results could be interpreted as the fact that both male and female teachers' perception of happiness in school was similar.

ANOVA results showed that the mean scores did not differ in terms of professional seniority. On the other hand, the mean scores differed significantly in terms of school type and teaching field variances.

In Table 8, the ANOVA results were given in terms of school type variable.

Table 8

ANOVA Results in terms of School Type Variable (N=484)

School Happiness		Sum of Squares	df	Mean Square	F	p	Significant Difference
Physical Equipment	Between Groups	4.741	3	1.580	2.797	.040*	1>2
	Within Groups	271.173	480	.565			
Learning Environment	Between Groups	22.134	3	7.378	24.819	.000*	1>2,3,4
	Within Groups	142.693	480	.297			
Collaboration	Between Groups	7.728	3	2.576	7.437	.000*	1>3,4
	Within Groups	166.240	480	.346			
Activities	Between Groups	37.658	3	12.553	21.427	.000*	3<1,2,4
	Within Groups	281,197	480	.586			
School Management	Between Groups	15.512	3	5.171	5.973	.001*	2>3,4
	Within Groups	415.483	480	.866			

p< .05

In Table 8, it was seen that the mean scores differed statistically in terms of school type variable in Physical Equipment sub-dimension [F $_{(3-480)}$ =2.797, p< .05], in Learning Environment sub-dimension [F $_{(3-480)}$ =24.819, p< .05], in Collaboration sub-dimension [F $_{(3-480)}$ =7.437, p< .05], in Activities sub-dimension [F $_{(3-480)}$ =21.427, p< .05] and in School Management sub-dimension [F $_{(3-480)}$ =5.793, p< .05]. According to Scheffe test results, the difference in Physical Equipment sub-dimension was between the mean scores of teachers working in primary schools (x̄=3.71, S=.76) and the teachers working in secondary schools (x̄=3.44, S=.78). The difference in Learning Environment sub-dimension was among the mean scores of teachers working in primary schools (x̄=4.27, S=.45) and the teachers working in secondary schools (x̄=3.99, S=.46), in vocational high schools (x̄=3.71, S=.63), and in academic high schools (x̄=3.80, S=.61). The difference in Collaboration sub-dimension was among the mean scores of teachers working in primary schools (x̄=4.24, S=.55), the teachers working in vocational high schools (x̄=3.98, S=.67), and the teachers working in academic high schools (x̄=3.96, S= .63). The difference in Activities sub-dimension was among the mean scores of teachers working in vocational high schools (x̄=3.20, S=.84) and the teachers working in primary schools (x̄=3.66, S=.75), the teachers working in secondary schools (x̄=3.88, S=.66) and the teachers working in academic high schools (x̄=3.89, S=.80). The difference in School Management sub-dimension was between the mean scores of teachers working in secondary schools (x̄=4.30, S=.70), the teachers working vocational high schools (x̄=3.85, S=1.0), and the teachers working in academic high schools (x̄=3.91, S=.98).

In Table 9, the ANOVA results were given in terms of teaching field variable.

Table 9

ANOVA Results in terms of Teaching Field (N=484)

School Happiness		Sum of Squares	df	Mean Square	F	p	Sig. Dif. (Scheffe)
Physical Equipment	Between Groups	4.632	4	1.158	2.045	.087	-
	Within Groups	271.281	479	.566			
Learning Environment	Between Groups	17.536	4	4.384	14.257	.000*	1>2,3,4,5
	Within Groups	147.291	479	.307			
Collaboration	Between Groups	5.651	4	1.413	4.020	.003*	1>4
	Within Groups	168.316	479	.351			
Activities	Between Groups	22.791	4	5.698	9.218	.000*	5<1,2,3,4
	Within Groups	296.065	479	.618			
School Management	Between Groups	6.701	4	1.675	1.891	.111	-
	Within Groups	424.294	479	.886			

p< .05

In Table 9, it was seen that the mean scores differed statistically in terms of teaching field in Learning Environment sub-dimension *[F$_{(4-479)}$ =14.257, p< .05],* in Collaboration sub-dimension *[F $_{(4-479)}$ =4.020, p< .05],* and in School Management *dimension [F $_{(4-479)}$ =1.891, p< .05].* According to Scheffe test results, the difference in Learning

Environment sub-dimension was among the mean scores of the classroom teachers ($\bar{x}$=4.26, S=.45), science teachers ($\bar{x}$=3.90, S=.51), social sciences teachers ($\bar{x}$=3.91, S=.59), arts, and sports teachers ($\bar{x}$=3.78, S=.65), and vocational education teachers ($\bar{x}$=3.68, S=.60). The difference in Collaboration sub-dimension was between the mean scores of the classroom teachers ($\bar{x}$=4.25, S=.55) and arts, and sports teachers ($\bar{x}$=3.96, S=.66). The difference in Activities sub-dimension was among the mean scores of the vocational education teachers ($\bar{x}$=3.10, S= .85), classroom teachers ($\bar{x}$=3.63, S=.55), science teachers ($\bar{x}$=3.75, S=.81), social sciences teachers ($\bar{x}$=3.79, S=.77), and arts, and sports teachers ($\bar{x}$=3.76, S=.80).

Discussion, Conclusion and Recommendations

In this research aiming to develop and implement the School Happiness Scale, the findings indicated that teachers perceived moderate level school happiness, in general. The highest scores were found in 'Collaboration' and 'School Management' sub-dimensions. The collaboration among teachers and the support of school administrators were the most important factors that increased school happiness of teachers. Similarly, in a study conducted by Boehm and Lyubomirsky (2008), the findings show that the employees feel more happiness if they receive more social support from their managers and colleagues, and encounter more cooperative approaches when interacting with others. In another studies, conducted by Aelterman, Engels, Van Petegem, & and Verhaeghe (2007), and Yildirim (2014), the findings show that supports of school principal and colleagues, the relationship with parents, cooperation among staff, fair and helpful assessment and feedback increase teachers' well-being in school. The results also show that the teachers perceive a moderate level of school happiness in 'Physical Equipment' and 'Activities' sub-dimensions. In numerous studies, physical equipment of school and the extracurricular activities are seen as the mediators for school well-being (Asici & Ikiz, 2018; Bakioglu & Bahceci, 2010; Bird & Markle, 2012; Demiriz & Ulutas, 2016; Engels, Aelterman, Petegem, & Schepens, 2004). These results are consistent with the findings of previous studies. Similarly, in a study conducted by Talebzadeh and Samkan (2011), the findings show that creating a green space in school, ensuring healthy foods at school, benefiting from various and suitable educational aids for more attraction in classroom, paying attention to the art in school programs, and establishing sport teams which include teachers and students, increase school happiness both for teachers and students.

The results revealed that school happiness scores did not differ statistically in terms of gender and professional seniority variables. These results can be interpreted as the fact that male and female teachers have similar perceptions in terms of school happiness. In addition, ANOVA results showed that school happiness scores did not differ statistically in terms of professional seniority variable. These findings can be interpreted as professional seniority has no significant effect on teachers' perceptions in terms of school happiness. ANOVA results also showed that the mean scores differed statistically in terms of school type variable, indicating that teachers working in primary schools felt happier than the teachers working in secondary schools in

terms of physical equipment. Actually, the mean scores did not differ in all school types in terms of Physical Equipment sub-themes, in general. These results can be interpreted as the fact that physical conditions have less significant impact on the school happiness of teachers. However, Talebzadeh and Samkan (2011) found that providing a green space in school, providing healthy foods at school, benefiting from various and suitable educational equipment for more attraction in classroom, increase school happiness in terms of elementary school teachers and students.

ANOVA results showed that teacher opinions differed significantly in Learning Environment, Collaboration and Activities sub-themes in terms of school type variable. Compared with other school types, it was seen that level of school happiness was higher in primary school teachers in Learning Environment sub-theme. Based on the findings it can be said that the education model, which is based on academic achievement, elimination and competition decreases the happiness level of teachers in secondary and high schools. Similar results are seen in previous studies. On the other hand, job stress affects teachers' school happiness negatively and job satisfaction affects it positively (Collie, Shapka & Perry, 2012; Klassen & Chiu, 2011). In a study, conducted by Ozkan (2017) the findings show that general job satisfaction level of the primary school teachers is higher than the teachers working in secondary schools. Similarly, Kumas and Deniz (2010) found that teachers working in primary schools have more job-satisfaction compared to teachers working in secondary schools and high schools. In Collaboration sub-theme, school happiness scores of the teachers working in primary schools were higher than the teachers working in vocational high schools and academic high schools. Actually, collaborative teacher attitudes increased teachers' effectiveness in their role of inspiring creativity as well as providing them to be role-model to students in terms of learning in a happy environment. In a research on school happiness conducted by UNESCO (2016), the participants state that teachers need to develop a sense of belonging and a collective identity, and that this could be developed through teamwork and a collaborative spirit, not only among teachers but also among students. In Activities sub-theme, the school happiness scores of teachers working in vocational high schools were lower than the teachers working in primary schools, secondary schools, and academic high schools. It is clear that social-cultural activities and sport activities in schools increase school happiness of teachers by creating a positive school climate. Findings of previous studies show that teachers working in vocational high schools have low job satisfaction and low well-being, in general. In a study conducted by Argon and Cicioglu (2017), teachers working in vocational high schools reported a very low level of educational beliefs and motivation for teaching. School happiness scores of teachers working in secondary schools were higher than the teachers working in vocational high schools, and academic high schools, in School Management sub-theme. The supportive attitudes of school managers affect teachers' school well-being. A study conducted by Aelterman, Engels, Van Petegem and Verhaeghe (2007) support that school principals increase teachers' well-being.

ANOVA results also showed that classroom teachers perceived more school happiness in Learning Environment sub-theme. It is clear that classroom teachers had

a positive judgment about their responsibility to nurture basic qualifications and endowments of the pupils. As a result of these assumptions, classroom teachers increased their intrinsic motivation and the commitment to teaching profession. In addition, the achievements of students affected teachers' well-being positively. In a study conducted by Caprara, Barbaranelli, Steca and Malone (2006), the findings show that if teachers can testify to the overall development of their students, this observation allows teachers to see the results of their efforts and increase their job satisfaction. Moreover, in activities sub-theme, the vocational education teachers felt lower level of school happiness. The highly stressful nature of the teaching profession as well as the high demands from teachers due to changes in curriculum constantly, insufficient resources, and inadequate supports decreased the success of school; and all these factors brought about high stress and low job satisfaction on teachers. Eventually, these factors caused to decrease the level of school happiness of teachers. In a study conducted by Yildirim (2014), teachers feel more school well-being when they apply student-oriented learning practices, and have adequate school climate and classroom climate.

The results showed that, teachers perceived school happiness at a moderate level in Physical Equipment and Activities sub-dimensions. In addition, they perceived higher school happiness in Collaboration and School Management sub-themes. The classroom teachers perceived higher level of school happiness than the teachers working in secondary schools, academic high schools and vocational high schools. The vocational education teachers felt lower school happiness in activities sub-theme. Based on the results, it can be suggested that physical conditions of schools should be improved. In order to increase the school happiness levels of teachers, school administrators should improve collaboration among the school community members. In schools, learning environment should be redesigned according to student interests. School principals should support teachers and display democratic and fair school management attitudes. In addition, social activities and sportive activities should be increased, and guidance for students should be improved in schools. In this study, school happiness was examined based on teacher opinions. Further studies can be conducted with different participants such as students, parents or school managers. In addition, further research can be planned applying different research models to determine participants' (teachers, students, school administrators, etc.) school happiness.

References

Acton, R., & Glasgow, P. (2015). Teacher wellbeing in neoliberal contexts: A review of the literature. *Australian Journal of Teacher Education, 40* (8), 99-115. Retrieved 18 July, 2018, from http://dx.doi.org/10.14221/ajte.2015v40n8.6

Aelterman, A., Engels, N., Van Petegem, K., & Verhaeghe, J. P. (2007). The wellbeing of teachers in Flanders: The importance of a supportive school culture. *Educational Studies, (33)* 3, 285-298.

