EURASIAN JOURNAL OF EDUCATIONAL RESEARCH

Bimontly Peer-Reviewed Journal, Year: 19 Issue: 82 / 2019
İki Ayda Bir Yayımlanan Hakemli Dergi, Yıl: 19 Sayı: 82 / 2019

EURASIAN JOURNAL OF EDUCATIONAL RESEARCH

Bimontly Peer-Reviewed Journal, Year: 19 Issue: 82 / 2019
İki Ayda Bir Yayımlanan Hakemli Dergi, Yıl: 19 Sayı: 82 / 2019

PUBLISHING MANAGER / *Sahibi ve Yazı İşleri Müdürü*
Anı Publishing Education and Consultancy Advertisement Stationary Industry Trade Co. Ltd. in the name of
Anı Yayıncılık Eğitim ve Danışmanlık Reklam Kırtasiye Sanayi Ticaret Ltd. Şti. adına
Özer DAŞCAN

EDITORIAL PRODUCTION ADMINISTRATOR / *Yayın Yönetim*
Özgür GÜLER

HEADQUARTER / *Yönetim Merkezi*
Anı Yayıncılık, Kızılırmak Sokak 10/A
06640 Bakanlıklar Ankara, TURKEY
ejer.editor@gmail.com
Tel: +90 312 425 81 50 pbx Fax: +90 312 425 81 11

Printing Date / *Basım Tarihi:* 31.07.2019
Printing Address / *Matbaa Adresi:* Vadi Grafik
İvedik OSB 770. Sokak No: 101 Yenimahalle-Ankara P: +90 312 395 85 71
Broadcast Type / *Yayın Türü:* Local Broadcast / *Yerel Süreli Yayın*
Cover Design / *Kapak Tasarımı:* Anı Publishing / *Anı Yayıncılık*
Composition / *Dizgi*: Göksel ÇAKIR

Eurasian Journal of Educational Research (Print ISSN 1302-597X, e-ISSN: 2528-8911) is a bimonthly (6 issues per year) peer-reviewed journal published by Anı Yayıncılık.

Eurasian Journal of Educational Research/**EJER** is abstracted and indexed in;

ESCI, *Emerging Sources Citation Index*
ERIC, *The Education Resources Information Center*
HERDC, *Higher Education Research Data Collection*
ERA, *Educational Research Abstracts*
SCOPUS *database*
EBSCO Host *database*
ULAKBİM *national index*
Social SciSearch
Journal Citation Reports / Social Sciences Edition

CONTENTS
İçindekiler

Effect of Using Reflective Diaries in Teaching Turkish on Bilingual Students' Academic Achievement and Writing Skills
Cansu CAN, Ayse Nur KUTLUCA CANBULAT 1-24

A Comparison of Kernel Equating Methods Based on Neat Design
Cigdem AKIN ARIKAN 27-44

Evaluation Model for Evaluating Vocational Skills Programs on Local Content Curriculum in Indonesia: Impact of Educational System in Indonesia
Samsul HADI, Dedek ANDRIAN, Badrun KARTOWAGIRAN 45-62

Using Rank-order Judgments Scaling to Determine Students' Evaluation Preferences
Durmus OZBASI 63-80

Gratitude, Hope, Optimism and Life Satisfaction as Predictors of Psychological Well-Being
Ferhat KARDAS, Zekeriya CAM, Mustafa ESKISU, Sedat GELIBOLU..... 81-100

Undergraduate Counseling Trainees' Perceptions and Experiences Related to Structured Peer Group Supervision: A Mixed Method Study
Gokhan ATIK, Zeynep ERKAN ATIK 101-120

The Development of an Instrument to Assess Visuo-Semiotic Reasoning in Biology
Lindelani MNGUNI 121-136

Factors Influencing Students' Research Self-Efficacy: A Case Study of University Students in Malaysia
Racheal POH, Abdul Ghani Bin KANESAN ABDULLAH 137-168

Direct and Indirect Effects of Learning Strategies and Reading Enjoyment on PISA 2009 Reading Performance
Ezel TAVSANCIL, Ozen YILDIRIM, Safiye BILICAN DEMIR 169-190

The Contribution of Test Type and Curriculum Difference on the Effect of the National Test Score at International Mathematic Test Score: The Challenge of IR 4.0 Curriculum
Setyabudi INDARTONO, Anwaril HAMIDY 191-202

REVIEWERS of the 82nd ISSUE
82. Sayı Hakemleri

A. Mehtap Sağocak

Ade Rusman

Ado Bıchı

Ayşe Gencer

Betül Kurnaz

Bünyamin Bavlı

Cemaliye Direktör

Eren Aybek

Fatma Erdoğan

Funda Nalbantoğlu Yılmaz

Gökhan Atik

Hatice Çiğdem Yavuz

Hatice Kumandaş Öztürk

Mimi Maya Badri

Mona Lutfi Alnajami

Mustafa Başaran

Nizamettin Koç

Nurullah Eryılmaz

Onur Özmen

Pınar Çağ

Rabia Sarıkaya

Sevil Hasırcı

Sevilay Kilmen

Tuncay Türkben

Eurasian Journal of Educational Research 82 (2019) 1-26

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

Effect of Using Reflective Diaries in Teaching Turkish on Bilingual Students' Academic Achievement and Writing Skills[1]

Cansu CAN[2] , Ayse Nur KUTLUCA CANBULAT[3]

ARTICLE INFO

Article History:

Received: 28 Jul. 2018

Received in revised form: 13 Jun. 2019

Accepted: 08 Jul. 2019

DOI: 10.14689/ejer.2019.82.1

Keywords

academic achievement test for Turkish Course, writing skills scoring rubric, reflective materials

ABSTRACT

Purpose: This study aimed to identify the effect of using reflective diaries in 4th grade Turkish classes on bilingual students' academic achievement and writing skills.

Research Methods: A one-group pretest-posttest design was employed with a study group of 4th graders identified according to convenience sampling method. Data were collected by utilizing the Academic Achievement Test for Turkish Course for measuring students' academic achievement in Turkish classes in addition to the Writing Skills Scoring Rubric for evaluating students' written work and writing skills. Both the test and the rubric were developed by the researcher. During the intervention phase of the research, students were asked to write their daily achievements in their reflective diaries for 13 months. Feedback and corrections were provided for these reflective diaries and the effect of reflective diaries on writing skills were explored.

Findings: Findings revealed significant relationships between students' academic achievement and writing skills. There were no significant differences in students' academic achievement and writing skills in terms of gender.

Implications for Research and Practice: Results of the research indicate that using reflective diaries increased bilingual students' academic success and improved their writing skills irrespective of gender.

[1]This paper was produced from the master thesis titled "The Effect of Learning Journal on the Students' Academic Achievements and Writing Skills in Turkish Education", which also was presented at EJER 2018 conference held in Antalya on May 02-05, 2018.

[2] MoNE, Primary school teacher, e-mail: cansu364@gmail.com, ORCID: https://orcid.org/0000-0003-2201-7922

[3] Corresponding author, Akdeniz University Faculty of Education Primary Education Department. /Antalya, e-mail. aysenur.canbulat@gmail.com, tel:+905333308815 ORCID: https://orcid.org/0000-0003-0010-7898

Introduction

The main purpose of language teaching is to improve individuals' skills in thinking, communication, comprehension, and expression. Achieving the purpose of language training in schools that provide formal education depends on the quality of curriculum and the effectiveness of teaching and learning process employed in the implementation of that curriculum (Hamzadayi, 2010). The Turkish Language Curriculum prepared by the Turkish Ministry of National Education (MoNE, 2015) aims to ensure that students can communicate, collaborate, make joint decisions, solve problems, read written texts correctly and fluently, get accustomed to reading by employing reading literacy skills, make a habit of writing, and develop writing skills of talented students. Additionally, it is also aimed that students can express their emotions, thoughts, wishes, dreams, ideas, and impressions on a subject using language and following the rules of written expression.

Learning and language are inseparable elements. Writing is crucial for learning especially considering that it is an important learning mechanism which enables exposure, organization and refining of ideas and feelings in addition to expressing them (Graham, 2008). The Turkish Language Curriculum (MoNE, 2015) aims to develop students' writing skills in order to ensure that students can express their feelings, thoughts, dreams, ideas, and impressions by using the possibilities of language and by complying with the rules of written expression. Additionally, it is also aimed to provide pupils with the habit of expressing themselves through writing and to develop writing skills of talented students. Previous research focused on explaining the writing process, development of writing skills and the effects of different writing practices on the writing process and the nature of the writing act which is developed in this process. (Alamargot & Fayol, 2009; Berninger, Fuller, & Whitaker, 1996; Bereiter & Scradamalia, 1987; De Beaugrande, 1984; Ellis, 1988; Fayol, 1997; Flower & Hayes 1997; Galbraith, 1999; Hayes, 2011; Margolin, 1984; Martlew, 1983; Segev-Miller, 2004; Tompkins & Jones, 2008; Van Galen, 1991).

The success of students in writing activities is related to their experiences with the writing process. However, writing activities are mostly provided by teachers included in native language classes. They are usually carried out through texts which are selected by the teachers to reflect pupils' feelings and thoughts on a given topic (Ungan, 2007). Writing activities do not always generate the desired outcome due to the shortcomings in the planning of the writing process and in the organization of the necessary information used in this process. Also, students are not provided with adequate opportunities to correct their errors due to lack of feedback provided to students about the quality of their written products. However, real actions, activities, and situations are required to develop skills rather than abstract rules and explanations (Maden, 2013).

Writing skill is a crucial linguistic ability for expression and may become difficult even for Turkish students who are native speakers, simply because writing skills are not acquired naturally like speaking skills but learned through training (Ozdemir, 2016). Acquiring writing skills can be even more difficult for students if the skill is to

be acquired through a language other than their native tongue. On the other hand, the educational system is responsible for accurate teaching of the national language to all individuals to prevent societal problems caused by inadequate use of language (Sari, 2001). Many studies were conducted in Turkey on various samples to explore the effect of various factors affecting writing skills. However, a study focusing on ways of improving the writing skills of primary school students whose native languages are not Turkish has not yet been conducted.

The literature review demonstrates that materials which enable students to reflect their thoughts and feelings in writing and which help develop their writing skills are compatible with the constructivist approach. These materials are being used in the teaching and learning processes.

Reflective writing materials are tools that allow students to reflect what they learn (Haigh, 2001) and enable them to become aware of themselves and their learning (Thorpe, 2004). In previous studies, researchers referred to these materials by various expressions such as reflective notebooks, reflective diaries, student diaries, science diaries/journals, and study journals (Abali, Ozturk, & Sahin, 2014; Demirci, 2016; Er & Sasmaz Oren, 2015; Guvenc, 2011a; Guvenc, 2011b; Israel, 2007; Langer, 2002; Park, 2003; Stephens & Winterbottom, 2010; Yildirim, 2012). Reflective writing materials are the records that reflect what students accomplish and learn at classes (Haigh, 2000; Nesbit et al., 2004; Uslu, 2009; Wallin & Adawi, 2018, Wormeli, 2004).

Previous research demonstrate that reflective diaries increase student interest towards lessons, enable students to express their perceptions about the learning processes and their own achievements, contribute to the development of sense of responsibility and higher levels of recall, facilitate permanent learning, increase academic achievement, enable self-assessment and develop their writing skills in addition to the benefits for teachers who have the opportunity to monitor student development (Akkuzulu, 2011; Aymen Peker, Tas, Apaydın, & Akman, 2014; Ajello, 2000; Arslan & Ilgin, 2011; Audet, Hickman, & Dobrynina, 1996; Cardak, 2010; Cavus, 2015; Cavus & Ozden, 2012; Demirci, 2016; Duman, 2004; Eker & Arsal, 2014; Eker & Coskun, 2012; Erduran Avci, 2008; Ersozlu & Kazu, 2011; Francis, 1995; Guvenc, 2010; Hasanoglu Tektas, 2004; Karaca, Armagan, & Bektas, 2015; Nesbit et al., 2004; O'Connell, & Dyment; Park, 2003; Polat & Uslu, 2012; Spalding & Wilson,2002; Tekin Aytas, & Ugurel, 2016; Unrau, 2008; Uslu, 2009; Uline, Wilson & Cordry,2004).

While students write in their diaries, they can become aware of their learning and their motivation to learn can be positively affected (Duman, 2004). Hence, the academic success of the student may increase (Eker & Arsal, 2014). Reflective diaries are important tools for students to realize that they are learning. They also provide feedback in regards to the impact of teaching activities provided by teachers (Nesbit et al., 2004). As students continue to keep reflective diaries, they will be able to see their mistakes and have the opportunity to correct them by the help of feedback provided on the diaries (Polat & Uslu, 2012).

The outcomes provided by reflective diaries overlap with the outcomes foreseen for Turkish writing since reflective diaries allow students to structure their learning

process by having them reflect on what they have learned during classes (Arslan & Ilgin, 2011). Reflective diaries, utilized by students to express in their own words what they have learned during the class and immediately after classes, are used in developing students' writing skills. Reflective writing diaries provide students with opportunities to take responsibility for their learning in generating knowledge and ensuring permanent learning rather than solely being the receivers of information. Therefore, this study examined the effect of reflective materials on bilingual students' academic achievement and writing skills. It is believed that this study will contribute to the literature and help classroom teachers and Turkish teachers in teaching writing skills to students with native languages other than Turkish.

This study aimed to identify the effect of using reflective diaries in 4th grade Turkish classes on bilingual students' academic achievement and writing skills. In this framework, answers were sought to the questions listed below:

Does the use of reflective diaries in 4th-grade Turkish classes affect bilingual students' academic achievement and writing skills?

In regards to Turkish instruction provided by using reflective diaries;

1. "What is the relationship between pre- and post- achievement test scores and pre- and post- writing skill scores of bilingual 4th graders?"
2. Is there a significant relationship between achievement test pretest and posttest results of bilingual 4th graders?
3. Is there a significant relationship between writing skills pretest and posttest results of bilingual 4th graders?
4. Are the achievement test pretest and posttest results and writing skills pretest and posttest results of bilingual 4th graders significantly differ in terms of gender?

Method

Research Design

This study utilized a one-group pretest-posttest quasi-experimental design for examining the effect of using the reflective diary in Turkish course on bilingual students' academic achievement and writing skills. There were two 4th grade groups: 4-A and 4-B in the school where the study was conducted. The first researcher was the classroom teacher of 4-B. The study was initially intended to be designed as experimental research with pre-test and post-test control group. However, it was decided to employ one-group pre-test-posttest design because of the change of teacher in 4-A, which was designated as the control group. The new teacher of 4-A was a non-education major temporary substitute teacher. The teacher is a very important factor in the teaching process. Since research findings might be affected by the difference in the competences and qualifications of teachers, the experimental design was dropped in favor of one-group pretest-posttest quasi-experimental design.

Research Sample

The study group was determined based on convenient sampling technique. The first researcher was the classroom teacher of class 4-B in Semdinli district of Hakkari province, where the native language of some students is not Turkish. Class 4-B consisted of 14 students (seven female and seven male). These students could not speak Turkish when they enrolled in primary school. They usually experienced the Turkish language first on the TV or in kindergarten. These bilingual students speak their native languages in their daily lives and begun to use Turkish only after a period of time studying in primary school. In the first year of her career, the researcher organized Turkish teaching activities on 3rd graders based on the curriculum but she observed that her students did not even have the outcomes they were supposed to have in earlier grades (1st and 2nd), they had difficulty in regards to writing skills, and the problem related to lack of writing skills was reflected in other classes as well. Her reflections made her realize that students needed a more student-centered practice to learn writing by practicing their writing skills. Hence the researcher determined to conduct her research with this specific group.

Research Instruments and Procedures

Academic Achievement Test for Turkish Course and the Writing Skills Scoring Rubric were used in the study as data collection tools.

Academic Achievement Test for Turkish Course. Academic Achievement Test for Turkish Course was developed to measure the achievement level of primary school 4th graders in Turkish classes. To ensure content validity for the instrument, a table of specifications was prepared by using Bloom's taxonomy revised by experts (Anderson, 1999).

Table 1 represents the level of outcomes in terms of the cognitive process in Bloom's revised taxonomy.

Table 1.

Outcomes in Bloom's taxonomy.

	Steps in Cognitive Domain						
Outcomes	Number of questions	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Uses punctuation marks accurately and appropriately.	9	1		8			
Uses orthographic rules in writing.	7			7			
Uses appropriate conjunctions when expressing emotions and thoughts (for instance: but, and, or, for example, especially, for, because, for this reason, as a result, more than …, as much as).	6			6			
Identifies emotional or exaggerated elements in what he/she reads	2			2			
Identifies between real and imaginary expressions.	1				1		
Distinguishes the meanings of homonymic words.	1				1		
Distinguishes synonyms and antonyms.	2				2		
Distinguishes the real, metaphoric and operand meanings of words.	3				3		
Derives words by using suffixes and prefixes.	6		6				
Uses different types of words appropriately in writing.	7		7				
Divides the words at the end of a line with a hyphen appropriately.	3	1		2			
Identifies the main emotions in poems.	1					1	
Selects appropriate titles for texts.	1						1
Uses various types of vocabulary in writing.	3					3	
Uses portrayals and descriptions in writing.	1			1			
Writes meaningful and grammatical sentences.	1						1

The achievement test included 58 items selected according to the extent of how wide the topics are covered. To ensure construct validity, the test was implemented on 120 5th graders with the supposition that they had the same outcomes the year before. After the piloting, the test was descriptively analyzed, and item difficulty index, item

discrimination index, variance, skewness, and kurtosis values were calculated for each item on the test. Table 2 depicts the results of the descriptive analysis for the test.

Table 2.

Descriptive Analysis for the Achievement Test.

Parameters	Values
N	117
Mean	30,4444
Median	29,0000
Mode	29,00
Standard Deviation	12,41970
Variance	154,249
Skewness	,196
Standard Error of Skewness	,244
Kurtosis	-1,159
Standard Error of Kurtosis	,444
Range	45,00
Minimum	10,00
Maximum	55,00

As Table 2 depicts, skewness, and kurtosis coefficients were close to 0. Additionally, mode, median, and mean values were very close to each other.

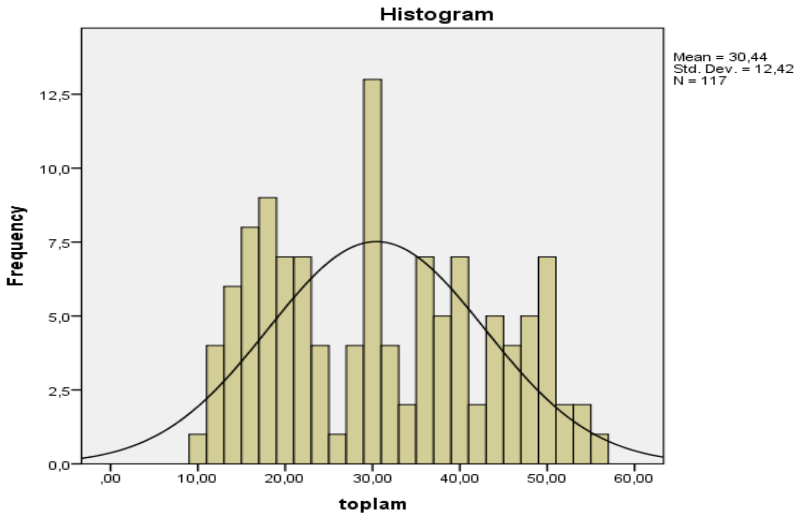


Figure 1. *Histogram of Achievement Test.*

The histogram for an ideal distribution should be symmetrical. The mode, median, and mean values were very close to each other as shown in Table 2 and Figure 1. Also, the histogram of the distribution was neither too sharp nor too kurtic. An opinion can be formulated on the data based on the proximity of skewness and kurtosis coefficients to 0 (Can, 2013). With the help of all these values, the distribution can be interpreted as ideal.

Item difficulty index (Pj) points to the percentage of respondents who provided correct answers. Because a participant can answer all items correctly, or no one can answer the item, Item difficulty index values are in the range of [0, 1]. Item discrimination index (Rj) indicates the extent to which success on an item corresponds to success on the whole test and it is the correlation between item scores and test scores. Item discrimination index can take values between [-1, +1] (Baykul, 2015). Negative values mean that the participants who received low scores from the test answered the item correctly whereas high scores mean the participants who received high scores from the test answered the item correctly. Since such items would not serve the purpose of the test, the items with negative item discrimination index values were excluded from the Academic Achievement Test for Turkish Course. After these items were excluded, the item difficulty index and item discrimination index were calculated to be 0.54 and 0.57 respectively. After these evaluations, the final form of Academic Achievement Test for Turkish Course was used with 54 items. KR-20 reliability coefficient, which is known as the internal consistency coefficient after the application, was found to be 0.986. These values indicate that the prepared test had moderate difficulty, was distinctive, and reliable. Table 3 displays the results of item analysis for some questions.

Table 3.

Item Analysis of Some of the Questions Included in the Achievement Test.

		Options				Statistical Analyses				
Item Number	**Group**	**A**	**B***	**C**	**D**	**Pj**	**Rj**	**Sj**	**skewness**	**kurtosis**
	Upper	1	30	2	0					
1						0,50	0,81	0,91	0,57	-0,49
	Lower	11	3	7	10					
Item Number	**Group**	**A***	**B**	**C**	**D**	**Pj**	**rj**	**Sj**	**skewness**	**kurtosis**
	Upper	31	0	0	1					
2						0,51	0,90	1,57	0,86	-0,62
	Lower	2	11	8	9					
Item Number	**Group**	**A**	**B***	**C**	**D**	**Pj**	**rj**	**Sj**	**skewness**	**kurtosis**
	Upper	1	31	0	0					
3						0,64	0,65	0,65	1,17	1,82
	Lower	5	10	10	5					

Note: **: Correct answer for the item. Pj: item difficulty index; Rj: item discrimination index; Sj: variance.*

Writing Skills Scoring Rubric. The rubric was developed to evaluate 4th graders' texts written before and after the treatment in terms of writing skills. Scoring rubrics are tools that help identify the criteria for specific outcomes and the levels for these outcomes from perfect to weak (Goodrich, 2001). They allow teachers to be more objective in evaluating students (Kutlu, Dogan, & Karakaya, 2010, p. 53) because they provide opportunities to monitor what areas need improvement and to observe successful areas (Aslanoglu, 2003).

According to Moskal (2000) and Goodrich (2000), the steps in developing a scoring rubric are as follows: identifying a goal in line with the expected competencies, listing criteria based on these competencies, determining the levels for these criteria, making decisions as to which type of scoring will be used, providing students information about the rubric, implementing the treatment and scoring it based on the rubric, giving feedback to students about the results and making corrections.

Writing goals of Writing Skills Scoring Rubric correspond to outcomes determined in Turkish Language Curriculum for 1st to 4th grades (MoNE, 2015). Writing outcomes that cannot be measured in a single intervention like the one used in this study, which require a lengthier process, were not included in the evaluation. Criteria were determined to assess the acquisitions. Expert opinion was sought after preparing the outcomes in order to ensure content validity. The rubric consists of three parts. The first part consists of expressions related to grammar; the second part consists of expressions related to the text, and the third part consists of expressions related to the appearance/view of the text.

The rubric has been organized according to the scoring system in which the following levels are identified: "highly inadequate (1)", "inadequate (2)", "partially adequate (3)", "adequate (4)", "highly adequate (5) for each writing acquisition. There are two types of scoring rubrics: holistic and analytic. Holistic scoring rubrics allow holistic performance assessment without separately evaluating parts of the performance (Nitko as cited in Bekiroglu, 2008). For example, when teachers are to score compositions with holistic scoring rubrics, they should decide how many points will be given to the whole text by taking the criteria into consideration (Kutlu, Dogan, & Karakaya, 2010).

Analytic scoring rubrics are usually used where a single response needs deep focus (Nitko as cited in Mertler, 2001). Analytic scoring rubrics are more specific compared to holistic scoring rubrics and provide opportunities for more in-depth evaluations. Therefore, Writing Skills Scoring Rubric was developed as an analytic scoring rubric. Thus, the texts and reflective diaries written by students were scored in more detail. According to the scoring rubric, a student can receive a minimum of 19 points and a maximum of 95 points. The opinions of five experts were consulted to determine how much the graded scoring key corresponds to the items. One professor, three associate professors and one assistant professor comprised the expert board. All experts were working in the field of Turkish education. The CGI value (0.95), which was calculated according to Lawshe (1975) technique, exceeded the criteria for a minimum value of content validity (0.87) that Wilson, Pan, and Schumsky (2012) determined for five

experts. Accordingly, the content validity of the analytical grade scoring key of Writing Skills Scoring rubric was considered to be statistically significant.

First of all, Spearman's rank correlation coefficient t was used to analyze inter-rater reliability in the analyses where students' writing skills scores were used in order to determine the reliability of the scores provided by the researcher. Table 4 depicts the results of Spearman's rank correlation coefficients used for observing the interrater reliability of Writing Skills Scoring.

Table 4.

Spearman's Rank Correlation Analysis for Interrater Reliability of Writing Skills Scoring.

		Spearman
Rater1pre – Rater2pre	r	,958**
	p	,000
	N	14
Rater1post –Rater2post	r	,955**
	p	,000
	N	14

** *p<.01*

Spearman's rank correlation coefficient values varied between -1 and +1. When the correlation coefficient value approaches +1, the relationship in the positive direction increases and this indicates that variables increase or decrease at the same time (Kilmen, 2015). The finding that the correlation coefficient for teacher scores was greater than 0.70 shows high-level consistency between teachers (Kutlu, Dogan, & Karakaya, 2010, p. 85). As Table 4 shows, there was a high-level positive correlation between raters.

Data Collection

The study started after obtaining the necessary permissions from the Hakkâri Provincial Directorate of National Education. At the outset of the intervention process, the validity and reliability studies were conducted for an academic achievement test and writing skills scoring rubric. Finalized form of Writing Skills Scoring Rubric was given to study group students as the pretest. Students were asked to write about "Summer Holiday" for the pretest which aimed to measure students' writing skills and the texts written by students were scored with the rubric by two classroom teachers one of whom was the researcher.

After the pretest, based on the treatment schedule; students wrote in their reflective diaries for a full class hour on a weekly basis each Thursday after the lesson was completed. As Maden (2013) stated, reflective diaries should be written at certain intervals. Reflective diaries were examined by the researcher on the evening of the day they were written and returned to students the next day to read what they had written accompanied by the teacher. Students were asked to think about their errors and they were provided opportunities to reflect on their mistakes. Students then were given the

necessary feedback and asked to control and correct their mistakes. The points they should take into consideration in their next writing were explained. Reflective diaries were evaluated by different teachers each week. The intervention lasted 13 weeks excluding the pretest and posttest sessions.

Writing Skills Scoring Rubric was given to study group students as a posttest. Students were asked to write about "Their Dreams" in the context of the posttest to measure their writing skills. Posttest can be implemented with the same measurement tool used in pretest or with a similar measurement tool in the dame format (Buyukozturk, 2016). Since the focus in the study was not the content but whether students could reflect on their writing skills, students were given a different topic during the posttest. Student texts were scored with the help of the Writing Skills Scoring Rubric by two teachers.

Data Analysis

According to literature, techniques that do not depend on the distribution should be selected in cases when the researcher is not sure whether the requirements for parametric techniques are met since impartment in hypotheses may increase in parametric tests when sample size decreases in small scale samples (when it is less than 30) (Akdag, 2011; Gamgam, 1989; Sipahi, Yurtkoru, & Cinko, 2008). For this reason, non-parametric techniques were preferred in all analyzes.

Results

Findings related to First Sub-Problem

Table 5 depicts the findings with regards to the following question: "What is the relationship between pre- and post- achievement test scores and pre- and post- writing skill scores of bilingual 4th graders?"

Table 5.

Correlations Between Achievement Test Pretest and Posttest And Writing Skills Pretest and Posttest.

	Test pretest total	**Test posttest total**	**Writing Pretest Total**	**Writing Posttest Total**
TEST Pretest Total	1,000	,800**	,850**	,833**
		,001	,000	,000
	14	14	14	14
TEST Posttest Total		1,000	,898**	,954**
			,000	,000
		14	14	14
Writing Pretest Total			1,000	,948**
				,000
			14	14
Writing Posttest Total				1,000
				14

**. p<.01

It was observed that there were strong significant correlations between academic achievement and writing skills. The positive correlations between pretest academic achievement score and both pre- and post- writing skill scores indicate that; those higher-achieving bilingual 4th graders have stronger writing scores and benefit more from the treatment. Moreover, those bilingual 4th graders who increase their academic achievement during the semester, seem to benefit even more from the treatment. This may also indicate that reflective diaries may contribute to the academic achievement of the students.

Findings Related to the Second Sub-Problem

Table 6 displays the findings related to the research question "whether there was a statistically significant relationship between achievement test pretest and posttest results of bilingual 4th graders".

Table 6.

Students' Pretest and Posttest Scores Attained From The Achievement Test.

Test posttest total Test pretest total	**N**	**Mean rank**	**Rank sum**	**Z**	**p**
Negative rank	1[a]	2,00	2,00	-3,173[d]	,002
Positive rank	13[b]	7,92	103,00		
Equal	0[c]				
Total	14				

a. Testsposttesttotal<Testpretesttotal
b. Testsposttesttotal > Testpretesttotal
c. Testsposttesttotal = Testpretesttotal
d. Dependent on negative rank

Table 6 shows that the level of significance (p) for Z value obtained as a result of the Wilcoxon Signed Ranks Test was lower than 0.05. Accordingly, there was a significant difference between students' pre and posttest achievement scores. According to this finding, it can be said that there was a significant increase in the academic achievement of students (z= -3,173; p<0, 05). This may also indicate that writing diaries may contribute to the academic achievement of the students. Since negative rank was 1, and the positive rank was 13, all except one student had an increase in their posttest scores (Kilmen, 2015, p. 249).