Ahn, N., Garcia, J. R., & Jimeno, J. F. (2004). The impact of unemployment on individual well-being in the EU, Europen Network of Economic Policy Research Institutes, Working paper, No: 29. Retrieved 11 July, 2018, from https://www.files.ethz.ch/isn/110596/WP%20029.pdf.

Argon, T., & Cicioglu, M. (2017). Meslek lisesi ogretmenlerinin egitime inanma düzeyleri ile öğretme motivasyonlari. [Vocational high school teachers' levels of belief in education and motivation to teach]. *The Journal of Academic Social Science Studies, 57* (1), 1-23. Retrieved 17 July, 2018, from http://dx.doi.org/10.9761/JASSS7108

Asici, E., & Ikiz, F. E. (2018). Okulda oznel iyi oluşun okul iklimi ve oz-yeterlik acisindan yordanmasi. [The prediction of subjective well-being in school in terms of school climate and self-efficacy]. *H. U. Journal of Education.* doi: 10.16986/HUJE.2018038523.

Aydin, A. (2016). *Mutluluk.* [Happiness]. (First Edition). Ankara: PegemA Yayincilik.

Bakioglu, A., & Bahçeci, M. (2010). Velilerin okul imajina ilişkin görüşlerinin incelenmesi. [Review of parents' perceptions in relation with school image]. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi, 31* (31), 25-55.

Bird, J. M., & Markle, R. S. (2012). Subjective well-being in school environments: Promoting positive youth development through evidence-based assessment and intervention. *American Journal of Orthopsychiatry, 82* (1), 61-66. DOI: 10.1111/j.1939-0025.2011.01127.x

Boehm, J., & Lyubomirsky, S. (2008). Does happiness promote career success? *Journal of Career Assessment, 16* (1), 101-116. Retrieved 11 July, 2018, from https://doi.org/10.1177/1069072707308140.

Buyuksahin-Cevik, G., & Yildiz, M. A. (2016). Pedagojik formasyon öğrencilerinde umutsuzluk ile mutluluk arasindaki iliskide benlik saygisinin aracilik rolu. [The mediating role of self-esteem on the relationship between hopelessness and happiness in pedagogy formation students]. *Dicle University Journal of Ziya Gokalp Faculty of Education, 27,* 96-107. Retrieved 11 July, 2018, from http://dx.doi.org/10.14582/DUZGEF.704.

Buyukozturk, S., Kilic-Cakmak, E., Akgun, O. E., Karadeniz, S., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri.* [Scientific research methods]. Ankara: PegemA Yayincilik.

Byrne, B. M. (2011). *Structural equation modeling with AMOS Basic concepts, applications, and programming (Multivariate Applications Series).* New York: Routledge.

Caprara G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44* (2006) 473-490.

Certel, Z., Bahadir, Z., Saracaloglu, S., & Varol, R. (2015). The investigation of the relation between the high school students' self-efficacy and subjective well-being. *Journal of Research in Education and Teaching, 4* (2), 307-318.

Clair, A. (2012). The relationship between parent's subjective well-being and the life satisfaction of their children in Britain. *Child Indicators Research, 5* (4), 631-650. Retrieved 11 July, 2018, from https://doi.org/10.1007/s12187-012-9139-5.

Chaplin, L. N. (2009). Please may I have a bike? Better yet, may I have a hug? An examination of children's and adolescents' happiness. *Journal of Happiness Studies, 10* (5), 541-562.

Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social-emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology, 104,* 1189-1204. Retrieved 18 July, 2018, from http://dx.doi.org/10.1037/a0029356

Cokluk, O., Sekercioglu, G., & Buyukozturk, S. (2012). *Sosyal bilimler için çok degiskenli SPSS ve LISREL uygulamaları.* [Multivariate SPSS and LISREL applications for social sciences.] Ankara: PegemA Yayincilik.

Comlekci, N. (2001). *Bilimsel arastirma yontemi ve istatistiksel anlamlilik sinamalari.* [Scientific research methods and statistical significance tests]. Ankara: Bilim Teknik Yayınevi.

Demir-Celebi, C., & Sezgin, O. (2015). Lise öğrencilerinin öznel iyi oluşlari ile ahlâkî olgunluk seviyeleri arasindaki ilişkinin incelenmesi. [Relationship between subjective well-being and moral maturity levels of high school students]. *Kalem Eğitim ve İnsan Bilimleri Dergisi, 5*(2), 99-146.

Demiriz, S., & Ulutaş, I. (2016). Çocuklar ne kadar mutlu? Bazı değişkenlere göre çocuklarda mutluluğun belirlenmesi. [How happy are children? Determining happiness according to some variables]. *Adnan Menderes Universitesi Egitim Fakultesi Egitim Bilimleri Dergisi, 7* (1), 16-24.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95* (3), 542-575. Retrieved 18 July, 2018, from http://dx.doi.org/10.1037/0033-2909.95.3.542.

Dogan, T., Sapmaz, F., & Akıncı-Cotok, N. (2013). Ozeleştiri ve mutluluk. [Self-criticism and happiness]. *Kastamonu Eğitim Dergisi, 21* (1), 391-400.

Engels, N., Aelterman, A., Petegem, K. V., & Schepens, A. (2004). Factors which influence the well-being of pupils in Flemish secondary schools. *Educational Studies, 30* (2), 127-143. Retrieved 18 July, 2018, from http://hdl.handle.net/1854/LU-214624.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics.* (3rd Ed.). Cambridge: Cambridge University Press.

Fraenkel, J., & Wallen, N. E. (2012). *How to design and evaluate research in education*. Columbus, OH: McGraw-Hill.

Fredrickson, B. L. (2013). Positive emotions broaden and build. In Patricia Devine, & Ashby Plant. (Eds). *Advances in experimental social psychology.* (pp. 1-53). Burlington: Academic Press.

Gundogdu, R., & Yavuzer, Y. (2012). Ogretmen adaylarinin öznel iyi olus ve psikolojik ihtiyaclarinin demografik degiskenlere göre incelenmesi. [Examining subjective well-being and psychological needs of students at the educational faculty according to demographic variables]. *Mehmet Akif Ersoy Universitesi Egitim Fakultesi Dergisi, 12* (23), 115-131.

Holder, M. D., & Klassen, A. (2010). Temperament and happiness in children. *Journal of Happiness Studies, 11* (4), 419-439. doi: 10.1007/s10902-009-9149-2.

Huebner, E. S. (1991). Correlates of life satisfaction in children. *School Psychology Quarterly, 6* (2), 103-111.

Karasar, N. (2017). *Bilimsel arastirma yontemi: Kavramlar, ilkeler, teknikler*. [Scientific research method: Concepts, principles, techniques]. Ankara: Nobel Akademi Yayinlari.

Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36,* 114-129. Retrieved 18 July, 2018, from http://dx.doi.org/10.1016/j.cedpsych.2011.01.002

Kline, R. B. (2005). *Principles and practice of structural equation modelling*. New York, NY: Guilford Press.

Koknel, O. (1992). *Dolu dolu yasamak.* [Live life fully]. (First Ed.). Istanbul: Altin Kitaplar.

Kumas, V., & Deniz, L. (2010). An investigation about job satisfaction of teachers. *M.U. Ataturk Egitim Fakültesi Egitim Bilimleri Dergisi, 32,* 123-139.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

López-Pérez, B., Sánchez, J., & Gummerum, M. (2015). Children's and adolescents' conceptions of happiness. *Journal of Happiness Studies, 17* (6), 2431-2455. doi: 10.1007/s10902-015-9701-1.

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology, 9* (2), 111-131. Retrieved 11 July, 2018, from http://dx.doi.org/10.1037/1089-2680.9.2.111.

Lyubomirsky, S., & King, L. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131* (6), 803-855, doi: 10.1037/0033-2909.131.6.803.

Mahon, N. E., & Yarcheski, A. (2002). Alternative theories of happiness in early adolescents. *Clinical Nursing Research, 11* (3), 306-323.

McCabe, K., Bray, M. A., Kehle, T. J., Theodore, L. A., & Gelbar, N. W. (2011). Promoting happiness and life satisfaction in school children. *Canadian Journal of School Psychology, 26* (3), 177-192. DOI:10.1177/0829573511419089.

Ozdemir, Y., & Koruklu, N. (2011). Universite öğrencilerinde degerler ve mutluluk arasindaki iliskinin incelenmesi. [Investigating relationship between values and happiness among university students]. *YYÜ, Egitim Fakultesi Dergisi, 8* (1), 190-210.

Ozkan, A. (2017). *Job satisfaction levels of the primary schools teachers and the secondary schools teachers (Balıkesir central district sample).* (Unpublished master's thesis). Balıkesir Üniversitesi, Sosyal Bilimler Enstitüsü, Balıkesir, Türkiye.

Ozturk, A., & Cetinkaya, R. S. (2015). Egitim fakultesi ogrencilerinin oznel iyi olus duzeyleri ile tinsellik, iyimserlik, kaygi ve olumsuz duygu düzeyleri arasindaki iliski. [Relationship between education faculty students' subjective well-being, spirituality, optimism, anxiety and negative affectivity]. *Marmara Universitesi Ataturk Egitim Fakültesi Egitim Bilimleri Dergisi, 42* (42), 335-356.

Ozturk, L., Meral, I. G., & Yılmaz, S. S. (2017). Lisans ogrencilerinin mutluluk ve dindarlik iliskisi: Kırıkkale Universitesi ornegi. [The relationship between happiness and religiosity for undergraduate students: A case of Kırıkkale University]. *Akademik Yaklaşımlar Dergisi, 8* (1), 23-39.

Pan, J., & Zhou, W. (2013). Can success lead to happiness? The moderators between career success and happiness. *Asia Pacific Journal of Human Resources, 51* (1), 63-80, doi:10.1111/j.1744-7941.2012.00033.x.

Park, N., & Peterson, C. (2006). Character strengths and happiness among young children: content analysis of parental descriptions. *Journal of Happiness Studies, 7* (3), 323-341. doi: 10.1007/s10902-005-3648-6.

Sarıcam, H. (2014). Belirsizliğe tahammulsuzlugun mutluluga etkisi. [The effect of intolerance of uncertainty on happiness]. *Sosyal Bilimler Dergisi, 4* (8), 1-11.

Schnittker, J. (2008). Happiness and success: Genes, families, and the psychological effects of socioeconomic position and social support. *American Journal of Sociology, 114* (1), 233-259.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338.

Seligman, M. E. P., Parks, A. C., & Steen, T. (2004). A balanced psychology and a full life. *Biological Sciences, 359* (1449), 1379-1381. doi: 10.1098/rstb.2004.1513.

Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being.* New York: Simon & Schuster.

Selim, S. (2008). Turkiye'de bireysel mutluluk kaynagı olan degerler üzerine bir analiz: Multinomial logit model. [An analysis on the values of individual happiness sources in Turkey: Multinomial logit model]. *Cukurova Universitesi Sosyal Bilimler Enstitusu Dergisi, 17* (3), 345-358.

Sumer, N. (2000). Yapisal esitlik modelleri. *Turk Psikoloji Yazıları, 3*(6), 49-74.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. (6th Ed.) Boston: Pearson.

Talebzadeh, F., & Samkan, M. (2011). Happiness for our kids in schools: A conceptual model. *Procedia-Social and Behavioral Sciences, 29* (2011), 1462-1471.

Telef, B. B. (2014). School children's happiness inventory: The validity and reliability study. *International Online Journal of Educational Sciences, 6* (1), 130-143.

Terzi, S. (2017). Oğretmenlerin is doyumlari ile mutluluk düzeyleri arasindaki iliskinin incelenmesi. [Analysis of the relationships between teacher's job satisfaction and their happiness levels]. *21. Yüzyılda Egitim ve Toplum, 6* (17), 475-487.

Thoilliez, B. (2011). How to grow up happy: An exploratory study on the meaning of happiness from children's voices. *Child Indicator Research, 4* (2), 323-351. DOI: 10.1007/s12187-011-9107-5.