Findings Related to the Third Sub-Problem

Table 7 represents the findings related to the research question "whether there was a statistically significant relationship between writing skills pretest and posttest results of bilingual 4th graders."

As Table 7 points out, the level of significance (p) for Z value obtained as a result of the Wilcoxon Signed Ranks Test was lower than 0.05. Accordingly, there was a significant difference between students' pre and post-test scores related to writing skills (z= -3,29; p<0,05). This may also indicate that writing diaries may contribute to

the writing skills of the students. Since negative rank was zero and positive rank was 14, all students had an increase in their posttest scores (Kilmen, 2015, p. 249).

Table 7.

Students' Pretest and Posttest Scores Attained From The Writing Skills Test.

WritingPosttesttotal - WritingPretesttotal	N	Mean rank	Rank sum	Z	p
Negative rank	0[a]	,00	,00	-3,297[d]	,001
Positive rank	14[b]	7,50	105,00		
Equal	0[c]				
Total	14				

a. WritingPosttesttotal <
b. WritingPosttesttotal >
c. WritingPosttesttotal =
d. Dependent on negative

Findings Related to the Fourth Sub-Problem

Table 8 depicts the findings related to the research question "whether the bilingual 4th grader's achievement test pretest and posttest results and writing skills pretest and posttest results statistically significant in terms of gender".

As shown in Table 8, there were no significant differences between the pre and posttest scores of students in the achievement test and writing skills based on gender ($p > 0.05$). As shown in Table 8, the mean scores of female students were higher than those of male students, but this difference was not found to be significant. According to this finding, it can be said that the course diaries may be contributing to an increase in the academic achievement and writing skills of students irrespective of gender.

Table 8.

Gender-based Difference Between Students' Achievement Test Pretest and Posttest Scores and Their Writing Skills Pretest and Posttest Scores.

	Gender	N	*Mean rank*	*Rank sum*	U	p
Test Pre Total	1,00	7	8,21	57,50	19,500	,522
	2,00	7	6,79	47,50		
	Total	14				
Test Post Total	1,00	7	9,64	67,50	9,500	,054
	2,00	7	5,36	37,50		
	Total	14				
Writing Pre Total	1,00	7	8,57	60,00	17,000	,336
	2,00	7	6,43	45,00		
	Total	14				
Writing Post Total	1,00	7	9,14	64,00	13,000	,141
	2,00	7	5,86	41,00		
	Total	14				

Note: *1 (female students) and 2 (male students).*

Discussion, Conclusion, and Recommendations

The findings of the study showed that, that there was a high level of relationship between students' pre-test and post-test scores in the achievement test and in the writing test. According to this finding, reflective diaries may contribute to the academic achievement of the students. While students write in their diaries, they may become aware of their learning, and their learning motivation can be positively affected (Duman, 2004). Therefore, academic success may increase (Eker & Arsal, 2014). Accordingly, it can be said that writing skills increase when students learn Turkish better. It can also be argued that expressing their feelings, thoughts, and impressions may affect students' achievement positively. Studies in the literature related to this field support these findings as well (Yorusun, 2013; Tekin, Aytas, & Ugurel, 2016).

The second finding of the study points to a significant difference between students' achievement test pre and post-test scores in favor of post-test scores. The reason for the difference may be the fact that students in the study kept their own learning process under control by writing a reflective diary. They reviewed the topics of study and their learning was supported by feedback and corrections. This result is in parallel with the studies which examined the effect of reflective diaries on student achievement (Akkuzulu, 2011; Arslan & Ilgın, 2011; Demirci, 2016; Dunlap, 2006; Eker & Arsal, 2014; Eker & Coskun, 2012; Ersozlu & Kazu, 2011; Guvenc, 2010; Hasanoglu Tektas, 2004; Liuoliene & Metiuniene, 2009; Park, 2003).

The third finding of the study indicates a significant difference between students' writing skills pre and post-test scores in favor of posttest scores. This finding may be associated with the opportunities provided by reflective diaries to enhance learning by writing. Writing is not a skill that can be learned in theory; it requires practice to enhance writing skills. The more students write, and the more they apply rules related to writing in their texts, the more they improve their writing skills. The literature review demonstrates that previous studies (Arslan & Ilgin, 2011; Butler & Nesbit, 2008) obtained results that were parallel with the finding related to this subproblem.

According to the fourth research finding, there were no significant differences based on gender in students' achievement test and writing skills pretest and posttest scores. This may be because the social environment in which both female and male students are raised and have identical general cultural expectations, upbringing styles, activities, and childcare. The school they attend has an identical approach to both genders in education (Capri & Celikkaleli, 2005). In previous studies which focus on the relationship between gender and academic achievement; while Ipek and Malas (2013) found no gender-related differences between students' pretest scores in academic achievement test, Coskun (2006) and Erduran Avci (2008) found that students' written expression skills differ significantly according to gender. Tulu (2009), who examined the factors affecting the language level of bilingual and monolingual students, found that gender affects the language development of bilingual students. On the other hand, it does not affect the language development of monolingual students.

This study aimed to determine the effect of using reflective diaries on students' academic achievement and writing skills in Turkish courses. Experimental research studies may further be conducted in order to determine the effect of using reflective diaries on students' reading and verbal communication skills. Further studies can be conducted by using an experimental design with pretest and posttest and control groups in order to better determine whether the use of reflective diaries has an effect on students' academic achievement and writing skills. Further studies may focus on the relationship of using reflective diaries with different school subjects at different levels. This study was conducted in a single class with a small study group. The same research can be replicated in other provinces from different regions in Turkey with larger study groups.

References

Abali Ozturk, Y. & Sahin, C. (2014). Alternatif olcme degerlendirme yontemlerinin akademik basari, kalicilik, ozyeterlilik algisi ve tutum uzerine etkisi [The effects of alternative assessment and evaluation methods on academic achievement, persistence of learning, self-efficacy perception and attitudes]. *Egitimde Kuram ve Uygulama*, 10(4), 1022-1046.

Ajello, T. (2000). Science journals: writing, drawing and learning. *Teaching Pre K-8*, February, 56-57.

Akdag, M. (2011). İstatistiksel analizlerde "n" sayıları. Inonu Universitesi. Retrieved from http://iys.inonu.edu.tr/webpanel /dosyalar/669/file/SPSS.

Akkuzulu, D. (2011). *Yedinci sinif ogrencilerinin fen ve teknoloji dersi cevre ve insan unitesinde yansitici gunlukleri tutmasinin basarı ve tutuma etkisi [The effect of reflective keeping dairies by 7th grade students' in science and technology course, environment and human unit, on students' achievements and attitudes]*. (Yayinlanmamis Yuksek Lisans Tezi). Gazi Universitesi Egitim Bilimleri Enstitusu, Ankara.

Alamargot, D. & Fayol, M. (2009). *Modeling the development of written composition*. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.) The Sage handbook of writing development (pp. 23-47). Los Angeles: Sage

Anderson, L. W. (1999). *Rethinking Bloom's taxonomy: implications for testing and assessment*. ERIC Document Reproduction, Service No. ED435630, TM 030 228.

Arslan, D. & Ilgin, H. (2011). Turkce dersinde ogrenci gunluklerinin degerlendirme araci olarak kullanilmasi [Using student journals as an assessment tool for Turkish lesson]. *Mustafa Kemal Universitesi Sosyal Bilimler Enstitusu Dergisi,* 8(16), 225-238.

Aslanoglu, E, A. (2003). *Ogretimde sunu becerilerinin degerlendirilmesinde dereceli puanlama anahtari (rubric) kullanilmasina iliskin bir arastirma [A research on rubric in evaluating the presentation skills in education]*. (Yayinlanmamıs Yuksek Lisans Tezi), Ankara Universitesi Sosyal Bilimler Enstitusu, Ankara.

Aymen Peker, E., Tas, E., Apaydin, Z. & Akman, E. (2014). Fen ve teknoloji dersi yasamimizdaki elektrik unitesi icin dusunme ajandasi (ogrenci gunlugu) tutulmasi ve tutulan ajandalarin ogrencilerin akademik basarisi uzerindeki etkileri [Keeping the thinking agenda (student diary) for electricity unit in science and technology course and the effects of agendas held on at the students' academic achievement]. *International Journal of New Trends in Arts, Sports & Science Education*, 3(3), 8-27.

Baykul, Y. (2015). *Egitimde ve psikolojide olcme: klâsik test teorisi ve uygulamasi*. (3. baski). Ankara: Pegem Akademi.

Bekiroglu, F. O. (2008). Performansa dayalı olcumler: teori ve uygulama [Performance based assessments: theory and practice]. *Turk Fen Egitimi Dergisi*, 5(1), 113-131.

Bereiter, C. & Scradamalia, M. (1987). *The psychology of written composition*. Hillsdale, New Jersey: Erlbaum.

Berninger, V., Fuller, F., & Whitaker, D. (1996). A process approach to writing development across the life span. *Educational Psychology Review*, 8, 193-218

Butler, M. B. & Nesbit, C. (2008). Using science notebooks to improve writing skills and conceptual understanding. *Science Activities*, 44(4), 137-145.

Buyukozturk, S., Kılıc, C. E., Akgun, O. A., Karadeniz, S. & Demirel, F. (2016). *Bilimsel arastirma yontemleri*. (22.baskı). Ankara: Pegem Akademi Yayıncılık.

Can, A. (2013). *SPSS ile bilimsel arastırma surecinde nicel veri analizi*. (1. baski). Ankara: Pegem Akademi Yayıncılık.

Capri, B. & Celikkaleli, O. (2005). Ilkogretim birinci kademedeki (7-11 yas grubu) cocukların korunum gelisim duzeylerinin cinsiyet ve sınıf degiskeni acısından incelenmesi [An investigation of elementary school students' (7–11-years-old) conservation development levels according to sex and grade]. *Mersin Universitesi Egitim Fakultesi Dergisi*, 1(1), 48-65.

Cardak, U. (2010). *Fen ve Teknoloji dersine iliskin gunluk tutmanın ogrenci basarısı ve tutumu uzerine etkisi [The effect of using a diary in science and technology teachin on the success and attitude of students to the lesson]*. (Yayinlanmamis Yuksek Lisans Tezi). Sakarya Universitesi Fen Bilimleri Enstitusu, Sakarya.

Cavus, E. (2015). *Fen ve Teknoloji dersinde fen gunlugu kullanımının ilkogretim ogrencilerinin bilisustu farkındalık ve akademik basarısına etkisi [In science and technology course use of journal science primary students' metacognitive awareness and academic success]*. (Yayinlanmamis Yuksek Lisans Tezi). Adiyaman Universitesi Fen Bilimleri Enstitusu, Adiyaman.

Cavus, E. & Ozden, M. (2012). İlkogretim ogrencilerinin fen ve teknoloji dersinde fen gunlugu kullanımına iliskin gorusleri [Primary school students' views about use of science journal in science and technology]. *Adiyaman Universitesi Egitim Bilimleri Dergisi,* 2(1).

De Beaugrande, R. (1984). *Text production: Toward a science of composition*. Norwood, NJ: Ablex.

Demirci, E. (2016). *Ilkogretim 7. sinif fen ve teknoloji dersi yasamimizdaki elektrik unitesinde ogrenci gunluklerinin kullaniminin ogrencilerin ust bilissel beceri gelisimine ve basarilarına etkisi [The effects of keeping the student diaries for electricity unit in 7th grade science and technology course on the students' metacognitive development and academic achievement].* (Yayinlanmamis Yuksek Lisans Tezi). Pamukkale Universitesi Egitim Bilimleri Enstitusu, Denizli.

Duman, B. (2004). *Ogrenme-ogretme kuramları ve surec temelli ogretim.* (1. baski). Ankara: Ani Yayincilik.

Dunlap, J. C. (2006). Using guided reflective journaling activities to capture students' chancing perceptions. Tech Trends, 50, 6, 20-26.

Eker, C. & Arsal, Z. (2014). Oz duzenleme stratejileri ogretimi surecinde ders gunlukleri kullanmanin oz duzenleme stratejileri ogretimine etkisi [Impact of teaching diaries on the use of students' self-regulation strategies]. *Karaelmas Journal of Educational Sciences*, 2, 80-93.

Eker, C. & Coskun, İ. (2012). Ders gunlugu yazmanin ilkokul 4. sinif ogrencilerinin sosyal bilgiler dersi akademik basarilarina etkisi [The effect of writing course diaries on academic achievement of elementary school fourth grade students]. *Mugla Sitki Kocman Universitesi Sosyal Bilimler Enstitusu Dergisi*, 29, 111-122.

Ellis, A. W. (1988). Normal writing processes and peripheral acquired dysgraphias. *Language and Cognitive Processes*, 3(2), 99-127.

Er, O. & Sasmaz Oren, F. (2015). Fen ve teknoloji dersi 7. sinif "isik" unitesinde alternatif degerlendirme yaklasimlari temelli ogretimin ogrencilerin akademik basari ve tutumlari uzerine etkisi [The effect of the alternative assessment approaches based education during the "light" unit of science and technology class in 7th grade on the academic achievements and attitudes of students]. *Celal Bayar Universitesi Sosyal Bilimler Dergisi*, 13(4), 135-164.

Erduran Avcı, D. (2008). Fen ve teknoloji egitiminde ogrenci gunluklerinin kullanilmasi [Using student diaries in science and technology education]. *Eurasian Journal of Educational Research*, 30, 17-32.

Ersozlu, Z. N. & Kazu, H. (2011). Ilkogretim besinci sinif sosyal bilgiler dersinde uygulanan yansitici dusunmeyi gelistirme etkinliklerinin akademik basariya etkisi [The effects of reflective thinking activities on the academic successes of fifth grade primary social studies students]. *Uludag Universitesi Egitim Fakultesi Dergisi*, 24(1), 141-159.

Fayol, M. (1997). On acquiring and using punctuation: A study of written French. In J. Costermans & M. Fayol (Eds.), Processing interclausal relationships: Studies in the production and comprehension of text (pp. 157–178). Mahwah, NJ: Lawrence Erlbaum

Flower L. &Hayes, J. R. (1997). A cognitive process theory of writing. Villanueva, Victor, Jr., (Ed.) Cross-talk in comp theory: a reader in (s. 251-276). Urbana: National Council of Teachers of English.

Francis, D. (1995). The reflective journal: A window to preservice teachers' practical knowledge. Teaching and teacher Education, 11(3), 229-241.

Galbraith, D. (1999). Writing as a knowledge-constituting process. Knowing what to write: Conceptual processes in text production, 4, 139-164.

Gamgam, H. (1989). *Parametrik olmayan istatistiksel teknikler*. (1. baski). Ankara: Gazi Universitesi Egitim Fakultesi Matbaasi.

Goodrich, A. H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.

Goodrich, A. H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4(4).

Graham, S. (2008). Research on writing development, practice, instruction, and assesment: Introduction to a special issue of reading and writing. *Read Write*, 21:1-2.