Turkdogan, T., & Duru, E. (2012). The role of basic needs fulfilment in prediction of subjective well-being among university students. *Educational Sciences: Theory & Practice, 12* (4), 2440-2446.

Turkmen, M. (2012). Öznel iyi oluşun yapisi ve anabab tutumlari, özsaygi ve sosyal destekle ilişkisi: Bir model sınaması. [The structure of subjective well-being and its relationship with parenting style, self-esteem and social support: A model test]. *Uşak Üniversitesi Sosyal Bilimler Dergisi, 5* (9), 41-73.

Ucan, A., & Kiran-Esen, B. (2015). Ergenlerin öznel iyi oluslarinin toplumsal konum ile ilgili risk alma degiskenine göre incelenmesi. [Analyzing the subjective well-being of teenagers in relation to social status-related risk taking]. *Mersin Universitesi Egitim Fakultesi Dergisi, 11* (2), 288-299. DOI: 10.17860/efd.22192.

UNESCO (2016). *Happy schools: A framework for learner well-being in the Asia-Pacific*. Retrieved 11 July, 2018, from http://unesdoc.unesco.org/images/ 0024/002441/244140e.pdf

Uusitalo-Malmivaara, L. (2012). Global and school-related happiness in Finnish children. *Journal of Happiness Studies, 13* (4), 601-619. doi: 10.1007/s10902-011-9282-6.

Uusitalo-Malmivaara, L., & Lehto, J. E. (2013). Social factors explaining children's subjective happiness and depressive symptoms. *Social Indicators Research, 111* (2), 603-615. doi: 10.1007/s11205-012-0022-z.

Unuvar, P., Calısandemir, F., Tagay, O., & Amini, F. (2015). Okul öncesi dönem çocuklarinin mutluluk algisi (Türkiye ve Afganistan Örneği). [Preschool children's perception of happiness (Turkey and Afghanistan sample)]. *Mehmet Akif Ersoy Universitesi Egitim Fakultesi Dergisi, 15* (34), 1-22.

Van Hal, G., Bruggeman, B., Aertsen, P., Gabriels, J., Marechal, E., Rotsaert, J., Mortelmans, W., & Van Dongen, S. (2014). A survey on happiness in primary school children in Flanders. *European Journal of Public Health, 24* (2), 247-248.

Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of Happiness Studies, 9* (3), 449-469. Retrieved 11 July, 2018, from http://dx.doi.org/10.1007/s10902-006-9042-1.

Weaver, R. D., & Habibov, N. H. (2010). Are Canadian adolescents happy? A gender-based analysis of a nationally representative survey. *US-China Education Review, 7* (4), 37-52.

Yaliz-Solmaz, D. (2014). Öğretmen adaylarinin öznel iyi oluş düzeyleri: Anadolu Üniversitesi Beden Eğitimi ve Spor Öğretmenliği Bölümünde bir araştirma. [The subjective well-being levels of teacher candidates: a research department of physical education and sports teaching at Anadolu University]. *Uluslararası Sosyal Araştırmalar Dergisi, 7* (35), 651-657.

Yildirim, K. (2014). Main factors of teachers' professional well-being. *Educational Research and Reviews, 9* (6), 153-163. DOI: 10.5897/ERR2013.1691.

Yucel, D., & Vogt-Yuan, A. S. (2016). Parents, siblings, or friends? Exploring life satisfaction among early adolescents. *Applied Research Quality Life, 11* (4), 1399-1423. DOI: 10.1007/s11482-015-9444-5.

Okul Mutluluğu: Ölçek Geliştirme ve Uygulama Çalışması

Atıf:

Özet

Problem Durumu: 20. Yüzyıl boyunca eğitim sistemleri, ekonomik kalkınma ve rekabetin artmasına bağlı olarak, öğrencilerin özellikle iş dünyası için üretken bireyler olarak hazırlanmasına odaklanmıştır. 21. Yüzyılda ise eğitim sistemlerinden artan nüfus hareketliliği, yaşam stresi, eşitsizlikler, küresel ısınma ve çevresel kaygıların yanı sıra yaşanan hızlı değişim ile baş edebilecek bireyler yetiştirmesi beklenmektedir. Toplumların mutsuzluk düzeyleri arttıkça okullar, eğitim hizmeti sunmayı amaçlayan kurumlar olmanın yanı sıra öğrencilerin sosyal ve duygusal gelişimlerine ve mutlu bir

şekilde öğrenmelerine olanak sağlayan kurumlar olarak yeniden yapılandırmaya gereksinim duymaktadır. Mutluluk sadece eğitimin bir amacı değil, aynı zamanda okul etkililiğinin de bir faktörüdür. Mutluluk, bireyin neşe, sevinç, umut, bedensel ve ruhsal esenlik gibi duyguları olarak tanımlanabilir. Başka bir tanımda mutluluk, yaşamın bir bütün olarak değerlendirmesi, genel olarak yaşamdan alınan memnuniyet, olumsuz duygulardan daha çok olumlu duygulara sahip olma durumudur. Mutluluk, yaşamın anlamını, olumlu duyguları, sorumluluğu, olumlu ilişkileri ve başarıyı içeren çok boyutlu bir yapı olarak da tanımlanır. Yetişkinlerin mutluluğu, sağlık, aile hayatı, sosyal ilişkiler, güvenlik, özgürlük, ahlaki değerler, gelir düzeyi, çalışma koşulları gibi birçok değişkene bağlı iken; çocukların mutluluğu, temel ihtiyaçlarının karşılama düzeyi, sevgi, güven, iletişim, tanınma, okul başarısı, sağlık veya oyun gibi pek çok değişkenle ilişkilendirilmektedir.

Okul mutluluğu, okul amaçları ile öğrencilerin, öğretmenlerin, okul yöneticilerinin kişisel ihtiyaçlarının karşılanması ve diğer çevresel faktörlere bağlı olarak çalışanlar arasındaki uyum sonucunda ortaya çıkan duygusal iyilik hali şeklinde ifade edilir. Öğretmenlerin okul mutluluğu, velilerle ilişkiler, meslektaşların desteği, öz-yeterlik, iş yükü, yeniliklere karşı olumlu tutumlar ve okul müdürünün desteği ile ilişkilendirilmektedir. Mutlu öğretmenler, öğrencilerin öğrenme sürecinde tehdit oluşturmayan bir ortam sağlayarak öğrenciler için kolaylaştırıcı, planlayıcı, öğretici, arabulucu ve açıklayıcı rol oynamaktadır. Öğretmenlik mesleğinin son derece stresli doğası, öğretim programlarının sürekli değişmesi nedeniyle öğretmenlerden yüksek beklentiler, kaynak azlığı ve okula sağlanan desteğin yetersizliği, okulların başarısını azaltmakta ve bunların hepsi öğretmenlerde yüksek strese, düşük iş doyumuna ve umutsuzluğa neden olmaktadır. Bu durum, öğretmenlerin düşük düzeyde okul mutluluğu algılamasına neden olmaktadır.

Araştırmanın Amacı: Bu çalışmanın amacı, öğretmenlerin okul mutluluğu düzeyinin belirlenmesidir. Bu amaçla 'Okul Mutluluğu Ölçeği' geliştirilerek, geçerlik ve güvenirlik çalışmaları yapılmıştır. Ölçek, öğretmenlerin okul mutluluk düzeyini belirlemek amacıyla öğretmenlere uygulanmıştır. Elde edilen verilere dayalı olarak öğretmenlerin okul mutluluğu düzeyi belirlenmeye çalışılmış, okul mutluluğu puanlarının öğretmenlerin cinsiyeti, mesleki kıdemleri, okul türü ve öğretim alanı değişkenleri açısından ne düzeyde farklılaştığı ortaya konulmaya çalışılmıştır.

Araştırmanın Yöntemi: Öğretmenlerin okul mutluluğu düzeyini öğretmen görüşlerine dayalı olarak belirlemeyi amaçlayan bu araştırma, tarama modelinde tasarlanmıştır. Tarama modeli, bir grubun belirli özelliklerini belirlemek için kullanılmaktadır. Bu çalışmada öğretmenlerin okul mutluluğu düzeyleri belirlenmeye çalışılmıştır. Araştırmanın ölçek geliştirme sürecinde, 450 öğretmenin görüşüne başvurulmuştur. 430 öğretmen tarafından ölçekteki tüm maddelere tam ve amaçlara uygun olarak verilen yanıtlar üzerinden Açımlayıcı Faktör Analizi gerçekleştirilmiştir. Doğrulayıcı Faktör Analizi için 449 öğretmenin görüşlerine başvurulmuştur. Araştırmanın verileri, araştırmacılar tarafından geliştirilen 'Okul Mutluluğu Ölçeği' aracılığı ile elde edilmiştir. Okul Mutluluğu Ölçeği, 26 maddeden oluşmakta ve madde toplam korelasyonları .54 ile .86 arasında değişmektedir. Beş faktörlü bir yapıya sahip olan ölçekte, beş faktör (fiziksel donanım, öğrenme ortamı, işbirliği, etkinlikler, okul

yönetimi) toplam varyansın % 65.09'unu açıklamaktadır. Öğretmenlerin okul mutluluğu düzeyini belirlemede, küme örnekleme yöntemi ile belirlenen 484 öğretmenin görüşüne başvurulmuştur. Analiz sürecinde önce verilerin dağılımın normalliği test edilmiştir. Dağılımın normal olduğu belirlendikten sonra verilerin analizinde, betimsel istatistik (ortalama, standart sapma, yüzde), bağımsız t testi ve ANOVA'dan yararlanılmıştır.

Araştırmanın Bulguları: Bulgular, öğretmenlerin genel olarak orta düzeyde okul mutluluğu algıladıklarını göstermektedir. En yüksek puanlar, 'İşbirliği' ve 'Okul Yönetimi' alt boyutlarındadır. Öğretmenler arasındaki işbirliği ve okul yöneticilerinin desteği, öğretmenlerin okul mutluluğunu arttıran en önemli faktörlerdir. Bulgular, öğretmenlerin 'Fiziksel Donanım' ve 'Etkinlikler' alt boyutlarında orta düzeyde okul mutluluğu algıladıklarını göstermektedir. Araştırmanın bulguları, okul mutluluk puanlarının cinsiyet ve meslek kıdem değişkenleri açısından istatistiksel olarak farklılık göstermediğini ortaya koymaktadır. ANOVA sonuçları, öğretmenlerin okul mutluluğu düzeylerinin Öğrenme Ortamı alt boyutunda, ilkokullarda görevli öğretmenler lehine istatistiksel olarak anlamlı bir farklılık gösterdiğini ortaya koymaktadır. Araştırmalardan elde edilen bulgular, sınıf öğretmenlerinin öğrencilerin temel niteliklerini ve yeteneklerini yetiştirme sorumluluğu konusunda olumlu bir yargıya sahip olduklarını ortaya koymaktadır. Bulgulara dayalı olarak, ilkokullarda görevli öğretmenlerin içsel motivasyonlarının ve öğretmenlik mesleğine olan bağlılığının yüksek olduğu söylenebilir. Ayrıca bu sonuçlar, akademik başarıya, eleme ve rekabete dayalı eğitim modelinin, ortaokul ve liselerde öğretmenlerin mutluluk düzeyini azalttığı şeklinde yorumlanabilir. Etkinlikler alt boyutunda, meslek liselerinde görev yapan öğretmenlerin okul mutluluğu puanları ilkokul, ortaokul ve akademik liselerde çalışan öğretmenlerden daha düşüktür. Yeterli sosyal-kültürel etkinlikler ve spor etkinlikleri, okullarda olumlu bir iklim yaratmakta, bu durum öğretmenlerin okul mutluluğunu artırmaktadır. İşbirliği alt temasında, ilkokul öğretmenlerinin okul mutluluğu puanları meslek lisesi ve akademik liselerde görevli öğretmenlerden yüksektir. İşbirlikli okul ortamı, öğretmenlerin yaratıcılığını artırmanın yanı sıra öğretmenlerin öğrencilere rol model olmalarını ve okulda mutlu bir öğrenme ortamı oluşturmalarını sağlar. Ayrıca, Etkinlikler alt boyutunda, meslek dersi öğretmenleri daha düşük düzeyde okul mutluluğu hissetmektedir. Öğretmenlik mesleğinin son derece stresli doğası ve öğretim programlarında sıklıkla gerçekleştirilen değişiklikler, öğretmenlerden yüksek beklentiler, okulların kaynaklarının yetersizliği ve öğretmenlere verilen desteğin yetersiz olması, okulun başarısını azaltmakta ve öğretmenler, yüksek stres ve düşük iş doyumu yaşamaktadır. Sonuç olarak, bu faktörler öğretmenlerin okul mutluluğunun azalmasına neden olmaktadır.