Guvenc, H. (2010). İsbirlikli ogrenme ve ders gunluklerinin ogretmen adayi ogrencilerin oz duzenlemeli ogrenmeleri uzerindeki etkileri [The effects of cooperative learning and learning journals on teacher candidates' self-regulated learning]. *Kuram ve Uygulamada Egitim Bilimleri*, 10(3), 1459-1487.

Guvenc, H. (2011a). Calisma gunluklerinin 6. sinif ogrencilerinin oz duzenlemeli ogrenmeleri uzerindeki etkileri [Effects of study diaries on sixth graders self-regulated learning]. *Hacettepe Universitesi Egitim Fakultesi Dergisi*, 41, 206-218.

Guvenc, H. (2011b). Yansitma materyalleriyle desteklenen isbirlikli ogrenmenin Turkce ogretmeni adaylarinin ozduzenlemeli ogrenmelerine etkileri [The effects of cooperative learning supported with reflective materials on Turkish teacher Candidates self-regulation]. *Egitim ve Bilim*, 36(159), 3-13.

Haigh, J. M. (2001). Constructing Gaia: Using journals to foster reflective learning. *Journal of Geography in Higher Education*, 25(2), 167-189.

Hamzadayi, E. (2010). *Butunlestirilmis ogrenme-ogretme yaklasiminin turkce ogretiminde okudugunu anlama ve yazili anlatim becerilerine etkisi [The influence of integrated learning -teaching approach in Turkish language teaching on reading comprehension and written expression skills]*. (Yayınlanmamis Yuksek Lisans Tezi). Ankara Universitesi Sosyal Bilimler Enstitusu, Ankara.

Hasanoglu Tektas, A. (2004). *Matematik gunluklerinin ogrencilerin matematik basarilari, matematige karsi olan tutumlari ve matematik kaygilari uzerinde etkileri [Effects of math learning journals on mathematics achievement, attitudes toward mathemathic, and mathematics anxiety]*. (Yayinlanmamis Yuksek Lisans Tezi). Bogazici Universitesi, Fen Bilimleri Enstitusu, İstanbul.

Hayes, J. R. (2011). Kinds of knowledge-telling: Modeling early writing development. *Journal of Writing Research*, 3(2), 73-92.

Ipek, J. & Malas, H. (2013). Bilgisayar destekli matematik dersinde STAR stratejisinin ilkogretim 2. sınıf ogrencilerinin matematik dersi basarilari ve problem cozme becerileri uzerindeki etkisi [The effects of star strategy of computer-assisted mathematics lessons on the achievement and problem solving skills in 2nd grade courses]. *Necatibey Egitim Fakultesi Elektronik Fen ve Matematik Egitimi Dergisi*, 7(2), 314-345.

Israel, E. (2007). *Oz duzenleme egitimi, fen basarisi ve ozyeterlilik [Self-regulation instruction, science achievement and self-efficacy]*. (Yayinlanmamis Doktora Tezi). Dokuz Eylul Universitesi, Egitim Bilimleri Enstitusu, İzmir.

Karaca, M., Oner Armagan, F. & Bektas, O. (2015). The use of the reflective diaries in science lessons from the perspectives of eighth grade students. *International Journal of Environmental &Science Education*, 11(2), 53-74.

Kilmen, S. (2015). *Egitim arastirmacilari icin SPSS uygulamali istatistik*. Ankara: Edge Akademi.

Kutlu, O., Dogan, C. D., ve Karakaya, İ. (2010). *Ogrenci basarisinin belirlenmesi performansa ve portfolyoya dayali durum belirleme*. (3. baskı). Ankara: Pegem Akademi.

Langer, A. M. (2002). Reflecting on practice: using learning journals in higher and continuing education. *Teaching in Higher Education*, 7(3), 337-351. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/13562510220144824

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.

Liuoliene, A. & Metiuniene, R. (2009). Learning through reflective journaling. Coactivity: *Philology Educology / Santalka: Filologija: Edukologija*, 17(4), 32-37.

Maden, S. (2013). Temel dil becerileri egitiminde kullanilabilecek aktif ogrenme ogretimsel istatistikleri [Instructional work and tactics in active learning used on basic language arts education]. *Ana Dili Egitimi Dergisi*, 1(2), 20-35.

Margolin, D. I. (1984). The neuropsychology of writing and spelling: Semantic, phonological, motor, and perceptual processes. *The Quarterly Journal of Experimental Psychology*, 36(3), 459-489.

Martlew, M. (1983). Problems and difficulties: Cognitive and communicative aspects of writing. *The Psychology of Written Language*, 295-333.

Ministry of National Education. (2015). Turkce dersi (1-8. Siniflar) ogretim programi [Turkish language curriculum for 1st to 8th grades]. Retrieved from http://ttkb.meb.gov.tr/

Myhill, D. (2009). Developmental trajectories in the mastery of paragraphing: Toward a model of development. *Written Language and Literacy* 12(1), 26-51.

Nesbit, C. R., Hargrove, T. Y., Harrelson, L., & Maxey, B. (2004). Implementing science notebooks in the primary grades. *Science Activities*, 40(4), 21-29.

O'Connell, T. S., & Dyment, J. E. (2011). The case of reflective journals: Is the jury still out? *Reflective Practice*, 12(1), 47-59.

Ozdemir, H. (2016). *Ortaokul besinci sinifta okuyan iki dilli ogrencilerin yazma becerilerinin degerlendirilmesi [Evaluation of secondary school 5th grade bilingual students writing skills]*. (Yayinlanmamis Yuksek Lisans Tezi). Yuzuncu Yil Universitesi, Egitim Bilimleri Enstitusu, Van.

Park, C. (2003). Engaging students in the learning process: the learning journal. *Journal of Geography in Higher Education*, 27(2), 183-199.

Polat, S. & Uslu, M. (2012). Fen ve teknoloji dersinde ust bilis stratejilerine dayali ogretim uygulamasinin 5. sinif ogrencilerinin erisilerine etkisi [The effect of teaching based on metacognitive strategies on 5th grade students' achievement in science and techonology course]. *Usak Universitesi Sosyal Bilimler Dergisi*, 5(3), 28-43.

Ruiz-Primo, M. A., Li, M., Ayala, C. & Shavelson, R. J. (1999). *Student science journals and the evidence they provide: classroom learning and opportunity to learn*. Paper presented at the NARST Annual Meeting, Boston.

Sari, M. (2001). *İki dilli cocuklarin cozumleme yontemiyle okuma-yazma ogrenirken karsilastiklari guclukler [Problems which bilingual children face while learning basic reading and writing with analysis method]*. (Yayinlanmamis Yuksek Lisans Tezi). Cukurova Universitesi, Sosyal Bilimler Enstitusu, Adana.

Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. L1-*Educational Studies in Language and Literature*, 4(1), 5-33.

Sipahi, B., Yurtkoru, E. S. & Cinko, M. (2008). *Sosyal bilimlerde SPSS'le veri analizi*. (2. baski). İstanbul: Beta Basim.

Spalding, E., & Wilson, A. (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record*, 104(7), 1393-1421.

Stephens, K. & Winterbottom, M. (2010). Using a learning log to supports tudents' learning in biology lessons. *Journal of Biological Education*, 44(2), 72-80.

Tekin Aytas, C. & Ugurel, I. (2016). Bir matematik sinifindaki yazma aktivitelerine dayali ogretim uygulamasinin ogrencilerin ogrenmeleri uzerindeki etkileri [The effects of an instruction practice based on the writing activities on students' learnings in a mathematics class]. *Milli Egitim*, 211, 113-146.

Tompkins, G. E., & Jones, P. D. (2008). *Teaching writing: Balancing process and product*. Pearson/Merrill Prentice Hall.

Thorpe, K. (2004). Reflective learning journals: From concept to practice. *Reflective Practice*, 5(3), 327-343.

Tulu, Y. (2009). *Ana dili Turkce olan ve ana dili Turkce olmayan (iki dilli) 4–7 yas cocuklarin dil duzeylerine etki eden faktorlerin incelenmesi [The examination of the factors which affect the language development of children at the age of 4-7 whose native language Turkish and native language not Turkish (bilingual)]*. (Yayinlanmamis Yuksek Lisans Tezi). Selcuk Universitesi, Sosyal Bilimler Enstitusu, Konya.

Uline, C., Wilson, J. D., & Cordry, S. (2004). Reflective journals: A valuable tool for teacher preparation. *Education*, 124(3).

Ungan, S. (2007). Yazma becerisinin gelistirilmesi ve onemi [Develepment of and importance of writing skills]. *Sosyal Bilimler Enstitusu Dergisi*, (23), 461-472.

Uslu, H. (2009). *Altinci ve yedinci sinif fen ve teknoloji ile matematik derslerinde gunluklerin kullanilmasina yonelik ogrenci goruslerinin belirlenmesi [Students' ideas about writing journals in science and technology and mathematics lessons in the sixth and seventh grades]*. (Yayinlanmamis Yuksek Lisans Tezi). Suleyman Demirel Universitesi, Fen Bilimleri Enstitusu, Fen Bilgisi Egitimi Anabilim Dali, Isparta.

Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science,* 10(2-3), 165-191.

Wallin, P., & Adawi, T. (2018). The reflective diary as a method for the formative assessment of self-regulated learning. *European Journal of Engineering Education*, 43(4), 507-521.

Wilson, F. R., Pan, W. & Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development,* 45(3), 197–210.

Wormeli, R. (2004). *Summarization in any subject: 50 techniques to improve student learning.* Retrieved from https://ebookcentral.proquest.com/lib/akdeniz-ebooks/reader.action?docID=3002108

Yildirim, C. (2012). *Bilimsel surec becerileri etkinliklerinin ilkogretim 7. sinif ogrencilerinin yansitici dusunmelerine etkisi [The effect of scientific process skills activities on elementary school 7th grade student's reflective thinking].* (Yayinlanmamis Yuksek Lisans Tezi). Pamukkale Universitesi Fen Bilimleri Enstitusu, Denizli.

Yildiz, A. & Buyukkasap, E. (2011). Ogretmen adaylarinin Compton olayini anlama duzeyleri ve ogrenme amacli yazma aktivitelerinin akademik basariya etkisi [Prospective teachers' levels of understanding Compton Effect and the impact of writing activities for learning purposes on academic success]. *Uluslararasi Insan Bilimleri Dergisi*, 8(1), 1643-1664.

Yorusun, S. (2013). *8. sinif ogrencilerinin dusunce yazisi yazma becerileri ile akademik basarilari arasindaki iliski uzerine bir inceleme [A research of the opinion writing skills and the academic achievement relation of 8. grade students: The example of Silifke Ataturk Elementary School].* (Yayinlanmamis Yuksek Lisans Tezi). Afyon Kocatepe Universitesi, Sosyal Bilimler Enstitusu, Afyonkarahisar.

Türkçe Öğretiminde Ders Günlüğü Kullanımının İki Dilli Öğrencilerin Akademik Başarıları ile Yazma Becerilerine Etkisi

Atıf:

Can, C., & Kutluca-Canbulat, A.N. (2019). Effect of using reflective diaries in teaching Turkish on bilingual students' academic achievement and writing skills. *Eurasian Journal of Educational Sciences*, 82, 1-26, DOI: 10.14689/ejer.2019.82.1

Özet

Problem durumu: Dil eğitiminin temel amacı, kişilerin düşünme, iletişim, anlama ve anlatma becerilerini geliştirmektir. Bu nedenle bütün ülkeler eğitim sistemlerinde dil eğitimine büyük önem verirler (Kavcar, 1983). Eğitimin formal boyutunun gerçekleştirildiği okullarda dil eğitiminin amacına ulaşabilmesi, öğretim programının niteliğine ve bu programın uygulanmasında işe koşulan öğrenme-öğretme sürecinin etkililiğine bağlıdır (Hamzadayı, 2010).

Türkçe Öğretim Programı'nın (MEB, 2015) öğrencilerin çevreleriyle iletişim kurmaları, iş birliği yapmaları, ortak karar vermeleri ve karşılaştıkları sorunları çözmeleri; yazılı metinleri doğru ve akıcı bir biçimde okuyabilmeleri, okuduklarını değerlendirip okumayı bir alışkanlık haline getirmeleri; duygu, düşünce, istek, hayal, tasarı ve izlenimleri ile bir konu hakkındaki görüşlerini dilin imkânlarından yararlanarak yazılı anlatım kurallarına uygun şekilde anlatmaları, yazmayı alışkanlığa dönüştürmeleri ve yeteneği olanların bu becerilerini geliştirmeleri gibi amaçları vardır. Bu çalışmada temel dil becerilerinden yazma öğrenme alanının öğretim süreci açısından etkililiği ve öğrenme sürecine olan etkileri ele alınmıştır.

Türkçe Öğretim Programı'nda (MEB, 2015) yazma becerisinin geliştirilmesiyle, öğrencilerin duygu, düşünce, hayal, tasarı ve izlenimlerini dilin imkânlarından yararlanarak ve yazılı anlatım kurallarına uyarak anlatmaları, yazmayı kendini ifade etmede bir alışkanlık haline getirmeleri ve yazmaya yeteneği olanların da bu becerilerini geliştirmeleri amaçlanmaktadır. Bir ifade etme becerisi olan yazma becerisi, konuşma becerisi gibi doğal bir potansiyelle gelmeyen, eğitim yoluyla öğrenilebilen (Özdemir, 2016) bir beceri olması nedeniyle anadili Türkçe olan öğrencilerin bile en çok zorlandıkları dil becerisidir. Anadili Türkçe olmayan öğrenciler için buna bir de anadilinin farklı olması faktörü eklenince yazma becerisini kazanmak daha zor olmaktadır. Eğitim sistemleri içinde bulunduğu toplumun bireylerine ulusal dilin bütün inceliklerini öğreterek, onların yaşamlarında dili yetersiz kullanmalarından kaynaklanan sorunlar yaşamalarına meydan vermemek olmalıdır (Sarı, 2001).

Türkiye'de çeşitli faktörlerin yazma becerisi üzerindeki etkilerinin farklı gruplar üzerinde incelendiği pek çok araştırma yapılmıştır. Ancak anadili Türkçe olmayan

ilkokul öğrencilerinin yazma becerilerini geliştirmeye yönelik yapılan bir çalışmaya rastlanmamıştır.

Alan yazın incelendiğinde öğrencilerin yazma becerisine yönelik kazanımlara ulaşabilmelerine yönelik olarak yansıtma materyalleri, yapılandırmacı yaklaşımın öğrenciden beklentileri ile uyumlu görülmektedir. Öğrencilerin Türkçe başarısı ve yazma becerisinin geliştirilmesi için öğrencilerin kendi öğrenmelerini takip etmeleri ve öğrendiklerini yansıtmalarına fırsat vermesi açısından ders günlüğü yazma uygun bir etkinlik olarak görülmüştür.

Araştırmanın Amacı: Bu araştırmanın amacı, dördüncü sınıf Türkçe dersinde ders günlüğü kullanımının iki dilli öğrencilerin akademik başarıları ve yazma becerileri üzerindeki etkisini belirlemektir.