Sonuç ve Öneriler: Sonuçlar, öğretmenlerin Fiziksel Donanım ve Etkinlikler alt boyutlarda orta düzeyde okul mutluluğu algıladıklarını göstermektedir. İşbirliği ve Okul Yönetimi alt boyutlarında ise öğretmenler, yüksek düzeyde okul mutluluğu algılamaktadır. Sınıf öğretmenleri ortaokul, akademik lise ve meslek liselerinde görev yapan öğretmenlerden daha yüksek düzeyde okul mutluluğu algılamaktadır. Meslek dersi öğretmenleri, Etkinlikler alt boyutunda daha düşük düzeyde okul mutluluğu

hissetmektedir. Araştırmanın sonuçlarına dayalı olarak, okulların fiziki koşullarının iyileştirilmesinin öğretmenlerin okul mutluluğunu artıracağı öngörülebilir. Öğretmenlerin okul mutluluğunu arttırmak için okul müdürleri, okul toplumu arasındaki işbirliğini geliştirmelidir. Okullarda öğrenme ortamı, öğrencilerin ilgi alanlarına göre yeniden tasarlanmalıdır. Okul müdürleri, öğretmenleri teşvik etmeli ve okulda demokratik ve adil bir yönetim sergilemelidir. Okullarda öğrencilerin daha iyi öğrenmeleri için sosyal ve sportif faaliyetler artırılmalı ve rehberlik hizmetleri geliştirilmelidir. Bu çalışmada, öğretmenlerin okul mutluluğu düzeyleri araştırılmıştır. Öğrenciler, veliler veya okul yöneticilerinin okul mutluluğunu belirlemek amacıyla farklı araştırmacılar tarafından farklı araştırmalar yürütülebilir. Ayrıca farklı yerleşim birimlerinde öğretmenlerin okul mutluluğunu belirlemek için farklı araştırmalar planlanabilir. Okul Mutluluğu Ölçeği, farklı araştırmacılar tarafından veri toplama aracı olarak kullanılarak geçerlik ve güvenirliği test edilebilir.

Anahtar Sözcükler: Mutluluk, okul mutluluğu, öğretmen, ölçek geliştirme, geçerlik, güvenirlik.

An Inquiry into the Underlying Reasons for the Impact of Technology Enhanced Problem-Based Learning Activities on Students' Attitudes and Achievement[*]

Yucel CETIN[1], Seref MIRASYEDIOGLU[2], Erdinc CAKIROGLU[3]

ARTICLE INFO	ABSTRACT
	Purpose: In the constantly changing and evolving world of today, there is a need to raise individuals who are able to renew themselves, learn autonomously, and have higher order thinking skills such as critical thinking and deducing. The purpose of the present study is to investigate the effects of Technology Enhanced Problem-Based Learning Activities (TEPLA), developed with regard to the aforementioned need, on learners' attitudes towards mathematics and academic achievement in mathematics through qualitative data in order to designate the underlying causes of these effects. **Method:** In the research, conducted as a case study, the analysis of the written and oral interviews were held

with students using "Student Interview Form" in order to analyze the effect of TEPLA on the attitudes towards mathematics and mathematics achievements of study group students who were applied TEPLA by the first author.
Findings: The analysis revealed two main dimensions affecting the association between TEPLA and learners' attitudes towards mathematics: Affection and Usefulness, which comprised five and three distinct reasons, respectively. On the other hand, ten factors were defined to account for the impact of using TEPLA on learners' mathematics achievement.
Implications for Research and Practice: The research findings suggested that learners found TEPLA interesting and meaningful. Thus, it can be suggested that the impact of TEPLA on learner motivation can also be investigated using a motivation scale. Moreover, learners stated during interviews that they achieved retention in learning thanks to TEPLA. Thus, the finding regarding TEPLA's impact on retention can be further investigated quantitatively.

---

[1] Corresponding Author, Etimesgut Science and Art Centre, TURKEY, e-mail: yucelhoca1@gmail.com, ORCID: https://orcid.org/0000-0003-0718-3746

[2] Baskent University, TURKEY, e-mail: serefm@baskent.edu.tr, ORCID: https://orcid.org/0000-0001-9492-5992

[3] 3 Middle East Technical University,TURKEY, e-mail: cakiroglu.erdinc@gmail.com ORCID: https://orcid.org/0000-0003-1182-0752

Introduction

Today Mathematics Education does no longer adopt the traditional function of teaching merely the equational skills and abstractions, but it is rather driven by the purpose of providing the learners with the ability to use mathematics in daily life. These new trends have brought up the issue of using different instructional methods and integrating technology into education. Cakır and Aztekin (2016) reported that authentic tasks contribute significantly to learners' motivation, as they show learners how the knowledge they have learned work in daily life. Likewise, National Council of Teachers of Mathematics (1989) emphasized the principle… "To learn mathematics is to apply mathematics".

On the other hand, in our world, which constantly change and evolve thanks to what technology brings into our lives, there is an increased need to raise individuals who are able to renew themselves, learn autonomously, and have higher order thinking skills such as critical thinking and deducing. Therefore, Problem-Based Learning (PBL) offers a method compatible with the purpose of raising individuals with those higher order thinking skills, since PBL is characterized by learning mathematics by doing, activating learners' higher order thinking skills, teaching how to learn, and enabling the learners to adapt knowledge to different situations. Furthermore, today the ease provided by technologies in instructional activities, and gradual increase in the prevalence of technology have made it necessary and inevitable to integrate technology into education (Bozkurt & Cilavdaroglu, 2011). Besides, Oberlander and Johnson (2004) stated that for students to learn better, it is necessary to integrate instructional technologies into PBL so as to create conditions that facilitate meaningful learning, enhance visuality, and provide opportunities for knowledge transfer into authentic tasks, feedback, and revision. Moreover, findings from many studies (e.g. Cheung, 1988; Erktin, 1993) indicated that learners' attitudes towards mathematics is a significant predictor of achievement, thus developing favorable attitudes among learners towards mathematics have become one of the primary objectives of mathematics education. Therefore, it is considered important in terms of mathematics education to better understand the underlying factors responsible for the impact of using PBL method and technology use in mathematics education on learners' mathematics achievements and attitudes towards mathematics classes. Likewise, Sheehan and Nillas (2010) further stress that technology integration has many benefits on learners, however, there is still a need for further studies to figure out the impact of technology on learning from all aspects. Present study is believed to contribute to the relevant literature by shedding light on this need.

Literature Review

Recently, a considerable number of studies have been conducted about the impact of technology and PBL method on learners' academic achievement and attitudes. Among them, let's first examine those investigating the impact of Geogebra software activities. For example, Baltacı, Yıldız and Kosa (2015) found that Geogebra can help learners develop positive attitudes towards mathematics, enhance their interests in mathematics, enjoy the lessons, and increase their motivation based on data obtained

from teachers' views. Likewise, Zengin (2017) concluded that Geogebra software contributed to the development of learners' mathematical communication skills through creating a learning environment with cooperation and discussions. In addition to this, Wah (2015) found positive and significant impact of Geogebra software on learners' academic achievement and motivation. Likewise, Takaci, Stankov and Milanovic (2015) inferred from interviews with students that using Geogebra software made it easy for the students to learn the topic, enabling them to communicate, cooperate and reinforce their understanding thanks to quick feedback. Moreover, Dayı (2015) argued that Geogebra software helped learners conceptually understand the definition of function better, they could easily model the real life problems with the help of Geogebra software removing any need to make symbolic definitions. As it can be understood from these research findings, using Geogebra software in mathematics education have favorable effects on learners.

Now we need to outline the findings of the studies on the effects of PBL. For example, Amalia, Surya and Syahputra (2017) reported that PBL has a positive and significant impact on learners' ability to solve mathematical problems. However, Han, et. al. (2016) suggested that while PBL has no significant impact on academic achievement, it ensured the learners to enjoy the lessons more, relieve their mathematics anxiety, and develop interest in advanced mathematics. Botty, Jaidin, Li, Shahrill and Chong (2016) found out that problem-based learning had a significant positive impact on learning mathematics, motivating learners to work in cooperation; and thus, reducing their dependence on teachers. It is understood from these studies that though different results regarding PBL's impact on student achievement have been obtained, positive effects of PBL on learners have been observed.

When we examine the studies conducted, it is evident that although there are many studies investigating the impacts of PBL and technology on students' achievement and attitudes based on students' self-report, research inquiring into the underlying reasons for their impacts on academic achievement and attitude have not been encountered. It is believed that this study will serve to better understand the impact of technology and PBL on attitude and achievement exploring the factors affecting achievement and attitude based on student views, and to develop multifaceted and content-rich instructional practices considering these factors.

Purpose

The purpose of the present study is to investigate in depth the effects of Technology Enhanced Problem-based Learning Activities (TEPLA), developed by the first author and implemented on 9th graders attending a state high school in Ankara, on learners' attitudes towards and achievement in mathematics through learners' views, and to designate the underlying reasons for these effects.

In this regard, this study will answer the following research questions:

1. How do the written and oral interviews with participants in the study group explain the effect of applied TEPLA on learners' attitudes towards mathematics?

2. How do the written and oral interviews with participants in the study group explain the effect of applied TEPLA on learners' achievement in mathematical functions topic?

Method

Research Design

Designed as a case study, which analyzes a certain process or individuals in depth (Creswell, 2014, s. 14), present study focused on and limited to the qualitative part of a large scale PhD dissertation (First Author, 2017). During the first phase of the abovementioned dissertation, using Technology Enhanced Problem-Based Learning Activities, (TEPLA) an experimental study was conducted on the treatment group while the control group was taught through Traditional Instruction (TI). This qualitative case study was conducted only on students in the treatment group. In order to analyze the impact of TEPLA on learners' attitudes and achievement regarding mathematics, qualitative data were collected using *Student Interview Form,* which included both open-ended written and oral interview questions.

It would be possible to reveal many factors affecting attitude and achievement we investigated in *the case study, which* is a qualitative research design providing the opportunity to investigate in depth the event. As a matter of fact, Yıldırım (1999) highlights that the field of education have been investigated predominantly with quantitative study, however, such studies fall short of explaining educational phenomena, and findings of these studies fail to regulate the actual practices in the field of education. Thus, we preferred to use the case study design to investigate in depth a phenomenon or an event which the researcher cannot take control of, by relying on the questions of "how" and "why" as asserted by Yıldırım and Simsek (2011).

Research Sample

Participants of the research comprised of 9th graders (25 students) attending a state high school affiliated to Provincial Directorate of National Education in Ankara during 2013-2014 school year. The school where the first author works, admits students through a selection exam and offers education on equally-weighted domain (combination of verbal and numerical skills). The group of the students, to whom the first author taught Mathematics, were selected as participants according to non-probability sampling method of convenient sampling.

Research Instrument and Procedure

In order to collect the detailed views of study group students about TEPLA used in this study, "Student Interview Form", including seven questions, was developed and used by the researcher. This form was examined by three academicians, and necessary regulations were executed.

The data were collected by using this Student Interview Form was developed by the first author in line with the views of an expert panel in order to understand the

students' detailed views about TEPLA. Using this form, students were asked to write down their views first. After the written feedbacks were analyzed, a total of twelve participants were selected from the students using the criterion sampling method based on the criteria of ability of self-expression, equal representation of both sexes, and different levels of mathematics achievement. Then, these students were interviewed in depth about their views on TEPLA. The interviews were recorded via a video camera.

First author taught mathematical functions to the participants using TEPLA. A total of twelve TEPLA were designed to help students understand the role and significance of functions in mathematics in real life through a combination of problem solving method and technology. These activities in TEPLA were aimed to ensure that students learned mathematics meaningfully by doing and experiencing with the help of technology. During these activities, learners were given examples of real life situations, and these stituations were transferred into problem situations as required by the subject matter. During the learning process, learners were expected to come to some mathematical conclusions through small and large group discussions. Participants were taught in three different formats. The first one included in-class group activities, the second one involved using smartboard and dynamic mathematics software in computer laboratory, and the third one was based on out-of-class activities and projects.

Validity and Reliability

To secure the validity of the qualitative data, following validity strategies were used as defined by Creswell (2014). First of these strategies is Peer Debriefing. Accordingly, the analysis of the participant students' views was examined by three specialized scholars, and revised duly. Moreover, these data were analyzed and approved by two peers doing their PhD in Mathematics education.

Furthermore, participant views running contrary to the general results of the research were also presented; thus, negative or discrepant information in the themes with opposite evidence were introduced. The validity of the results increases to the extent that the researcher has experiences and interactions with the participants. As a validity strategy, data were collected spending a prolonged time, a month (20 lesson hours), in the field.

Another validity strategy is to explain the researcher's bias he/she brings to the study in order to clarify the researcher's characteristics in the background in qualitative research. As a matter of fact, qualitative research is mainly characterized by reflectiveness. Therefore, all aspects of the researcher that could affect the study were defined in the "Researcher Role".

Researcher Role: All TEPLA were applied by the first author. The relevant researcher had studied the use of technology in teaching mathematics in his master's thesis, which facilitated him to teach participants the technology-based applications better. Moreover, the participant interviews were conducted by the first author, in

person. To avoid biased responses from the participants, they were encouraged to be relaxed at the beginning of the interviews, and the interviews were recorded on video.

Data Analysis

Participants' written and oral responses to the questions of Student Interview Form were coded via open coding method, and analyzed and evaluated. Participants' written and oral responses were labeled with letter "d" and "D", respectively. The analysis of the participant students' views was examined by three specialized scholars, and revised duly. Moreover, these data were analyzed and approved by two peers doing their PhD in Mathematics education.

Results

Results Regarding the First Research Question

The results about the first research problem "How do the written and oral interviews with participants in study group explain the effect of applied TEPLA on learners' attitudes towards mathematics?" were presented below. Participants' views on the underlying reasons for the impact of TEPLA on learners' attitudes towards mathematics were analyzed in terms of "Affection" and "Usefulness". Heijden (2003) defines Affection factor as one's enjoying something in its essence. Gürol (2008), on the other hand, defines Usefulness as the degree of one's belief that using a certain system will improve his/her work performance.

Affection factor: Five underlying reasons under "Affection" factor which explain the impact of TEPLA on learners' attitudes towards mathematics were listed in *Themes* column accompanied by quotations from participants in Table 1.

The first theme in Table 1, which emerged from the students' views suggested that group works made the lesson enjoyable. Data extracts from participants including D9 ("It was enjoyable that everyone tried to do something and cooperated with their classmates" and D15 ("When we do by giving hands to each other and cooperate, it becomes more intimate.") also stressed that group work made the lessons enjoyable. The second theme explaining the impact of TEPLA in terms of affection factor suggested that TEPLA facilitated learning the lessons. For example, participant D14 illustrated how they do so by saying "Since they are related with daily life, it facilitated my comprehension." The third factor was that TEPLA improved self-confidence, responsibility, communication competences, and alike. Asserting that "Out-of-class tasks enhanced our senses of responsibility." participant D16 emphasized that Out-of-class tasks enhanced their senses of responsibility, while participant D15 suggested how TEPLA strengthened their communications skills and self-confidence saying "It has strengthened our communication… I have gained confidence in this topic." Fourth factor suggested that learning how to learn among learners improves the fluency of lesson. Participant D16 indicated learning autonomously by saying "the lessons are more easygoing since we learn and produce formulas". The last factor suggested that technology made the lesson interesting. Participant D7 explained the role of technology

in making the lessons interesting by saying "More appealing the technology is (to the youth), more interesting the lessons are. "

Table 1

Participant Views on Underlying Reasons Regarding the Impact of TEPLA under Affection Factor

Themes	Quotations
Group work makes the lesson enjoyable	D9- *"It ensured that the lessons were not dull anymore and lessons were productive. It was enjoyable that everyone tried to do something and cooperated with their classmates. I think other lessons should be taught in this way. It makes mathematics classes more enjoyable and interesting."* D15: *When we do by giving hands to each other and cooperate, it becomes more intimate.*
TEPLA make learning easier	D10 -*"I have been afraid of functions since last year. I realized that I don't need to be afraid of them when it is taught this way. Functions are not that difficult to understand."* D14: *Since they are related with daily life, it facilitated my comprehension. I have learned what function means during the activities. I know its relation with the daily life. I can solve the problems more easily as I can understand them by reasoning.*
TEPLA improve self-confidence, responsibility, communication competences and alike	D16-*"Out-of-class tasks enhanced our senses of responsibility. Communication competence has improved. As we competed with our friends like in a race, I gained self-confidence, saying 'I know it, I must do it first'."* D15: *It has strengthened our communication. We become closer friends. I have gained confidence in this topic.*
Learning how to learn improves the fluency of lesson	D16 -*"While we used to yearn for the end of the lesson since the lessons were very boring and monotonous, now the lessons are more easygoing since we learn and produce formulas."* d2: *"Involving the daily life while teaching the lesson content pleased me since it enabled me to apply mathematics into life".*
Technology makes the lesson interesting	d23 - *"Technological applications contributed to us. They attracted my attention more. I listened to the teacher more and engaged to the lesson more."* d7 - *"More appealing the technology is (to the youth), The more interesting the lessons are"*

Usefulness factor: The three reasons under "*Usefulness*" *factor* which explained the impact of TEPLA on learners' attitudes towards mathematics were listed under *Themes* column accompanied with quotations from participants in Table 2:

Table 2

Participant Views on Underlying Reasons Regarding the Impact of TEPLA Under Usefulness Factor

Themes	Quotations
Increasing eagerness to learn	D16 -*"I used to find it boring. Students always ask a question: "What does Mathematics even work for us? However, when functions, which students generally find very complicated, were applied into daily tasks, we admitted that it was so connected to real life. Mathematics is something reasonable. It is necessary in daily life. Thus, I became more devoted to study. At least we didn't get bored. We used to get bored previously. We could not even understand anything from the test books."* D13 – *I realized that mathematics can be enjoyable. We have learned that it may work in daily life, seeing that I can use it in daily life. I became more willing to learn.*
Developing affection for mathematics	D20 -*"Thus, One likes mathematics much more from now on. (It becomes like a game. As we solve our own problems by working regularly on it, we also solve mathematics by working on it, too. Thus, we kind of identify mathematics with our life and consider it as something we must do."* D1 - *"It presented us a different point of view. It made us love mathematics. People who dislike mathematics generally do not believe in its usefulness. Actually, they are right. The practical use of mathematics is not taught in traditional education ."*
Satisfaction by students	D10 -*"When we measured the height of the school building, we both enjoyed a lot and applied mathematics directly to our lives. We didn't pretend, we really witnessed that mathematics is feasible, and we entertained while doing."* D9- *"It ensured that the lessons were not dull anymore and lessons were more productive. It was enjoyable that everyone tried to do something and cooperated with their classmates."*

It was seen in Table 2 that first factor explaining the impact of TEPLA on learners' attitudes towards mathematics in terms of usefulness was increasing learners' eagerness to learn. Regarding this factor, participant *D13* explained how TEPLA increased learners' eagerness to learn by saying *"I realized that mathematics can be enjoyable. We have seen that it can be used in daily life, and we have learned that it may work in daily life. I beca*me more willing to learn." Second factor suggested that TEPLA developed affection for mathematics. Participant D20 asserted that TEPLA helped him love mathematics by identifying himself with it by saying *"Thus, we associate mathematics with our life, and consider it as something we must do."* Third factor defined the satisfaction TEPLA gave to the learners. Learners expressed their satisfaction by saying "*It was productive, it was entertaining*".

Results Regarding the Second Research Question

The results about the second research problem "*How do the written and oral interviews with participants in study group explain the effect of applied TEPLA on learners' achievement in mathematical functions…?*" were presented below. Ten underlying reasons explaining the impact of TEPLA on learners' mathematics achievement were listed under *Themes* column accompanied with quotations from participants in Table 3.

Table 3

Participant Views on Underlying Reasons Regarding the Impact of TEPLA on Mathematics Achievement

Themes	Quotations
Group work	D15- *"When we do the group work, we cannot always ask you questions. But I can easily ask to my classmates. When I am solving the problems, I come to some point where I get stuck. I need someone to give me a hand and show me so that I can progress. I was able to progress myself in the following questions. I was perfect for us."* D1- *"There was a pleasant competition among us during group work. Everybody concentrated better on the questions. Hence, we could understand the topic better."*
Difficulties of group work in learning	D2-*"As I tried to involve my peers in the group, I missed the lesson."* D1- *"Although our classmates explained us the difficult parts, we cannot understand as they are not expert."*
Verbal expression	D14 - *"When it is stated, we can exactly understand how to do."* D10- *"Expressing what we have learned at the end of the activity has improved our mathematical expression ability."*
Visuality	D11- *"Because we could visualize the graphic as we see on the computer, it became easier to understand. I was able to solve the problem faster upon seeing it. I mean it made the topic more concrete. After seeing the graphic we were able to find the equation more easily."* D23- *"The program we used enabled me to do and understand the activity more easily as I could see the changes on the graphic."*
Lack of technology acceptance	D16-*"It takes some time to get used to the technology. There was a problem with perception. I could not focus on the computer. Paper-and-pencil is better for me. There was some humming noise in the laboratory."* D25- *"I am not quite good at technology and computers. I believe we can learn the content without using the technology."*
Learning by discovering, experiencing and involving more senses via technology	D14 - *"It is more effective to learn myself with daily life activities, because in this way we open our senses and challenge ourselves more thus gaining more permanent knowledge."* D20- *"When I do it myself, I struggle, pay attention and spend effort. This is easier to remember."*
Daily life tasks' making it easy to comprehend and associate	D20- *"In the new method we learned technological authentic tasks and this attached us to mathematics. Using mathematics in daily life and applying them by seeing… both makes us happy and let us understand and comprehend things better."* D8- *"I think using daily life tasks is a better method, because we can understand better when we are in it."*
Meaningfulness and increased speed of learning by discovering and experiencing	D16 – *"We have learned on our own by doing. We learned faster. We didn't learn stereotyped things, but we discovered the solutions by finding our own stereotypes."*
The opportunity to spend the time to improve higher order thinking skills thanks to the ease of processing via technology	D14 -*"I think our graphic exercises in the laboratory improved our ability to read graphics. It was beneficial because we did everything with technology by seeing in daily life activities. It was both easy and we saved extra time. We didn't waste any time with calculations. The essential knowledge was produced. We could understand what actually we needed to learn. This let us understand the topic better. We understand better by concentrating on doing rather than wasting time with calculations."*

Table 3 Continue

Themes	Quotations
Self-Learning	D21-"*In fact, one hardly forgets something he/she has learned by himself/herself. One remembers it in his/her own way. It would be a great desire to discover something by ourselves. We can remember better later on, since we have learned it by ourselves. It would be permanent knowledge.*" D20-"*We can understand, comprehend and remember something we do, learn and apply by ourselves.*"

First factor explaining the TEPLA's relationship with learners' mathematics achievement was "*Group work*". Participants suggested that one aspect of the contribution of group work was cooperation expressing their views with such sentences as "*…we cannot always ask you questions. But I can easily ask to my classmates.*" Participants defined another contribution of group work as encouraging competition by saying "*There was a pleasant competition among us during group work.*" Second factor was "*Difficulties of group work on learning*". Participants stressed that they had difficulty in following the lesson due to the group work, and cooperation within group was not enough at higher levels. Third factor was "*Verbal expression*". Participants stated that TEPLA improved learners' verbal expression skills. Fourth factor emerged from participant views was "*Visuality*". Learners asserted that they were able to comprehend the content more easily and faster. The fifth factor explaining the TEPLA and mathematics achievement was "*Failure to accept technology*". Participants stated that they had difficulty in getting used to the computers and had problems in laboratory, since they were not quite good at computers. The sixth factor emerged from learners' views was "*Learning by discovering, experiencing and involving more senses via technology.*" Participants explained that learning by exploring and doing enhanced their capacity to perceive, increased their interests, thus helped to achieve retention in learning outcomes. Seventh factor was "*Daily life tasks' making it easy to comprehend and associate*". Participants emphasized that daily life activities involved them in actual learning and helped them comprehend the topic better by understanding its rationale. Eighth factor that emerged from participant views was "*meaningfulness and increased speed of learning by discovering and experiencing*". Participants suggested that learning by discovering, doing and making sense was faster. Ninth factor was about "*The opportunity to spend the time to improve higher order thinking skills thanks to the ease of processing via technology*". Participants' views revealed that technology accelerated the unnecessary processes and saved time, thus provided the opportunity to improve the higher order thinking skills. The tenth and the last factor emerged was "*Self-Learning.*" Participants emphasized that by learning themselves, they comprehended better and retained what they had learned. All factors explaining the underlying reasons for the impact of TEPLA on learners' attitudes towards mathematics and academic achievement are summarized in Table 4.