Araştırmanın Yöntemi: Bu çalışmada tek grup ön test-son test deneysel desen kullanılmıştır. Araştırmanın çalışma grubunu kolay ulaşılabilirlik esasına göre belirlenen, araştırmacının görev yaptığı okulda okuttuğu 4.sınıf öğrencileri oluşturmaktadır. Çalışma grubundaki öğrencilerin uygulamaların başında ve uygulama sonunda Türkçe dersine ilişkin akademik yeterliklerini ve yazma becerilerini belirlemeye yönelik olarak uygulamalar yapılmıştır. Araştırma verileri, öğrencilerin Türkçe dersindeki akademik başarılarını belirlemek amacıyla araştırmacı tarafından hazırlanan Türkçe Dersi Akademik Başarı Testi ve öğrencilerin yazma becerilerini ölçmek amacıyla öğrencilerin yazdığı yazılar ve bu yazıları değerlendirmek için araştırmacı tarafından geliştirilen Yazma Becerileri Dereceli Puanlama Anahtarı ile toplanmıştır. Araştırmanın uygulama sürecinde, on üç hafta boyunca öğrencilere her ders sonunda derste öğrendiklerini yansıttıkları ders günlüğü yazdırılmıştır. Yazılan ders günlüklerine ilişkin gerekli dönüt ve düzeltmeler yapılmıştır. Araştırma verileri, Wilcoxon İşaretli Sıralar Testi, Mann-Whitney U ve Spearman Sıra Farkları Korelasyon Katsayısı testleri ile analiz edilmiştir.

Araştırma Bulguları: İki dilli öğrencilerin başarı testi ön puanları ile son puanları ile yazma becerileri ön ve son test puanları arasında yüksek düzeyde bir ilişki olduğu saptanmıştır. Buna göre, öğrencilerin duygu, düşünce ve izlenimlerini yazarak ifade etmelerinin başarılarını olumlu yönde etkileyebileceği söylenebilir. Öğrencilerin başarı testinden aldıkları ön ve son test puanları arasında son test puanları lehine anlamlı bir fark olduğu sonucuna ulaşılmıştır. Bu farkın sebebi öğrencilerin kendi öğrenme sürecini ders günlüğü yazarak kontrol altında tutmaları, konuyu tekrar etmiş olmaları, öğrenmelerinin öğretmenlerinin dönüt ve düzeltmelerle desteklenmesi olabilir. Öğrencilerin yazma becerileri ön ve son ölçümlerden aldıkları puanlar arasında son test puanları lehine anlamlı bir fark olduğu sonucuna ulaşılmıştır. Ulaşılan bu sonuç, ders günlüklerinin yazmayı yazarak öğrenmeye fırsat vermesi ile ilişkilendirilebilir. Çünkü yazma, teoride öğrenilecek bir beceri değildir. Öğrenciler, nekadar çok yazarsa ve yazılarında bu yazma kurallarını nekadar çok uygularsa o ölçüde yazma becerilerini kazanabilecektir. Öğrencilerin başarı testinden aldıkları ön ve son test puanları ile yazma becerilerinden aldıkları ön ve son test puanlarının cinsiyet açısından anlamlı bir farklılık göstermediği sonucuna ulaşılmıştır. Bunun bulgunun sebebi, hem kız hem de erkek öğrencilerin bulunduğu sosyal çevrenin,

öğrenim gördükleri okulun, yetiştirilme tarzının aynı olması sebebiyle gerek genel kültürel beklenti ve aktivitelerin gerekse de ilgilendikleri doğal aktiviteler ve çocuklar büyütülürken yapılan bakım uygulamalarının benzer olması olabilir (Çapri ve Çelikkaleli, 2005).

Araştırma Sonucu ve Öneriler: Bu çalışmada ders günlüğü yazma etkinliğinin iki dilli öğrencilerin akademik başarılarını arttırdığı ve yazma becerilerini geliştirdiği sonucuna ulaşılmıştır. Ayrıca öğrencilerin akademik başarılarının ve yazma becerilerinin cinsiyete göre değişkenlik göstermediği saptanmıştır.

Ders günlüğü kullanılarak yazmanın öğrencilerin okuma ve sözlü iletişim becerileri üzerindeki etkisini belirlemeye yönelik veya daha güçlü bir değişken olabilmesi için ön test-son test kontrol gruplu deneysel desende başka çalışmalar yapılabilir. Farklı derslerde ve sınıf seviyelerinde ders günlüğü tutmanın etkililiği araştırılabilir. Bu araştırma, küçük bir çalışma grubuyla tek bir sınıfta gerçekleştirilmiştir. Benzer bir araştırma daha büyük bir çalışma grubuyla tüm dördüncü sınıf Türkçe öğretimi için gerçekleştirilebilir.

Anahtar Kelimeler: Türkçe dersi akademik başarı testi, yazma becerileri dereceli puanlama anahtarı, yansıtma materyalleri

Eurasian Journal of Educational Research 82 (2019) 27-44

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

A Comparison of Kernel Equating Methods Based on Neat Design*

Cigdem AKIN ARIKAN[1]

ARTICLE INFO

Article History:
Received: 12 Nov. 2018
Received in revised form: 22 Apr. 2019
Accepted: 29 May. 2019
DOI: 10.14689/ejer.2019.82.2

Keywords
Equating, equipercentile, linear, RMSD SEED, SEE

ABSTRACT

Problem Statement: Equating can be defined as a statistical process that allows modifying the differences between test forms with similar content and difficulty so that the scores obtained from these forms can be used interchangeably. In the literature, there are many equating methods, one of which is Kernel equating. Trends in International Mathematics and Science Study (TIMSS) aims to find out the knowledge and skills gained by the fourth and eighth-grade students in the fields of mathematics and science. TIMSS have different test forms, and these forms are equated through common items.

Purpose of the Study: This research aimed to compare the equated score results of the Kernel equating (KE) methods, which are chained, and post-stratification equipercentile and linear equating methods under NEAT design.

Methodology: TIMMS Science data were used in this study. The study sample consisted of 865 eighth-grade examinees who were given the Booklets 1 and 14 during the TIMSS application in Turkey. There were 39 items in Booklet 1, and 38 items in Booklet 14. Firstly, descriptive statistics were calculated and then the two Booklets were equated according to NEAT design based on Kernel chained, Kernel post-stratification equipercentile, and linear equating methods. Secondly, the equating methods were evaluated according to some criteria such as DTM, PRE, SEE, SEED, and RMSD.

Findings and Results: It was seen that results based on equipercentile and linear equating methods were consistent with each other, except for a high range of the score scale. PRE values demonstrated that KE equipercentile equating methods better matched with the discrete target distribution Y, and distribution of SEED revealed that KE equipercentile and linear methods were not significantly different from each other according to DTM.

* This study was presented at the 6th International Congress on Measurement and Evaluation in Education and Psychology in Prizren-Kosova, 5-8 September 2018.

[1] Ordu University, Faculty of Education, Ordu, TURKEY, e-mail: akincgdm@gmail.com, ORCID: https://orcid.org/ 0000-0001-5255-8792

Introduction

Equating can be defined as a statistical process that allows modifying the differences between test forms with similar content and difficulty so that the scores obtained from these forms can be used interchangeably (Kolen, 1988). For about 100 years, equating methods have attracted the attention of psychometrics and the development of new methods has not stopped. Equating methods include methods based on equipercentile equating, linear equating methods, IRT observed-score and true score equating, van der Linden local equating, Levine nonlinear method, and Kernel equating (von Davier, 2013). As for the Kernel equating, an observed-score equating method was defined by Holland and Thayer (1989) and then improved by von Davier, Holland and Thayer (2004). In traditional equipercentile equating methods, cut-off score distribution is made continuous by using linear estimates. On the other hand, Kernel equating employs the Gaussian Kernel approach after which it is also named. In the latter, discrete distributions are made continuous so that scores are equated on the basis of the continuous distributions (Lee & von Davier, 2011, Ricker & von Davier, 2007). KE is a flexible family of equipercentile-like equating functions that include the linear equating function as a special case (von Davier, Holland & Thayer 2004).

In the Kernel equating model, test forms are equated in five steps: presmoothing, estimation of score probabilities, continuization, equating, and standard error of equating. The first step is presmoothing that refers to using the log-linear statistical model for smoothing of score distributions. The goal of presmoothing is to achieve decreased sampling errors. In this step, the estimation of score probabilities varies depending on the score equating design. Equivalent groups design is a univariate distribution; however, common-item test design is a bivariate distribution in nonequivalent groups. Von Davier et al. (2004) indicated four statistical properties in the selection of estimating point probabilities as;

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

The second step is the estimation of score probabilities of X and Y scores according to the equating design that is obtained from step one. The third step is continuation where Gaussian Kernel approach is used to make the cut-off score distributions continuous at the relevant stage. In this step, the choice of bandwidths is essential. Von Davier et al. (2004) suggest the penalty function to automatically select the bandwidths. In addition to Gaussian Kernel approach, Lee and von Davier (2011) recommend logistics and uniform kernel approaches as alternatives. The fourth step

is equating. When the first three steps are done, test forms are equated by using continuous distributions. The last step is the standard error of equating (SEE). SEE is dependent on presmoothing, computing r and s from the smoothed data and equating function (von Davier et al., 2006).

Kernel equating can be used in single-group, equivalent groups, and non-equivalent groups (von Davier et al., 2004). Non-Equivalent groups Anchor Test-NEAT is used when the test form is applied more than once due to test safety. In NEAT design, both forms have common items and equating the relationship between the test forms is established through common items (Kolen & Brennan, 2004). In Kernel equating in NEAT pattern; Post- stratification (PSE), Levine observed-score linear, and Chained Equating (CE) methods are used (von Davier et al., 2004). In NEAT pattern, two different groups take two different test forms (X and Y) and the common test form (A). PSE uses the common test form to estimate the distribution of test forms across a group I and group II. In CE, the common test is used as a chain and the test form X is first connected to the common test form for group I. Then the common test form is connected to the version Y for group II (von Davier et al., 2004). Kernel equating includes both linear and equipercentile equating functions by manipulating bandwidths. If optimal bandwidths are selected, KE approximates the equipercentile equating function, and if large bandwidths are selected, KE approximates the linear equating (von Davier et al., 2006). The equating methods used in this paper are given in Table 1.

Table 1

Equating Methods

Linear	PSE-with large bandwidths
	CE –with large bandwidths
Equipercentile	PSE-with optimal bandwidths
	CE- with optimal bandwidths

One of the criteria for determining which method performs better in equating is the error. The equating method with a smaller rate of error can be said to be more appropriate. Furthermore, KE provides some measures, percent-relative error (PRE) and standard error of equating difference (SEED) when evaluating the equating results. PRE is a tool that assesses how well an equating function matches the discrete target distribution Y. SEED can be defined as a difference between the two equating functions and the range of ±2 SEED shows that the differences are because of sampling variability (Liu & Low, 2007). The equating methods are evaluated according to certain criteria: Difference That Matter (DTM), PRE, standard error of equating (SEE), SEED, and Root Mean Squared Difference (RMSD).

DTM: DTM is used to evaluate the difference between equated scores obtained from two distinct equating functions. Despite not being an established rule, it is generally determined to be .5, which is half of the raw point unit (1). If the difference between the two equated scores is less than .5, the scores are regarded similar; if the difference is bigger than .5, the equated scores are considered distinct (Holland & Dorans, 2006).

PRE: The percent-relative error (PRE) is a tool that compares the distribution of Y with the equated values, eY(X) and assesses how well an equating function matches the discrete target distribution Y (Von Davier et al., 2004). The PRE is calculated by the following formula.

$$PRE(p) = 100\frac{\mu_p\,(eY\,(X)) - \mu p(Y\,)}{\mu p(Y\,)} \qquad (1)$$

KE compares the first 10^{th} moment of Y and eY(X). If continuization step has been done cautiously, then the PRE values are frequently small (von Davier et al., 2004).

SEE: In Kernel equating, standard error of equating depends on three factors. The first is the combination of pre-smoothing, the second is the computation of smoothed data, and the third is the mathematical form of the smoothing and equating function (von Davier et al., 2006).

SEED: It is used to determine the accuracy of the difference between the two equating functions and suggest which synchronization function is more appropriate. SEED is also used to choose either linear or non-linear equating functions (Von Davier et al., 2004). Furthermore, ±2 SEED band is available in order to determine how the two equating functions vary depending on sample variability (Von Davier et al., 2004). If the variance between equating functions does not exceed the ± 2 SEED range, this means that the variance is due to sampling error (Liu & Low, 2007).

$$SEED_Y(x) = \sqrt{Var\left(\hat{e}_1(x) - \hat{e}_2(x)\right)} \qquad (2)$$

RMSD: Equating error is used to define the accuracy of equating. RMSD coefficient is used for the equating error.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{k-1} fi(X_E - X_{Crit})^2}{\sum_{i=1}^{k} f_i}} \qquad (3)$$

Xcrit: The raw score number i in test D

XE : The score obtained with equating methods and equal to the raw score number i in test X

fi: The frequency of the raw score number i in test D

Purpose of the Study

International tests applied in Turkey include TIMMS (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment), and PIRLS (Progress in International Reading Literacy Study). TIMSS is a test held every four years since 1995, but Turkey did not participate in 1995 and 2003. TIMSS aims to find out the knowledge and skills gained by the fourth and eighth-grade students in the fields of mathematics and science (MEB, 2016). TIMSS 2015 Turkey test consists of 24 blocks with 14 test booklets for science and mathematics. The 24 blocks were placed in 14 test booklets, two in science and two in mathematics, and one of two blocks in science and mathematics is common to two of the booklets (MEB, 2016). To compare trends between the years, TIMSS assessments were converted into the same metrics. For that, TIMMS uses item response theory (IRT) scaling with concurrent calibration (Mullis, Martin & Foy, 2016). However, it is of great importance which equating method is chosen. For the purpose of the test, the equating method should be determined by taking into account the strengths and weaknesses of the methods. It is needed because the choice of an inappropriate equating method increases the equating errors, leading to unfair decisions. KE methods can be used especially when IRT (true score) equating methods are not favorable (Godfrey, 2007; Meng, 2012; Norman Dvroak, 2009). TIMMS didn't use Kernel equating methods for converting the scores into the same metrics. In Turkey, several tests such as KPSS and ALES hold different validity of periods and project subjects take different test forms in Measurement and Evaluation of Academic Skills (MEAS-ABIDE). Since test forms must be equated in order to compare or use the scores interchangeably, several studies have been used Kernel equating (Choi, 2009; Grant, Zhang & Damiano, 2009; Godfrey, 2007; Holland, von Davier, Sinharay & Han, 2006; Mao, 2006; Mao, von Davier & Rupp, 2005; Meng, 2012; Moses & Holland, 2007; Norman Dvorak, 2009; Ricker & von Davier, 2007; von Davier et al., 2006). When the literature is examined, it is seen that articles about Kernel equating are very limited in Turkey (e.g. Akın Arıkan, 2017; Akın Arıkan & Gelbal, 2018). Therefore, it is thought that this study will contribute to the other studies which can use KE when the assumptions of IRT equating methods are not meet.