Table 4

All Factors Explaining the Underlying Reasons for the Impact of TEPLA *on Learner Attitudes and Academic Achievement*

Factors explaining the relationship of learner attitudes and TEPLA		Factors explaining the relationship of academic achievement and TEPLA
Themes explaining the affection factor	**Themes explaining the usefulness factor**	
Group work makes the lesson enjoyable	*Increasing eagerness to learn*	*Group work*
TEPLA make learning easier	*Developing affection to mathematics*	*Limitations of group work on learning*
TEPLA improve self-confidence, responsibility, communication competences and alike	*Satisfaction by students*	*Verbal expression*
Self-learn improves the fluency of lesson		*Visuality*
Technology makes the lesson interesting		*Failure to accept technology*
		Learning by discovering, experiencing and involving more senses via technology
		Daily life tasks' making it easy to comprehend and associate
		Meaningfulness and increased speed of learning by discovering and experiencing
		The opportunity to spend the time to improve higher order thinking skills thanks to the ease of calculation via technology

Discussion, Conclusion and Recommendations

The results of the present research are supported by many previous studies in the literature, which investigated using problem-based learning, technology, and dynamic mathematics software in particular. For example, investigating a software prototype to teach mathematical functions, Ceylan (2003) found based on interviews with learners that almost all learner agreed that they could learn mathematics more meaningfully and permanently thanks to the technology-aided instructional methods. Likewise, Wah (2015) reported that Geogebra software program had positive and significant effect on learners' achievement and motivation. Moreover, findings from Takaci et al. (2015) also support the results of the present study. As a result of the analysis of the participant interviews in this study, it was concluded that Geogebra applications enabled the learners to understand the topic better, to devote more time to analyze the relationship between formulas and graphics rather than the calculations,

and to enjoy an effective and enriched learning environment. Furthermore, learners stated that they had the opportunity to communicate, cooperate and get quick feedback, thus deepening their knowledge thanks to Geogebra. Sheehan and Nillas (2010) found out that learners could participate in lessons more actively, experience higher order learning, and associate different mathematical representations in case they were primary users of technology. Fies (2007) argued that technology helps many students visualize topics in mathematics, which help them learn more easily, and that using technology enhances learners' motivation and participation into lessons. In addition to this, Arbain and Shukor (2014) reported that Geogebra applications had significant and positive effects on learners, and enhanced learners' motivation, self-confidence, willingness to learn, and critical thinking skills.

The findings of many previous researches about PBL method also comply with the findings of the present study. However, the results regarding the effect of PBL on academic achievement are different to some extent. For example, Usta (2013) found out that problem-based learning method had positive effects on learners in terms of improving higher order thinking skills like problem solving, association, and deductive reasoning, and other outcomes like communication skills and self-efficacy. On the other hand, Han et al. (2016) reported that while problem-based learning method had no significant effect on the academic achievement of 9th grade students, thanks to PBL method learners found the lessons more enjoyable, their level of mathematics anxiety decreased, and their interests in advanced mathematics increased. Botty et al. (2016) reported that problem-based learning had significant positive effect on learning mathematics, motivated them to work in cooperative groups, thus decreased their dependence on teachers. Moreover, Amalia et al. (2017) found out that problem based learning had positive and significant effect on developing learners' ability to solve math problems.

An analysis of the research results above apparently reveals that there is a consensus about the positive effects of technology on academic achievement, while there are controversial results regarding the effect of problem-based learning on academic achievement. Therefore, it can be interpreted that integrating technology and problem-based learning into our TEPLA can have a positive and significant impact on academic achievement. As a matter of fact, as specified in Gestalt theory, the whole can have an effect different from the sum of its parts (Cevik, 1991). In addition to this, apart from the previous researches in the literature, results of the present study regarding TEPLA's effect might contribute to understanding the effect of TEPLA on attitude and academic achievement better, and enable future researchers and practitioners to prepare multifaceted and successful instructional activities considering these factors. However, it should be noted that the present study was limited to the 9th grade students attending a high school in Ankara and to teaching mathematical functions only.

This study investigated how students have explained the impact of TEPLA (Technology Enhanced Problem-Based Learning Activities) on learners' attitudes towards mathematics and academic achievement in mathematics through written and oral interviews.

The analysis revealed two main factors accounting for the impact of using TEPLA on learners' attitudes towards mathematics: *Affection* and *Usefulness,* which comprised five and three distinct reasons, respectively. Another aspect of the study was about the impact of TEPLA on learners' achievement in mathematics. As a result of the analysis of the participants' views, ten factors, mentioned above, were defined accounting for the impact of using TEPLA on learners' mathematics achievement.

It was found in the present study that learners found TEPLA interesting, meaningful, and entertaining. Therefore, considering that learning motivation refers to the extent learners find learning activities meaningful, interesting, and helpful in making use of the learning outcomes, whether TEPLA affect learner motivation can be investigated using a motivation scale.

Contrary to the positive results, it was also found that one of the factors explaining the relationship between TEPLA and learners' academic achievement in mathematical functions was the failure to accept technology and adapt to the practices since they were not good at technology. Therefore, learners' readiness regarding technology and their attitudes towards technology can be measured before applying TEPLA, and thus the effect of technology on learners can be investigated comparing their backgrounds.

Since present study was restricted to mathematical functions, activities like TEPLA can be used to teach different topics in mathematics at different class levels to investigate the factors affecting attitude and academic achievement in a more comprehensive manner. Besides, in the present study it was found that TEPLA were accepted and considered useful by the learners. Therefore, it can be useful to teachers, and candidate teachers can be trained about TEPLA to use them in their actual or future schools. As a matter of fact, following the present research, 15.000 mathematics teachers were trained online about dynamic mathematics software by a team of experts where the first author was also involved within a project conducted by Ministry of Education Directorate General of Innovation and Educational Technologies (YEGİTEK). Moreover, 100 mathematics teachers were given a face-to-face teacher training program about teaching mathematics with technology. In addition to this, learners stated that they could retain what they have learned by exploring, spending effort, and using what they had learned in daily life. Thus, whether TEPLA ensure retention in learning can be investigated using quantitative research methods.

References

Amalia, E., Surya, E., & Syahputra, E. (2017). The effectiveness of using problem based learning in mathematics problem solving ability for junior high school students. *International Journal of Advance Research and Innovative Ideas in Education, 3*(2), 147-155.

Arbain, N., & Shukor, N. (2015). The effects of geogebra on students' achievement, *Procedia Social and Behavioral Sciences, 172,* 208–214.

Baltaci, S., Yildiz, A., & Kosa T. (2015). Analitik geometri öğretiminde Geogebra yazılımının potansiyeli: Ogretmen adaylarının gorusleri (Potential of Geogebra software in teaching analytical geometry: Views of teacher pre-service teachers). *Turkish Journal of Computer and Mathematics Education, 6*(3), 483-505.

Botty, R.H., Jaidin, J.H., Li, H. , Shahrill, M., & Chong, S.F. (2016). The implementation of problem-based learning in a year 9 mathematics classroom: a study in Bruneidarussalam. *International Researchs in Education.* 4(2), s. 34-47.

Bozkurt, A., & Cilavdaroglu, A. (2011). Matematik ve sınıf ogretmenlerinin teknolojiyi kullanma ve derslerine teknolojiyi entegre etme algıları (Using mathematics and primary school teachers to use technology and integrate technology into their lessons). *Kastamonu Journal of Education, 19*(3), 859-870

Cakir, S., & Aztekin S. (2016). Matematik dersinde probleme dayalı Öğrenme yonteminin ogrencilerin motivasyonlarına ve matematik kaygı duzeylerine etkisi (The effect of Problem-based learning method in mathematics on students' motivation and mathematics anxiety levels). *Abant Izzet Baysal University Journal of the Faculty of Education, 16*(2), 377-398.

Cevik, S.(1991). *Mekan-kimlik-kimliklendirme, Trabzon sokagı ornegi (Place-identity-identification, Trabzon Street example)* (Doctorate Thesis). K.T.Ü. Institute of Science, Trabzon.

Ceylan, E. A. (2003). *Matematik egitimine uygun bir ogretim yazılımı ve prototipi gelistirilmesi, calısma yaprakları ile uygulanması (Development of a teaching software and prototype suitable for mathematics education, application with worksheets)* (Doctorate Thesis). Dokuz Eylül University Institute of Educational Sciences, İzmir.

Cheung, K.C. (1988). Outcomes of schooling: Mathematics achievement and attitudes Towards mathematics learning in Hong Kong. *Educational Studies in Mathematics, 19,* 209-219.

Creswell, J.W. (2014). *Araştırma deseni (Research design).* (Translator: Demir, S.B.). Ankara: Egiten Kitap.

Dayı, G. (2015). *How geogebra contributes to middle grade algebra Students' conceptual understanding of functions* (Doctorate Thesis). Florida State University, Florida.

Erktin, E.(1993). *The Relationship between math anxiety attitude toward mathematics and classroom environment.* In 14. International Conference of Stress and Anxiety Research Society (STAR), Cairo, Egypt.

Fies, C. (2007). Research summary: Digital technologies in mathematics and science education. Retrieved May 13, 2017, from http://ww.nmsa.org/ResearchSummaries/DigitalTechnology/tabid/1486/Default.aspx

Han, S. H. , Shahrill, M.,Tan, A., Tengah,K.,A. Jaidin, J.,H., & Jawawi, R. (2016). *Administering Problem-Based Learning (PBL) Approach in the Teaching of College Level Mathematics.* Presentation in International Conference On New Horizons in New Education. Vienna.

National Council of Teachers of Mathematics Commission on Standards for School Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston VA: The Council.

Oberlander, J., & Talbert-Johnson, C. (2004). Using technology to support problem-based learning, *Action in Teacher Education, 25*(4), 48-57.

Sheehan, M., &Nillas,L. A. (2010) "Technology integration in secondary Mathematics classrooms: Effect on students' understanding" *Journal of Technology Integration in the Classroom*, *2*(3),67-83.

Takaci, D. , Stankov, G., & Milanovic, I. (2015). Efficiency of learning environment using geogebra when calculus contents are learned in collaborative groups. *Computers and Education, 82*, 421–431.

Usta, N. (2013). *Probleme dayalı ögrenmenin ortaokul ogrencilerinin matematik basarısına, matematik ozyeterligine ve problem cozme becerilerine etkisi (The effect of problem based learning on mathematics achievement, mathematics self-efficacy and problem solving skills of middle school students)* (Doctorate Thesis), Gazi University Institute of Educational Sciences, Ankara.

Wah, K. L. (2015). The effects of instruction using the arcs model and geogebra on Upper secondary students' motivation and achievement in learning combined transformation. *Asia Pacific Journal of Educators and Education, 30*, 141–158.

Yildirim, A., & Simsek, H. (2011). *Sosyal bilimlerde nitel araştırma yontemleri (Qualitative research methods in social sciences)* (8. Edition). Ankara: Seckin Publishing.

Yildirim, A. (1999). Nitel arastırma yontemlerinin temel ozellikleri ve egitim arastırmalarındaki yeri ve onemi (Basic features of qualitative research methods and place and importance in education research). *Education and Science*, *23*(112), 7-17.

Zengin, Y. (2017). The potential of geogebra software for providing mathematical communication in the light of pre-service teachers' views, *Necatibey Faculty of Education Electronic Science and Mathematics Education Journal. 11*(1), 101-127.

Teknolojiyle Zenginleştirilmiş Probleme Dayalı Öğrenme Uygulamalarının Öğrencilerin Tutum ve Başarılarına Etki Sebeplerinin İncelenmesi

Atıf:

Özet

Problem Durumu: Anlamlı öğrenmeye elverişli koşullar sağlayan, görsellik, geri bildirim ve gözden geçirmeye imkân sunan öğretim teknolojisi ortamlarının öğrencilerin yeteneklerini geliştirmek üzere tasarlanmasını sağlayan teknolojinin matematiği gerçek yaşamla ilişkilendiren PDÖ'ye entegrasyonu öğrencilerin daha iyi öğrenmeleri açısından önemlidir. Fakat teknoloji ve PDÖ'nün öğrenciler üzerindeki etki veya faydalarının tüm yönleriyle derinlemesine inceleyen çalışma pek bulunmamaktadır. Dolayısıyla, matematik öğretiminde PDÖ'nün ve teknoloji kullanımının matematik başarısına ve öğrencilerin matematik dersine yönelik tutumlarına etkisinin altında yatan sebeplerin açığa çıkarılması matematik öğretimi açısından yararlı olacaktır.

Araştırmanın Amacı: Bu çalışmanın amacı Teknoloji Destekli Probleme Dayalı Öğrenme Uygulamalarının (TPU)'nun öğrencilerin matematiğe yönelik tutumları ve matematik başarıları üzerindeki etkisinin altında yatan sebepleri belirlemektir.

Araştırmanın Yöntemi: Büyük ölçekli bir doktora araştırmasının (Birinci Yazar, 2017) sadece nitel kısmına odaklanan bir durum çalışması olarak gerçekleştirilen bu araştırmada Öğrenci Görüşme Formu kullanılarak öğrencilerle gerçekleştirilen yazılı ve sözlü mülakatların analizleri yapılmıştır. Araştırmada çalışma grubu; Milli Eğitim Bakanlığına bağlı olarak birinci yazarın görev yaptığı, sınavla öğrenci alan ve eşit ağırlık alanında eğitim veren bir devlet lisesinde bulunan bir 9. Sınıfın öğrencileri seçilmiştir.

Fonksiyonlar konusunun öğretiminde teknoloji destekli probleme dayalı öğrenme uygulamaları (TPU) adı altında problem çözme ve teknolojiyi kullanarak matematiğin günlük hayattaki yerini ve anlamını görebilecekleri bir içerikte birinci yazar tarafından hazırlanan 12 etkinlik, çalışma grubu öğrencilerine birinci yazar rehberliğinde uygulanmıştır.

Çalışma grubu öğrencilerinin teknoloji destekli probleme dayalı öğrenme uygulamaları hakkında düşüncelerini ayrıntılı biçimde algılamak için birinci yazar tarafından uzman görüşleri doğrultusunda bir Öğrenci Görüşme Formu düzenlenerek 7 soru içerecek biçimde son şekli verilmiştir. Birinci yazar tarafından geliştirilen bu

form kullanılarak önce öğrencilerin yazılı görüşleri alınmıştır. Daha sonra bu yazılı görüşler incelenerek kendini ifade edebilen, kız-erkek ve matematik başarısı dengeleri gözetilerek bu öğrencilerden amaçlı örneklemeden ölçüt örnekleme yöntemi ile belirlenen 12 öğrenci ile okulda özel bir odada mülakat yolu ile TPU hakkında detaylı görüşleri alınmıştır.

*Araştırmanın Bulguları***:** Öğrencilerin matematik dersine yönelik tutumları ile TPU'nun ilişkisi "Hoşlanma" ve "Kullanışlılık" alt boyutları bazında incelenmiştir. "Hoşlanma" boyutu ile TPU ilişkisini açıklayan sebepler aşağıda sıralanıp ilgili öğrenci görüşleri verilmiştir. TPU'nun *hoşlanma* boyutu ile ilişkisini açıklayan nedenler *Grup Çalışması, TPU'nun ders işleyişini kolaylaştırması, TPU'nun kişisel becerileri geliştirmesi, kendi kendine öğrenmenin dersi akıcı kılması ve teknolojinin dersi ilgi çekici hâle dönüştürmesi* olarak ortaya çıkmıştır. "Kullanışlılık" *boyutunun* uygulanan TPU ile ilişkisini açıklayan etmenler *dersi öğrenme isteğini artırması, matematiği sevdirmesi ve matematiğin kullanışlılığını görmenin öğrencilere verdiği memnuniyet* olarak ortaya çıkmıştır.

TPU ile matematik dersi akademik başarısı ilişkisini açıklayan on etmen aşağıdaki başlıklar altındaki öğrenci görüşleriyle belirlenmiştir. TPU ile matematik dersi başarısı ilişkisini açıklayan etmenler; *grup çalışması, grup çalışmasının öğrenme zorlukları, sözlü ifade etme, görsellik, teknolojinin kabullenilmemesi, teknolojiyle daha çok duyu organlarının kullanılması, günlük hayat uygulamalarının ilişki kurmada ve anlamada kolaylığa sebep olması, Keşfederek, uygulayarak öğrenmenin anlamlı ve hızlı olması, Teknolojinin sağladığı işleyiş kolaylığının zamanı üst seviyedeki düşünme becerilerinin geliştirilmesine kullanmaya imkân tanıması ve Kendi kendine öğrenme* olarak belirlenmiştir.

Araştırmanın Sonuçları ve Önerileri: Öncelikle çalışma grubunun öğrencileriyle yapılan yazılı ve yüz yüze görüşmeler, uygulanan TPU'nun öğrencilerin matematik dersine yönelik tutumlarına etkisini nasıl açıkladığı incelenmiştir. Öğrencilerin matematik dersine yönelik tutumları ile TPU'nun ilişkisinin incelendiği iki alt boyuttan biri olan *Hoşlanma* boyutu ile TPU ilişkisini açıklayan sebepler yukarıda belirtilen beş başlık altında, diğer alt boyut olan *Kullanışlılık* boyutunun uygulanan TPU ile ilişkisini açıklayan etmenler ise 3 başlık altında incelenmiştir. Çalışmada öğrenci görüşlerinin incelendiği diğer bir boyut; TPU ile öğrencilerin matematik başarıları arasındaki ilişkidir. İncelenen öğrenci görüşleri sonucunda TPU ile matematik dersi başarısı ilişkisini açıklayan etmenler on başlık altında incelenmiştir.

Bu çalışmada öğrencilerin TPU'yu ilgi çekici, anlamlı ve eğlenceli buldukları belirlenmişti. Dolayısıyla, öğrenme motivasyonu, öğrenen bireyin, öğrenme etkinliklerini anlamlı, ilgi çekici bulması ve bunlardan yarar sağlaması olarak düşünüldüğünde TPU'nun öğrenci motivasyonunu etkileyip etkilemediği bir motivasyon ölçeği kullanılarak nicel olarak da araştırılabilir.

Bununla birlikte, öğrencilerin matematik dersi fonksiyon konusundaki başarıları ile TPU ilişkisini açıklayan bir etmenin teknolojinin kabullenilmemesi olduğu ve öğrencilerin teknolojiye adapte olmakta güçlükler yaşadıklarından uygulamaları benimseyemedikleri belirlenmişti. Bu yüzden TPU çalışması öncesinde öğrencilerin bilgisayar hazır bulunuşluk düzeyleri veya teknolojiye yönelik tutumları belirlenip

teknolojinin öğrenciler üzerindeki etkisi teknoloji geçmişleriyle karşılaştırılarak irdelenebilir.

Bu çalışma sadece fonksiyon konusuyla sınırlı olduğundan farklı matematik konularında ve farklı seviyelerdeki sınıflarda TPU benzeri uygulamalar kullanılarak tutum ve başarıya etki eden faktörler daha geniş bir yelpazede irdelenebilir. Bunun yanı sıra, çalışmamızda TPU'nun öğrenciler tarafından benimsendiği ve faydalı görüldüğü belirlenmişti. Bu yüzden TPU öğretmenlere veya öğretmen adaylarına ders verilip bu uygulamaları görev yaptıkları okullarda uygulamaları sağlanması öğrenciler açısından faydalı olabilir.

Öğrenciler akılda kalıcılığa ilişkin görüşlerinde kendilerinin keşfederek, emek sarf ederek, günlük hayatta uygulayarak ve kullanarak kalıcı öğrenme gerçekleştirdiklerini ifade etmişlerdir. TPU'nun akılda kalıcılık sağladığı görüşü nicel olarak da araştırılabilir.

Anahtar Kelimeler: Matematik eğitimi, Geogebra, Günlük hayat uygulamaları, Durum çalışması.

Submission Checklist

Makale Sunumu Kontrol Çizelgesi

Indicate that this submission is ready to be considered by this journal by checking off the following.

Aday makalenin değerlendirilmeye hazır olduğunu aşağıdakilerin her biri ile karşılaştırarak kontrol ediniz.

Manuscript Agreement:

I/We hereby accept that, the article after being accepted for publication in the Eurasian Journal of Educational Research (EJER), the author (s) as, all rights related to the article has been transferred to the ANI PUBLISHING under the laws of the, "copyright transfer".

Makale Eurasian Journal of Educational Research (EJER) dergisinde basıma kabul edildikten sonra, yazar(lar) olarak; makale ile ilgili tüm hakları, "Telif Hakkı Devir" yasaları uyarınca, ANI YAYINCILIK'A devrettiğimizi kabul ediyoruz.

Manuscript Submission:

I have prepared my manuscript based on the criteria listed below and I accept all submission conditions.

Makalemi aşağıda belirtilen kriterlere göre hazırladım ve makale gönderme koşullarının tamamını kabul ediyorum.

Indicate that this submission is ready to be considered by this journal by checking off the following.