The main purpose of this study was to compare Kernel equating methods with real data under NEAT design based on equipercentile and linear methods so as to detect the most appropriate equating method. For this main purpose, research questions were as follows:

1) What is the relationship between raw scores and equivalent scores obtained from different equating methods?

2) How do PRE, DTM, SEE, SEED and RMSD values differ according to equating methods?

3) Which is the best Kernel equating method to equate TIMSS science subtests under NEAT design?

Method

Research Design

In this study, TIMSS 2015 science tests (Booklet number 1 and 14) were equated with Kernel equating methods and the obtained equating results were compared with each other. In terms of this, this research was a descriptive study.

Research Sample

During the period when the TIMSS 2015 research was conducted, there were a total of 1,108,572 students at the 4th grade and another 1,187,893 students at the 8th grade in Turkey. Out of the population; 6456 of 4th graders and 6079 of 8th graders participated in the TIMMS application (MEB, 2016). The study sample consisted of 865 eighth-grade examinees who were given the Booklets 1 and 14 during the TIMSS application in Turkey.

Research Instrument and Procedures

For data analysis, the data set was used consisting of the pattern of responses given by the 8th-grade examinees to science literacy items in the TIMSS 2015 Turkey. In this study, the items in Booklet number 1 and 14 were used among fourteen booklets included in the TIMMS application. There were 39 items in Booklet number 1, and 38 items in Booklet number 14. The wrong and missing values were coded as 0 and the partial credit scores and all the correct answers were coded as 1 yielding the final data for analysis.

Data Analysis

The booklets were equated according to the methods of Kernel CE and Kernel PSE. The kequate package (Andersson, Branberg & Wiberg, 2013) was used for kernel equating methods analyses (R Core Team, 2017).

Results

In the first phase of data analysis, descriptive statistics were calculated and the findings are presented in Table 2.

Table 2

Raw Score Descriptive Statistics of Booklet 1 and Booklet 14

Descriptive Statistics						
TEST	N	Mean	Std. Dvt.	Variance	Skewness	Kurtosis
K1	435	18.58	7.58	57.38	0.20	-0.63
Anchor-K1	435	5.84	3.48	12.10	0.44	-0.30
K14	430	12.74	4.65	21.60	-0.15	-0.67
Anchor-K14	430	6.17	3.71	13.78	0.42	-0.73

Table 2 shows the mean and standard deviation values for both booklets according to the total tests and anchor tests. Anchor test of booklet 14 mean scores were higher than the anchor test of Booklet 1 mean scores. Moreover, since the skewness coefficient of score distribution in Booklet 1, the common test of Booklet 1, and the common test of Booklet 14 was positive, the distribution seemed to be skewed to the right of what was normal. In addition, since the skewness coefficient of score distribution of Booklet 14 was negative, it can be said that the distribution was skewed to the left than normal. It can be suggested that the distributions had kurtosis compared to normal because the kurtosis coefficients of score distribution of both forms were negative.

The bandwidths values were automatically calculated by kequate package. The obtained values for KE PSE equipercentile (PSE EQ) method were .6327 for hX and .6318 for hY; for KE PSE linear (PSE L) method, it was 7611.23 for hX and 7306.05 for hY. As for KE CE equipercentile (CE EQ) equating, the values are .633 for hX and .6322 for hY. Finally, 7575.07 for hX and 7342.40 for hY in KE CE linear (CE L) method. Table 3 displays PRE values for KE PSE and KE CE (equipercentile and linear) equating methods.

Table 3

The PRE Values for the KE Optimal and KE Linear for Equating X to Y

Pth Moment	**Post- stratification Equating (PSE)**		**Chained Equating (CE)**			
	PRE EQ	PRE L	CE EQ		CE L	
			X to A1	A1 to Y1	X to A1	A1 to Y1
1	0.000	0.000	-0.010	0.257	0.000	0.000
2	-0.001	0.000	-0.002	-0.443	0.000	0.000
3	-0.007	-0.217	-0.108	-0.144	-8.243	1.633
4	-0.018	-0.646	-0.087	-0.165	-10.372	3.564
5	-0.035	-1.281	0.014	-0.608	-10.577	4.516
6	-0.059	-2.112	0.191	-0.751	-7.683	3.822
7	-0.092	-3.126	0.439	-0.801	-2.471	1.328
8	-0.134	-4.309	0.758	-0.820	5.244	-2.795
9	-0.188	-5.645	0.845	-0.879	15.486	-8.221
10	-0.254	-7.117	0.902	-0.970	28.531	-14.573

PRE = Percent relative error, EQ= Equipercentile

Table 3 indicates that the PRE values stated a good match for PSE and CE equipercentile equating methods but a poorer match for both KE linear equating methods between the equating function computed at the discrete values of X and the

target distribution of Y. Both equipercentile and linear equating PRE(p) values for PSE were smaller than for both CE methods, indicating good matching of the moments of the distributions. Booklet 1 and Booklet 14 were equated according to Kernel chained (EQ –L) and Kernel post-stratification (EQ-L) equating methods. Table 4 displays the results of equating method.

Table 4

Equivalent scores of Booklet 14 corresponding to raw scores of Booklet 1

Booklet 1 Raw Score	PSE EQ	PSE L	CE EQ	CE L
0	-0.16	-0.69	-0.16	-0.58
1	0.69	0.27	0.69	0.36
2	1.55	1.23	1.55	1.31
3	2.43	2.19	2.43	2.25
4	3.33	3.15	3.33	3.20
5	4.25	4.11	4.24	4.14
6	5.18	5.07	5.16	5.09
7	6.11	6.03	6.09	6.03
8	7.04	6.99	7.03	6.98
9	7.98	7.95	7.96	7.92
10	8.93	8.91	8.90	8.87
11	9.87	9.87	9.84	9.81
12	10.82	10.83	10.78	10.76
13	11.77	11.79	11.72	11.70
14	12.72	12.75	12.66	12.65
15	13.67	13.71	13.60	13.60
16	14.62	14.67	14.54	14.54
17	15.58	15.63	15.47	15.49
18	16.54	16.59	16.41	16.43
19	17.50	17.55	17.35	17.38
20	18.46	18.51	18.29	18.32
21	19.42	19.47	19.22	19.27
22	20.38	20.43	20.16	20.21
23	21.35	21.39	21.11	21.16
24	22.32	22.35	22.05	22.10
25	23.28	23.31	23.01	23.05
26	24.25	24.27	23.97	23.99
27	25.22	25.23	24.93	24.94
28	26.19	26.19	25.91	25.88
29	27.16	27.15	26.89	26.83
30	28.13	28.11	27.87	27.77
31	29.11	29.07	28.86	28.72
32	30.09	30.03	29.86	29.66
33	31.08	30.99	30.87	30.61

Table 4 Continue…

Booklet 1 Raw Score	PSE EQ	PSE L	CE EQ	CE L
34	32.08	31.95	31.89	31.55
35	33.09	32.91	32.93	32.50
36	34.14	33.87	33.99	33.44
37	35.23	34.83	35.11	34.39
38	36.40	35.79	36.31	35.33
39	37.67	36.75	37.62	36.28

Table 4 showed that the raw scores from Booklet 1 got values from 0 to 39, but the results of PSE EQ equating showed that equivalent scores of Booklet 14 got points between -0.16 and 37.67, PSE L equating showed the values of -0.16 to 36.75, CE EQ equating yielded values from -0.16 to 37.62 and CE L equating showed values between -0.58 and 36.28. All raw scores of Booklet 1 were greater than Booklet 14 equivalent scores. This implies that Booklet 1 was easier than Booklet 14 throughout the score scale and there was a linear relationship between the raw scores and equivalent scores. Figure 1 and Figure 2 show the differences between the equivalent scores obtained according to the equating methods. Differences KE PSE EQ and KE PSE L and differences between KE CE EQ and KE CE linear are shown in Figure 1.

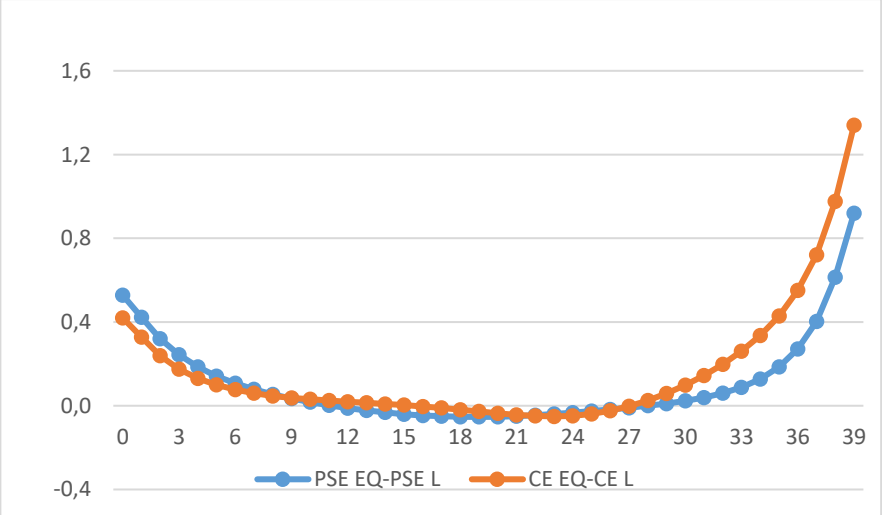


Figure 1. *Differences between KE PSE EQ and KE PSE L and differences between KE CE-EQ and KE CE L*

Figure 1 shows the raw-to-raw equating differences between KE PSE equipercentile and KE PSE linear and differences between KE CE- equipercentile and

KE CE linear, respectively. The results indicated that KE PSE equipercentile produced very similar results to KE PSE linear, except a high range of the score scale. KE CE equipercentile produced very similar results to KE CE linear, except between the scores of 36 and 39. The differences between KE PSE equating methods were smaller than DTM below 38 raw score points and the differences between KE CE equating methods were smaller than DTM below the raw score point of 36. Differences between KE PSE EQ and KE CE EQ and between KE PSE linear and KE CE linear are shown in Figure 2.

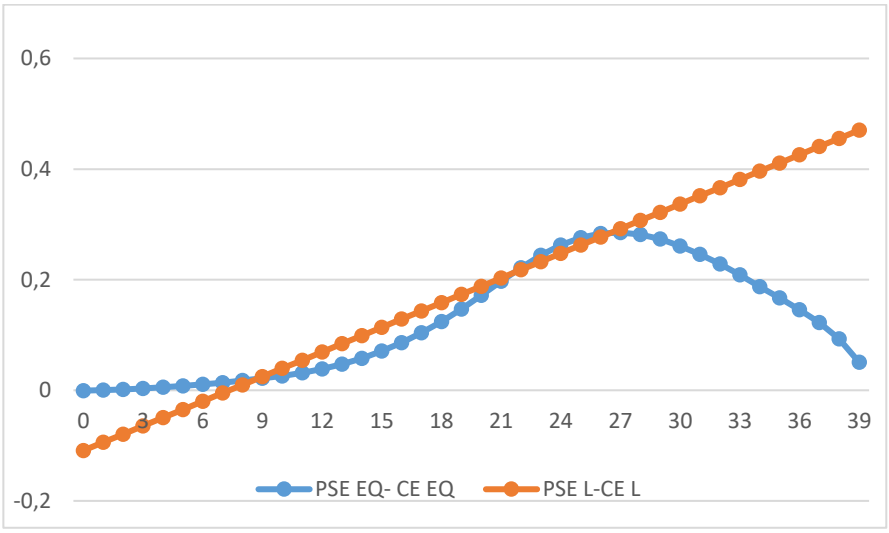


Figure 2. *Differences KE PSE EQ and KE CE EQ and differences between KE PSE L and KE CE L*

Figure 2 shows the raw-to-raw equating differences between KE PSE and KE CE equipercentile methods and differences between KE PSE and KE CE linear methods, respectively. The results indicated that KE PSE equipercentile method produced very similar results to KE CE equipercentile and KE PSE linear produced very similar results to KE CE linear. The differences between all equating methods were smaller than DTM. Figure 3 shows the values of the SEE obtained for each raw point from Kernel equipercentile and Kernel linear equating methods. The mean SEE values were found as .511 for KE PSE equipercentile; .573 for KE PSE linear; .526 for KE CE equipercentile, and .598 for KE CE linear methods.

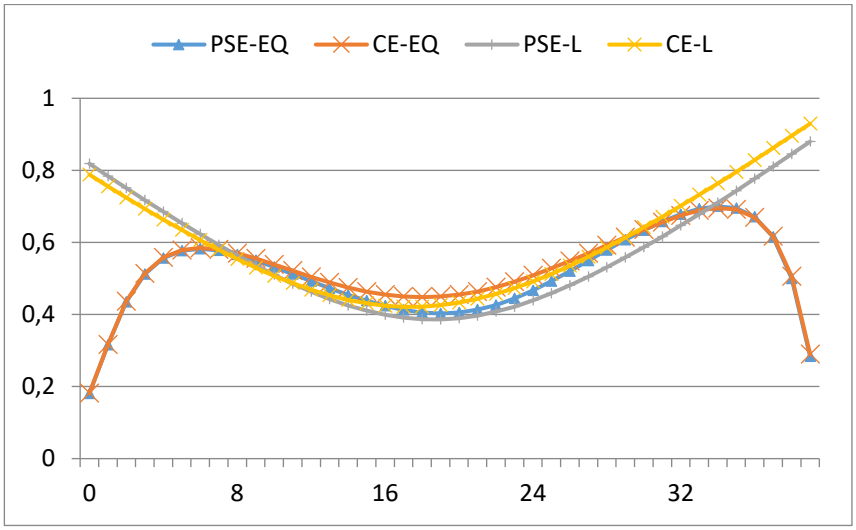


Figure 3. *SEE for each equating methods*

Figure 3 reveals that the standard error values were close to each other in the middle of the raw score scale (range of 8-32 points). On the other hand, at extreme points, Kernel equipercentile equating methods showed lower levels of standard errors while linear equating methods had higher standard errors. The SEE values for both equipercentile equating methods were nearly the same and the SEE values for both linear equating methods were close to each other. When we compared all equating methods, PSE method has a slightly smaller SEE for the middle of the raw score scale. SEED values between KE PSE EQ and KE PSE L, and between KE CE EQ and KE CE L were shown in Figure 4 and Figure 5.