Aday makalenin değerlendirilmeye hazır olduğunu aşağıdakilerin her biri ile karşılaştırarak kontrol ediniz.

| 1 | ☒ | *The manuscript is a report of original educational research or a discussion article on research topics in education.* |

		Aday makale, eğitim alanıyla doğrudan ilgili bir araştırma ya da tartışma yazısıdır.
2	☒	*The submission has not been previously published, nor sent to another journal for consideration.* Sunulan çalışma daha önce herhangi bir yerde yayınlanmamıştır, başka bir derginin incelemesinde değildir.
3	☒	*Within a year, I have not submitted a manuscript to EJER as an author or co-author for review other than this manuscript.* Son bir yıl içerisinde, yazarı olduğum ya da yazarları arasında bulunduğum başka bir çalışma değerlendirilmek üzere EJER'e sunulmamıştır.
4	☒	*All authors of this manuscript are subscribers of the EJER and accept to pay 300 EURO for the cost of proofreading and 10 loose copies of their articles if the manuscript is accepted.* Aday makalenin kabul edilip basılması halinde makalelerinin İngilizce son okuma (proofreading) ve 10 serbest kopya gönderimi nedeniyle oluşan **1275 TL** maliyeti Anı Yayıncılık'a ödemeyi kabul ederler.
5	☒	*The entire manuscript is written in English.* Aday makalenin bütününün yazım dili İngilizcedir.
6	☒	*The original manuscript is typed on A4 paper. The margins are 2.5 cm.* Aday makale kenar boşlukları 2.5 cm olan A4 kağıda yazılmıştır.
7	☒	*Each paragraph is longer than two sentences.* Her bir paragraf en az üç cümle içermektedir.
8	☒	*The entire manuscript - including quotations, references, author note, content footnotes, figure captions, and all parts of tables – is double-spaced.* Aday makalenin tamamı, alıntılar, kaynakça, şekil ve tablo başlıkları da dâhil olmak üzere çift aralıklı yazılmıştır.
9	☒	*The submission file is in Microsoft Word document file format. 12-point Times New Roman font is used in entire manuscript.*

		Aday makale, tamamında 12 punto Times New Roman yazı tipi kullanılarak hazırlanmış bir Microsoft Word dokümanıdır.
10	⊠	*The text has had the authors' names removed. If an author is cited, "Author" and year are used in the bibliography and footnotes, instead of author's name, paper title, etc. The author's name has also been removed from the attached document.* **Aday makale, yazar adları çıkarılarak sunulmuştur**. Eğer yazar kendisine atıfta bulunduysa yazarın adına ve çalışma başlığına yer verilmeyecek, sadece "Author" yazılarak çalışmanın yılı belirtilecektir. Eklenen dosyada yazar adı belirtilmeyecektir.
11	⊠	*The title is 10 to 12 words.* Aday makalenin başlığı 10-12 sözcük uzunluğundadır.
12	⊠	*The maximum length of the manuscript-including structured abstract in English, tables, and references is 6000 words. This limitation does not include Turkish extended abstract (750-1000 words) which is placed after the references section.* Aday makale, İngilizce abstract, tablolar ve kaynakça vb. tüm ögeler dâhil olmak üzere en fazla 6000 sözcüktür. Kaynakça'nın ardından yer verilen uzun Türkçe özet (750-1000 sözcük) bu sayıya dâhil değildir.
13	⊠	*The article is preceded by* ***English Structured Abstract of not more than 250 words*** *and not less than 200 using five required headings:* **Purpose**: *State the problem in field. Then explain the purpose of the study.* **Method:** *Specify the research design, sample, and research instrument and data analysis in brief.* **Findings**: *Highlight the significant, interesting or surprising results.* **Implications for Research and Practice**. (These headings may need some adaptation in the case of discussion papers: *Background, Purpose of Study, Sources of Evidence, Main Argument, and Conclusions*). More information available from http://www.tandf.co.uk/journals/authors/rereabstracts.asp Yapılandırılmış İngilizce öz 200-250 sözcük uzunluğunda olup, aday makalenin başında yer almakta ve **Purpose** (*İlk önce alanda karşılaşılan sorunu belirtelim. Daha sonra araştırmanın amacını bir cümle ile veriniz*),

		Method *(Araştırma deseni, örneklem, very taoplama aracı ve verilerin analizinini kısaca açıklayınız),* **Findings** *(En önemli ve çarpıcı araştırma bulgularını verelim)* **Implications for Research and Practice,** *(Uygulama ve ileriye dönük araştırmalar için olası çıkarımlarınız)* başlıklarını içermektedir. Bu başlıklar tartışma yazıları için: *Çalışmanın Temeli, Çalışmanın Amacı, Kanıt Kaynakları, Ana Tartışma ve Sonuçlar* şeklinde olabilir. Daha fazla bilgi için; http://www.tandf.co.uk/journals/authors/rereabstracts.asp adresine başvurunuz.
14	☒	*Following the structured abstract in English four to six keywords are included. They should* ***represent*** *the content of your manuscript and* ***be specific*** *to your field or sub-field. Avoid using keywords form the title of the paper.* Yapılandırılmış İngilizce özden sonra 4-6 anahtar sözcüğe yer verilmiştir. Anahtar kelimeler çalışmanızı temsil etmeli ve kendi alanınıza ya da alt alanlara özgü olmalıdır. **Makale adındaki kavramları anahtar kelime olarak seçmekten <u>kaçınınız</u>**.
15	☒	*An extended (750-1000 words) Turkish structured abstract is placed following the "References" section using five required headings: Problem Statement, Purpose of Study, Methods, Findings and Results, and Conclusions and Recommendations.* (These headings may need some adaptation in the case of discussion papers: *Background, Purpose of Study, Sources of Evidence, Main Argument, and Conclusions*). More information available from http://www.tandf.co.uk/journals/authors/rereabstracts.asp Kaynakça'dan sonra 750-1000 sözcükten oluşan Türkçe yapılandırılmış öze yer verilmiştir. Türkçe yapılandırılmış öz *Problem Durumu, Araştırmanın Amacı, Araştırmanın Yöntemi, Araştırmanın Bulguları, Araştırmanın Sonuçları ve Önerileri* başlıklarını içermektedir. Bu başlıklar tartışma yazıları için: *Çalışmanın Temeli, Çalışmanın Amacı, Kanıt Kaynakları, Ana Tartışma ve Sonuçlar* şeklinde olabilir. Daha fazla bilgi için; <u>http://www.tandf.co.uk/journals/authors /rereabstracts.asp</u>
16	☒	*Following the Turkish structured abstract, four to six keywords are included.*

		Uzun Türkçe özetten sonra 4-6 anahtar sözcüğe yer verilmelidir.
17	⊠	*References are not cited in the structured abstracts in English and in Turkish.* İngilizce abstract ve Türkçe öz içerisinde atıfta bulunulmamıştır.
18	⊠	*The format of headings, tables, figures, citations, references, and other details follow the **APA 6 style** as described in the Publication Manual of the American Psychological Association, 6th edition, available from http://www.apa.org* Aday makalenin başlıkları, tabloları, şekilleri, atıfları, kaynakçası ve diğer özellikleri tamamen APA altıncı baskıda belirtildiği şekildedir.
19	⊠	*All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.* Aday makalenin şekilleri ve tabloları metin içerisinde bulunmaları gereken uygun yerlere yerleştirilmiştir. Makale sonunda sunulmamıştır.
20	⊠	Citations in the text of the document include the author's surname, the year of publication, and, when there is a specific quote from a source used, a page number where the quote is located in the text. **Example:** Nothing seemed so certain as the results of the early studies (Tatt, 2001, p. 445). It was precisely this level of apparent certainty, however, which led to a number of subsequent challenges to the techniques used to process the data (Jones & Wayne, 2002, p. 879). There were a number of fairly obvious flaws in the data: consistencies and regularities that seemed most irregular, upon close scrutiny (Aarns, 2003; West, 2003, p. 457). With studies by two authors, always include both author names: (Anderson & Bjorn, 2003) As Anderson and Bjorn (2003) illustrated in their recent study As recently as 2003, a prominent study (Anderson & Bjorn) illustrated When a study has 3, 4, or 5 authors, include the names of all the authors the first time the work is cited: (Anderson, Myers, Wilkes, & Matthews, 2003)

		For all subsequent citations of this work, use "et al.": (Anderson et al., 2003)
		When a work has 6 or more authors, use et al.: (Bell et al., 2003)
		For unsigned works, include the title, enclosed in parentheses. Put quotation marks for short work titles, and italicize the titles of reports, books, and other significant works:
		("Recent Developments," 2004) (Dictionary of Tetrathalocigistic Diseases, 2004)
		Metin içindeki atıfları üstte verilen örneklere uygundur.
21	☒	Three levels of headings are used: Level 1, Level 3 and Level 4. The headings are formatted as follows: Centered Uppercase and Lowercase Heading (Level 1)
		Flush Left, Italicized, Uppercase and Lowercase Side Heading (Level 3)
		Indented, italicized, lowercase paragraph heading ending with a period. Start writing after the period (Level 4).
		Aday makale içerisinde üç farklı düzey başlık kullanılmıştır. Düzey 1, Düzey 2, Düzey 3. Başlıklar bu düzeylere uygun olarak aşağıdaki şekilde biçimlendirilmiştir:
		Ortalı ve Her Sözcüğün İlk Harfi Büyük Yazılmış Başlık (Düzey 1)
		Tam Sola Dayalı, İtalik ve Her Sözcüğün İlk Harfi Büyük Yazılmış Başlık (Düzey 3)
		İçeriden, italik, tamamı küçük harflerle yazılmış ve nokta ile biten başlık. Noktadan sonra normal metin yazımına devam edilmeli (Düzey 4).
22	☒	References are listed in alphabetical order. Each listed reference is cited in text, and each text citation is listed in the References. Basic formats are as follows:
		Haag, L., & Stern, E. (2003). In search of the benefits of learning Latin. *Journal of Educational Psychology, 95,* 174–178.
		Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Johnson, D. W., & Johnson, R. T. (1990). Cooperative learning and achievement. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 173–202). New York: Praeger.

Turkish References Only:

Çınkır, Ş., & Çetin, S. K. (2010). Öğretmenlerin okullarda mesleki çalışma ilişkileri hakkındaki görüşleri [Teachers' opinions about the professional working relationships in schools]. *Kuram ve Uygulamada Eğitim Yönetimi, 1 6*(3), 353-371.

Article in an Internet-only journal/Periodical, database

Fredrickson, B. L. (2000, March 7). Cultivating positive emotions to optimize health and well being. *Prevention & Treatment, 3,* Article 0001a. Retrieved November 20, 2000, from

http://journals.apa.org/prevention/volume3/pre0030001a.html

More information is available from:

http://citationonline.net/CitationHelp/csg04-manuscripts-apa.htm#references

Kaynakça'nın yazımı üstte verilen örneklere uygundur.

23	⊠	***Order of the main parts in the manuscript is as follows:*** *Main title in English (max. 12 words)*
		Structured abstract (min. 200- max.250 words length)
		Keywords (in English, min. four-max. six)
		Main text
		References
		Main Title in Turkish (max. 12 words
		Extended structured abstract (min.750-max.1000 words length in Turkish)
		Keywords (in Turkish, min. four-max. six)
		Aday makaleyi oluşturan ana öğeler aşağıdaki düzendedir:
		İngilizce Ana Başlık (En fazla 12 sözcük)
		Yapılandırılmış İngilizce Abstract (En az 200, en fazla 250 sözcük)

		Anahtar Sözcükler (İngilizce, en az dört, en fazla altı)
		Ana Metin
		Kaynakça
		Türkçe Ana Başlık (En fazla 12 sözcük)
		Yapılandırılmış Türkçe Öz (En az 750, en fazla 1000 sözcük)
		Anahtar Sözcükler (Türkçe, en az dört, en fazla altı)
24	☒	**Structure of the Manuscript should be as follows:**

Structure of the Manuscript should be as follows:

Manuscript Title

English Abstract (200-250 words) Sub-headings:

> *Purpose:…*
>
> *Purpose of Study:…*
>
> *Method:….*
>
> *Findings:….*
>
> *Implications for Research and Parctice:…*
>
> *Keywords:...*

Introduction Method (sub-headings):

> *Research Design*
>
> *Research Sample*
>
> ⬩*Research Instrument and Procedure*
>
> *Data Analysis*

Results, Discussion and Conclusion, References

Extended Turkish Abstract (750-1000 words) Sub-headings:

> *Problem Durumu:…*
>
> *Araştırmanın Amacı:…*

⬩ Reliability and the validity of the research instrument used or adapted in the work must be provided, and explained in detail.

		Araştırmanın Yöntemi:… *Araştırmanın Bulguları:…* *Araştırmanın Sonuçları ve Önerileri:…*
25	☒	***EJER editor reserves the right to reject manuscripts and not to publish it that do not comply fully with the EJER Manuscript Submission Guidelines.***

SUBSCRIPTION/ABONELİK

Annually / Yıllık

Individual / Kişi Aboneliği ....: 120 TL

Institution / Kurum Aboneliği ....: 280 TL

Abroad / Yurtdışı Aboneliği: ....: 250 $

Student / Öğrenci Aboneliği: ....: 70 TL

For single issue / Tek sayılar için

Individual / Kişiler ....: 25 TL

Institution / Kurumlar ....: 75 TL

Payment / Ödeme

İş Bankası Kızılay Şubesi, Anı Yayıncılık 4214-0962561

Account number / Numaralı hesap