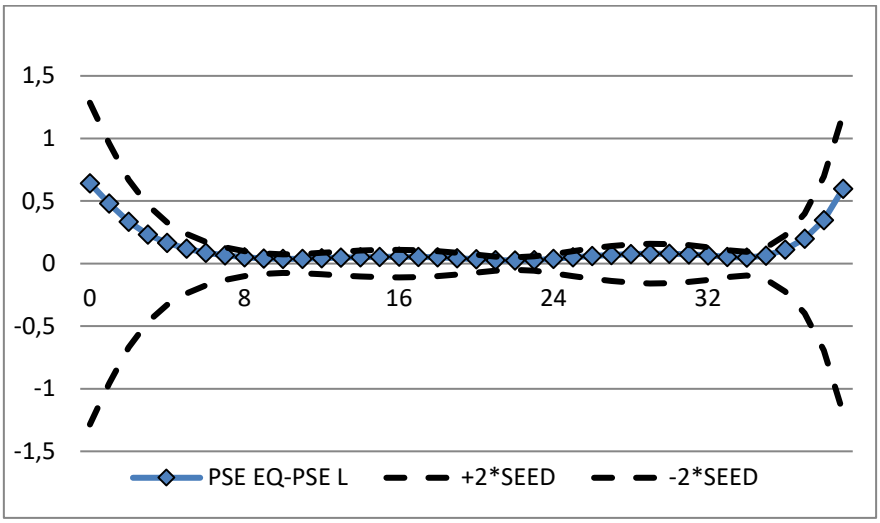


Figure 4. *SEED for equating methods: KE PSE EQ versus KE PSE L*

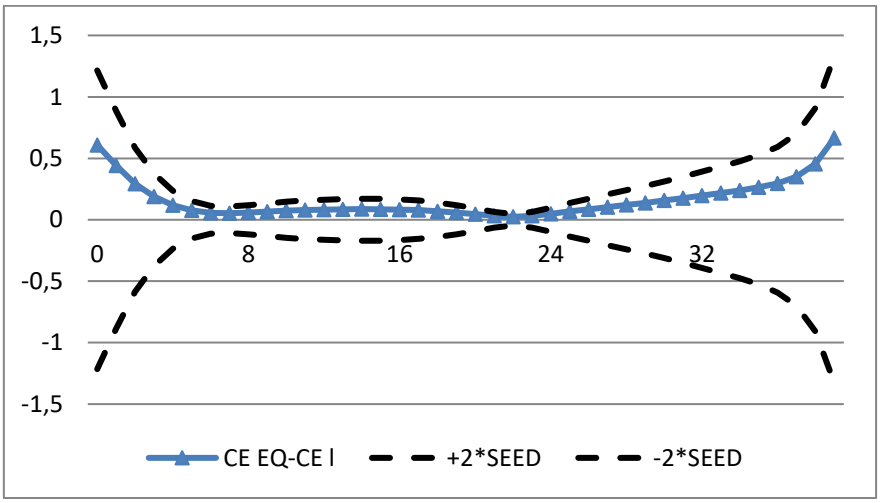


Figure 5. *SEED for equating methods: KE CE EQ versus KE CE L*

Figure 4 and 5 plot the differences and the SEED between PSE EQ and PSE linear, CE EQ and CE linear functions, respectively. Both plots indicated that the line was above the zero line for all score scale. EQ methods equated higher converted scores than linear equating methods, that is to say, EQ methods measured new test X as being harder than linear methods did (i.e., the X form was harder than the Y form). The

differences were smaller than one DTM, expected at the lower end and the higher end of the score scale. Moreover, the difference between the equating functions lied within ±2 SEED across the entire score range, in other words, EQ and linear functions were not significantly different from each other. RMSD coefficient was calculated to evaluate the random involved in the equating methods. The resulting coefficients are given in Table 5.

Table 5

The RMSD Values for Equating Methods

Equating methods	RMSD
PSE-EQ	2.044
PSE-L	2.043
CE-EQ	2.483
CE-L	2.528

It was seen in Table 4 that the equal RMSD coefficients existed in scores equated with KE PSE equipercentile and linear equating methods. The smallest RMSD (2.044 and 2.043) coefficients were obtained from scores equated with PSE method, while the largest RMSD coefficients were obtained through KE CE linear equating method. It can be inferred that whereas the least random error was yielded by KE PSE method, the maximum random error was given by chained linear equating method.

Discussion, Conclusion, and Recommendations

In this study, two Booklets (Booklet 1 and 14) used in TIMMS 2015 science test were equated by using the methods of KE PSE linear, KE PSE equipercentile, KE CE equipercentile and KE CE linear equating methods, and the resulting PRE, SEE, SEED and RMSD values were compared. When reviewing PRE values of Kernel PSE equipercentile and PSE linear, Kernel CE equipercentile and linear; PRE values demonstrate that equipercentile equating methods exhibit lower values than linear equating methods. In other words, it better matches the discrete target distribution Y. Distribution of SEED reveals that the difference between the equating functions lies within ±2 SEED across the entire score range. To put in another way, EQ and linear functions are not significantly different from each other. When the raw-to-raw equating differences between equating methods were examined, the results indicated that KE equipercentile seemed to produce very similar results to KE linear, except the high range of the score scale, and differences between KE PSE and KE CE equating methods were smaller than DTM, except the high range of the score scale. Comparison of the RMSD coefficients based on KE PSE and CE equating methods implies that post-stratification equating method offers the least random error, whereas chained linear equating method yields the maximum random error rates.

When Kernel equating methods were compared against mean SEE, linear equating methods had slightly higher than equipercentile methods. This finding seems incompliant with the findings of Choi (2009) and Liou, Cheng and Johnson (1997). In

his study, Choi (2009) compared the variables of sample size, test length, bandwidth, and presmoothing parameter with Kernel equating and traditional equating methods. He found out that linear Kernel equating methods yield lower standard errors than equipercentile methods. Apart from that, Liou, et al. (1997) found out that the Gaussian Kernel method reduces the standard error with wide bandwidth. While the same study revealed that selection of the parameter h decreases the standard error values, our study found out that the parameter h increased slightly the mean standard error. This difference may be due to the use of simulation data or large sample size in other studies. It was also found out that the KE linear equating methods yielded higher standard error rates at extreme points than the average scores. The results seem to be in conformity with findings of Mao (2006) and Mao, von Davier and Rupp (2006). The latter explained the higher standard errors at extreme values in Kernel equating methods with the use of the Gaussian Kernel method for the continuization of the cumulative score distribution. In the Gaussian Kernel continuization method, the score scale ranges from + ∞ to – ∞ and this leads to arising of increased mean error rates from extreme scores. When the RMSD coefficients obtained based on the KE, PSE, and CE equating methods were compared, the method with the least random error was found to be the post- stratification equating method, while the method with the most random errors was the chained linear equating method.

In this study, Booklets 1 and 14 in the TIMMS 2015 science test were equated in the NEAT design by using Kernel equating methods. A similar study can be carried out by means of equating methods based on the Item Response Theory and the Classical Test Theory, and the results can be compared to the results of this study. A similar study can also be performed for different subtests.

References

Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software, 55*(6), 1-25.

Akin-Arikan, Ç. (2017). *Kernel Eşitleme ve Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması* [Comparison of kernel equating and item response theory equating methods] (Unpublished doctoral dissertation). Hacettepe University, Ankara, Turkey.

Akin-Arikan, Ç. & Gelbal, S. (2018). A Comparison of Traditional and Kernel Equating Methods. *International Journal of Assessment Tools in Education, 5*(3), 417-427. doi: 10.21449/ijate.409826

Choi, S. I. (2009). *A Comparison of Kernel Equating and Traditional Equipercentile Equating Methods and the Parametric Bootstrap Methods for Estimating Standard Errors in Equipercentile Equating* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, United States.

Godfrey, K. E. (2007). *A comparison of Kernel equating and IRT true score equating methods* (Unpublished doctoral dissertation). The University of North Carolina, United States.

Grant, M. C., Zhang, L., & Damiano, M. (2009). An Evaluation of Kernel Equating: Parallel Equating with Classical Methods in the SAT Subject Tests [TM] Program. (ETS RR-09-06). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (pp. 187–220). Westport, CT: Praeger Publishers.

Holland, P., von Davier, A., Sinharay, S., & Han, N. (2006). *Testing the untestable assumptions of the chain and post-stratification equating methods for the NEAT design* (ETS RR-06-17). Princeton, NJ: Educational Testing Service.

Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice, 7*(4), 29-37. doi: 10.1111/j.1745-3992.1988.tb00843.x

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

Lee, Y. H., & von Davier, A. A. (2010). Equating through alternative kernels. In A.A. von Davier (Ed.) *Statistical models for test equating, scaling, and linking* (pp. 159-173). Springer New York.

Liou, M., Cheng, P. E., & Johnson, E. G. (1997). Standard errors of the Kernel equating methods under the common-item design. *Applied Psychological Measurement, 21*(4), 349-369. doi: 10.1177/01466216970214005

Liu, J., & Low, A. C. (2007). An Exploration of Kernel Equating Using SAT® Data: Equating to a Similar Population and to a Distant Population. (ETS RR-07-17). Princeton, NJ: Educational Testing Service.

MEB (2016). TIMSS 2015 Uluslararası Matematik ve Fen Eğilimleri Araştırması: TIMMS 2015 Ulusal Matematik ve Fen Bilimleri On Raporu 4. ve 8. Sınıflar. MEB Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. Ankara. [Çevrim-içi http://timss.meb.gov.tr/wp-content/uploads/TIMSS_2015_Ulusal_Rapor.pdf] Erişim Tarihi: 15 Mart 2018.

Meng, Y. (2012). *Comparison of Kernel Equating and Item Response Theory Equating Methods* (Unpublished doctoral dissertation). University of Massachusetts Amherst, United States.

Mao, X. (2006). *An investigation of the accuracy of the estimates of standard errors for the Kernel equating functions* (Unpublished doctoral dissertation). University of Iowa, Iowa City, United States.

Mao, X., von Davier, A. A., & Rupp, S. (2006). Comparisons of the Kernel equating method with the traditional equating methods on PRAXISTM data (ETS RR-06-30). Princeton, NJ: Educational Testing Service.

Moses, T., & Holland, P. (2007). Kernel and traditional equipercentile equating with degrees of presmoothing (ETS RR-07-15). Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., Cotter, K. E., Centurino, V. A. S., Fishbein, B. G., & Liu, J. (2016). Using Scale Anchoring to Interpret the TIMSS 2015 Achievement Scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 14.1-14.47). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-14.html

Norman Dvorak, R. K. (2009). *A comparison of Kernel equating to the test characteristic curve methods* (Unpublished doctoral dissertation). University of Nebraska, Lincoln, United States.

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

Ricker, K., & von Davier, A. A. (2007). *The impact of anchor test length on equating results in a nonequivalent groups design* (ETS RR-07-44). Princeton, NJ: Educational Testing Service.

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

von Davier, A. A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). *An evaluation of the Kernel equating method. A special study with pseudotests constructed from real test data* (ETS RR-06-02). Princeton, NJ: Educational Testing Service.

Kernel Eşitleme Yöntemlerinin Denk Olmayan Gruplarda Ortak Madde Test Deseninde Karşılaştırılması

Atıf:

Akin-Arikan, C. (2019). A comparison of kernel equating methods based on neat design. *Eurasian Journal of Educational Research,* 82, 27-44, DOI: 10.14689/ejer.2019.82.2

Özet

Problem Durumu: Eşitleme benzer içerik ve güçlük düzeyinde geliştirilen test formları arasındaki farklılıkları düzenleyerek, bu formlardan elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olarak tanımlanabilir (Kolen, 1988). Test eşitleme yöntemleri yaklaşık 100 yıldır psikometristlerin dikkatini çekmekte ve yeni yöntemler geliştirilmektedir. Eşitleme yöntemleri eşit yüzdelikli eşitlemeye dayalı yöntemler, doğrusal eşitleme yöntemleri, MTK gözlenen ve gerçek puan eşitleme, van der Linden yerel eşitleme, Levine doğrusal olmayan metot ve yeni bir yaklaşım olan Kernel eşitlemeyi kapsar (von Davier, 2013). Tek grup, eşdeğer grup ve denk olamayan gruplarda ortak madde test deseninde kullanılır (von Davier et al., 2004). Denk olmayan gruplarda ortak madde deseni (Non-Equivalent groups Anchor Test-NEAT), test güvenliği nedeniyle test formunun birden daha fazla uygulandığı durumlarda kullanılır. NEAT deseninde, her iki formda ortak maddeler yer alır ve test formları arasındaki eşitleme ilişkisi de ortak maddeler üzerinden kurulur (Kolen ve Brennan, 2014). Kernel eşitleme doğrusal ve eşit yüzdelikli eşitleme yöntemlerini içerir. NEAT deseninde zincirleme eştileme (doğrusal ve eşit yüzdelikli), son tabakalama (eşit yüzdelikli ve doğrusal), Levine gözlenen puan doğrusal eşitleme yöntemleri bulunmaktadır. Yeni bir yaklaşım olan Kernel eşitleme yöntemlerinin geleneksel eşitle yöntemleri ve Madde Tepki Kuramı eştilem yöntemleri ile karşılaştırıldığı çalışmalar bulunmakdadır. Bu çalışmanın amacı ise, Türkiye'nin de yer aldığı TIMMS fen datasındaki Kernel eşitleme yöntemlerine göre eşitlenmesidir.
Araştırmanın Amacı: Bu araştırmanın amacı, TIMMS fen datasındaki 1. Ve 14. Kitapçıklarının Kernel eşitleme yöntemlerinden zincirleme ve son tabakalama eşitleme yöntemlerine göre eşitlenerek, en iyi eşitleme yönteminin belirlenmesidir.

Araştırmanın Yöntemi: TIMSS 2015 araştırmasının yapıldığı dönemde Türkiye'de toplam 1.108.572 4. sınıf öğrencisi, 1.187.893 de 8. sınıf öğrencisi bulunmaktadır. 6456, 4. sınıf öğrencisi ve 6079, 8. Sınıf öğrencisi TIMMS uygulamasına katılmıştır. Araştırmanın örneklemini ise Türkiye'deki TIMMS uygulamasına katılan 8 sınıf öğrenciler arasından, bu uygulama esnasında 1. ve 14. kitapçıkları alan 865 öğrenci oluşturmaktadır. Veri analizi için TIMMS 2015 uygulanmasına katılan Türkiye'deki 8. sınıf öğrencilerin fen okuryazarlığı maddelerine verdiği cevap örüntülerinden oluşan veri setinden yararlanılmıştır. Bu çalışmada TIMMS uygulamasında yer alan 14 kitapçıktan 1 ve 14 nolu kitapçıklarda yer alan maddeler kullanılmıştır. 4 nolu

kitapçıkta 39, 14 nolu kitapçıkta 38 madde yer almaktadır. Yanlış ve kayıp veriler 0 ve kısmi puanlanan ve doğru cevapların hepsi 1 olarak kodlanarak analiz edilecek veri hazırlanmıştır. Verilerin analizinin birinci aşamasında, Kernel zincirleme ve Kernel son tabakalama eşit yüzdelikli ve doğrusal eşitleme yöntemlerine göre kitapçıklar eşitlenmiştir. Daha sonra eşitleme yöntemleri DTM, PRE, SEE, SEED ve RMSD kriterlerine göre değerlendirilmiştir.

Araştırmanın Bulguları: Kernel zincirleme eşit yüzdelikli, zincirleme doğrusal, son tabakalama doğrusal ve son tabakalama eşit yüzdelikli eşitleme yöntemlerine göre kitapçıklar eştilendiğind eilk olarak PRE değerleri elde edilmiştir. KE zincirleem eşit yüzdelikli ve son tabakalama eşit yüzdelikli eşitleme yöntemlerine datanın daha iyi uyum sağladığı elde edilmiştir. Eşitleme yöntemleri karşılaştırıldığında, eşit yüzdelikli eşitleme yöntemlerinin ve doğrusal eşitleme yöntemlerinin birbiriyle benzer sonuçlar ürettiği ve aralarındaki farkın DTM'den küçük olduğu elde edilmiştir. Eşitleme yöntemlerine göre SEE değerleri karşılaştırıldığında, orta puan ölçeğinde bu değerlerin birbirlerine yakın olduğu görülmektedir. Uç puanlarda ise Kernel eşit yüzdelikli eşitleme yöntemleri düşük, doğrusal eşitleme yöntemleri ise yüksek standart hatalara sahip olduğu elde edilmiştir. Eşitleme yöntemlerine göre SEED değerleri karşılaştırıldığında, eşitleme yöntemleri arasındaki farkın DTM'den küçük olduğu ve ±2 SEED çizgisi arasında kaldığı bulunmuştur. Eşitleme yöntemlerine karışan random hatayı değerlendirebilmek için RMSD katsayısı hesaplanmıştır. En az random hata içeren eşitleme yöntem son tabakalama eşitleme yönteminde iken en fazla random hata içeren yöntemin zincirleme doğrusal eşitleme yönteminde olduğu elde edilmiştir.

Araştırmanın Sonuçları ve Önerileri: Kernel eşitleme yöntemleri ortalama SEE açısından karşılaştırıldığında, doğrusal eşitleme yöntemlerinin eşit yüzdelikli yöntemlere göre daha yüksek ortalama SEE sahip olduğu bulunmuştur. Bu bulgu Choi (2009) ve Liou ve diğerlerinin (1997) bulgularıyla tutarlı olmadığı görülmektedir. Elde edilen bu sonuç diğer çalışmalarda simülasyon data veya geniş örneklem büyüklüğünün kullanılmasından kaynaklı olabilir. Ayrıca KE doğrusal eşitleme yöntemlerinde uç puanlarda orta puanlara göre daha yüksek standart hata verdiği bulunmuştur. Bu bulgu literatürdeki çalışmaları desteklemektedir. RMSD katsayıları karşılaştırıldığında en az random hata içeren yöntem son tabakalama eşitleme yöntemi iken en fazla random hata içeren yöntemin zincirleme doğrusal eşitleme olduğu görülmüştür. Elde edilen bu sonuçlardan hareketle, gelecek çalışmalarda farklı kriterler kullanılarak farklı eşitleme yöntemleri kullanılabilir ve bu çalışmanın sonuçlarıyla karşılaştırılabilir.

Anahtar Sözcükler: eşitleme, eşit yüzdelikli, doğrusal, SEED, SEE

Eurasian Journal of Educational Research 82 (2019) 45-62

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

Evaluation Model for Evaluating Vocational Skills Programs on Local Content Curriculum in Indonesia: Impact of Educational System in Indonesia

Samsul HADI[1], Dedek ANDRIAN[2], Badrun KARTOWAGIRAN[3]

ARTICLE INFO

Article History:
Received: 24 Feb. 2019
Received in revised form: 25 Jun.2019
Accepted: 13 Jul.2019
DOI: 10.14689/ejer.2019.82.3

Keywords
evaluation model, vocational skills programs, local content curriculum

ABSTRACT

Purpose: The impact of decentralization on education systems requires the local government to develop local content curriculum to preserve regional characteristics and local uniqueness through education. Practice in the field revealed problems in terms of implementation so that the development of an evaluation model in accordance with the characteristics of the region is necessary. The purpose of this study was to produce an evaluation model for evaluating vocational skills programs on the local content curriculum.

Method: This was a developmental study consisting of three phases: 1) The initial investigative phase was conducted by interviewing vice-principals and vocational skills programs teachers;2) The design phase was carried out by making an instrument, evaluation guide, evaluation procedure and the effectiveness of the model assessment sheet; 3) The trial phases were carried out by performing readability test, small-scale trial, and large-scale trial.
Results: The results of the readability test indicated that evaluation instrument, evaluation guide, and evaluation procedure were feasible to use, and according to experts the model was very effective to evaluate vocational skills programs on the local content curriculum. The result of the small-scale trial showed that the 24 items developed were valid and reliable. Large-scale trial results indicated that the instrument had acceptable validity and reliability of the construct.
Implications for Research and Practice: Evaluation models that have been developed, validated by experts, validated by practitioners and trials on the small and large-scale can contribute to improving vocational skills programs on the local content curriculum developed by local governments throughout Indonesia and even other countries.

[1] Corresponding Author, Postgraduate Program, Yogyakarta State University, Email: samsul_hd@uny.ac.id ORCID: https://orcid.org/0000-0003-3437-2542

[2] Mathematics Education , Riau Islamic University, Email: dedekandrian@edu.uir.ac.id .ORCID:https://orcid.org/ 0000-0001-5192-2222

[3] [3]Postgraduate program, Yogyakarta State University, Email: : kartowagiran@uny.ac.id ORCID: https://orcid.org/0000-0002-8536-5417

Introduction

Efforts to conserve natural resources or local uniqueness have been thought by the central government. The central government has established an organized system in a school curriculum that is called the local content curriculum. This is stipulated in the regulation of the Minister of Education and Culture No. 22 of 2006 in content standards stating that local content and self-development activities are an integral part of the curriculum structure at elementary and secondary education levels. This is based on the facts that the Unitary State of Republic of Indonesia consists of various regions of diverse geographical conditions, natural resources, and communities with different historical and cultural backgrounds (Depdiknas, 2006).

The central government has made a policy to accommodate the needs of local governments in efforts to develop natural resources or uniqueness of the region. The policy is run through education contained in the curriculum that is packaged in learning activities. Government policies facilitate regions to develop natural resources or potential areas into a product that can improve the economy of the society. Local content becomes an effective means to preserve and develop regional potentials. Local content curriculum to bridge learners to adopt the potential of the region so that through the bridge the learner can develop and preserve the environment into something useful. The introduction of local culture through teaching and classroom activities can essentially help students to recognize their cultural identity (Prastiwi, 2013). The local content curriculum makes learners able to generate useful creativity in improving the economy if the learners do not continue to university education.

The importance of the role of local content implemented in schools is undeniable. Local content is very important in maintaining the local wisdom of a region (Saputra, 2013). Local content can help the government in realizing the local vision to enrich students with the local cultural knowledge and improve students' ability to socialize it to other areas (Prastiwi, 2013). Therefore, the activity of local content learning that runs continuously in a learning system is expected to contribute overall to make the formation of students' character who love local culture.

Many studies, which were inspired by Gardner's AMTB, were conducted in maintaining the regional uniqueness and preserving the local culture, local government from Riau Province developed a local content curriculum by implementing vocational skills programs. Vocational skills programs are expected to be a weapon and an effective way to preserve the uniqueness and culture through educational programs. Vocational programs are the main programs to introduce regional and cultural characteristics to learners. Learners as a young generation are expected to maintain uniqueness and develop a regional culture with the existence of a vocational skills program. Vocational skills programs are expected to contribute greatly to preserving regional uniqueness and culture. Vocational skills program on the local content curriculum is expected to be a program that can develop the uniqueness of the region and culture so that it can improve the economy of students, schools, and regions.

The curriculum is a sum of the learning activities and experiences that a student has under the auspices or direction of the school (Fink, C.R .& Crunkilton, 1999; Howell, K.W. & Nolet, 2000). The curriculum is an educational process undertaken from determining the process, methods, teaching strategies, means, and infrastructure, where the educational function is taught (Sorenson, Goldsmith, Mendez& Maxwell, 1979). This curriculum refers to planned learning experiences that the educational institution intends to provide for its learners (Uys & Gwele, 2005). The curriculum is a learning experience planned by educational institutions that are provided for learners (Syomwene, Kitainge & Mwaka, 2013). The curriculum is an important part of the education system because the curriculum is a framework that expresses goals, expectations, and means (Suryadi, Ekayanti, & Euis, 2017). The curriculum is total means through students by school guidance.

Curriculum evaluation is an integral part of curriculum development activities. The developed curriculum needs to be evaluated to see the achievement of the objectives of the curriculum. However, curriculum evaluation is not only used to obtain information about achievement but also as a measure of the value or effectiveness of any particular activity in education. The curriculum evaluation refers to the collection of information in which judgment might be made about the worth and the effectiveness of a particular program (Hussain, Dogar, Azeem & Shakoor, 2011). Curriculum evaluation refers to the process of studying the merit or worth of some aspects, or the whole curriculum (Bharvad, 2010). Curriculum evaluation has a very important role in gaining information about the achievement of the developed curriculum. Curriculum evaluation is a systematic way of examining all components of a curriculum whose results are evaluative inferences (Uys & Gwele, 2005). Curriculum evaluation is an important aspect that can answer fundamental questions about the achievement of the purpose of developing a curriculum (Akker & Verloop, 1994). From the description of curriculum evaluation, it can be concluded that evaluators or stakeholders can check the success of the curriculum through basic questions with reference to the objectives of the developed curriculum. Curriculum evaluation has an important role in detecting the success of learning because the problems of learning results that are not achieved are often addressed in the curriculum. Curriculum evaluation has an important role in identifying issues occurring in the curriculum and ensuring the quality of schools and influencing the success of the curriculum (Yeung, 2010). Curriculum evaluation is a phase of selecting information, obtaining, analyzing, transferring, using and making decisions to improve the quality of the curriculum (Hakan & Seval, 2011).

The National Research Council (2004) states that three framework components must be determined to evaluate the curriculum; (1) program materials and principles in curriculum design, (2) quality of curriculum implementation, (3) curriculum impact on student achievement. These three components can show how effectively the curriculum has been developed. The evaluation of the curriculum is to analyze the process and impact of activities running in the school curriculum (Figueiredo, Leite & Fernandes, 2016). Through curriculum evaluation, the understanding, practice of

teaching, learning and assessment can be improved (Leathwood& Phillips, 2000). This shows that the curriculum is the basis of the success of teaching and learning in class.

An understanding of the problem in learning can be developed by investigating or evaluating the curriculum being applied (Haghparast, Sedghizadeh, Shuler, Ferati & Christersson, 2007). Significant attention to improving the curriculum is something that must be given at the level of state, nation and region. Concrete steps should be taken by evaluating the curriculum. Interest in evaluating the curriculum should grow as a major effort to improve the curriculum. curriculum evaluation provides illustrations of problems experienced in learning activities, then schools can create a formula or problem-based model to solve the problem (Harlacher, Sakelaris & Kattelman, 2014). Schools can explicitly explore approaches that can be used to improve the applied curriculum (Adin-Surkis, 2016). Curriculum evaluation is an important point to show how far students have achieved the competence of the learning process (Merritt, Blake, McIntyre & Packer, 2012). Based on this it can be concluded that the good or bad standard of a curriculum can be seen by evaluation. The gaps in curriculum implementation can be seen clearly by conducting curriculum evaluation.

Vocational skills programs are the impacts of local content curriculum implementation developed by the local government. Vocational skills programs become an effective means of improving the economy. Vocational skills programs applied through the education system can improve the economy of society (Akshay et al., 2013). Vocational skills programs that are still in continuous training can improve the economic productivity of a region (Powell, Bernhard & Graf, 2012). Vocational skills program in the application can improve communication skill, interpersonal skills, group work, and problem-solving skills of the students (Hassall, Joyce, Montaño, & Anes, 2005; Montano, Donoso, Hassall & Joyce, 2001). Vocational skills programs taught sustainably throughout education can develop the ability to plan, encourage sustainable work practices and develop strategies in the workforce (Brown, 2013)

Based on the definition of experts it can be concluded that the evaluation of vocational skills programs on the local curriculum is an important thing to do. The success of the curriculum can be observed through curriculum evaluation. The curriculum evaluation will examine in detail the factors that may affect the success of students' vocational skills in the local curriculum. Evaluation of a vocational skills program in the local content curriculum can provide a comprehensive picture of government-developed curriculum achievement. Vocational skills programs in the local content curriculum should be thoroughly evaluated. This is because the vocational skills programs in the local content curriculum are developed based on the objectives, background, policy and needs analysis equipped with infrastructure and supporting elements to achieve the objectives. The components underlying the vocational skills in the local curriculum need to be evaluated so that known weaknesses could be corrected.

Method

Research Design

Evaluation model for evaluating the implementation of vocational skills programs in the local content curriculum in Indonesia: Impact The education system in Indonesia uses a research and development model from Borg and Gall (1983) consisting of 10 steps simplified into three steps: (1) initial investigation phase, (2) design phase, and (3) trials phase. The initial investigation was conducted with a qualitative research approach that is phenomenology. Data were collected (FGD) with 16 Participants consisting of 8 vice-principals and 8 vocational skills programs teachers. The purpose of FGD was to find problems about the implementation of vocational skills programs on the local content curriculum and to find the constructs used to develop the instrument. From of the FGD four constructs have obtained that need to be evaluated in terms of the implementation of vocational skills programs on the local content curriculum namely, learning facilities, teacher preparation, learning implementation, and assessment. The design phase was conducted by developing a questionnaire and an interview guide. The small-scale trial aims to see the validity of the contents empirically to the items of the questionnaire that has been developed. The large-scale trial aims to see the validity and reliability of the construct of the questionnaire that has been developed

Research Sample

The sample used in this research is the vice principal, vocational skill program teachers, and senior high school students applying a vocational skills program on the local content curriculum. The initial investigative phase involved 8 vice-principals and 8 vocational skills programs, teachers. The small-scale trial phase involved 180 senior high school students in Riau Province. The large-scale trial phase involving 300 senior high school students. The sampling technique used is purposive sampling because it can select the samples correctly according to the criteria and the number of samples to be used for the research (Eğmir, Erdem & Koçyiğit, 2017).

Research Instruments and Procedures

The instrument used in this study is a questionnaire and an interview guide. The questionnaire and interview guide that has been developed is then validated by experts and practitioners. The questionnaire developed amounted to 24 items consisting of 6 items of Learning facilities, 5 items of teacher preparation, 7 items of curriculum implementation, 6 items of assessment. The interview guide developed amounted to 12 questions. The development of the instrument was conducted through the results of the initial investigation.

Data Analysis

There were three phases of data analysis performed in this research. Data analysis of the initial investigation phase, data analysis of small-scale trial data and data analysis of the large-scale trial. Data analysis at the initial investigative phase used the method proposed by Miles & Huberman (1994): data reduction, data display, and

conclusion. Data analysis at the small-scale trial phase was held using exploratory factor analysis to see the validity of content empirically. The data analysis of the large-scale trial phase also used confirmatory factor analysis to see the validity of the constructs.

Results

Design Phase

The design phase was conducted by developing a questionnaire and an interview guide. Questionnaire and interview guide were assessed by experts and practitioners. The expert and practitioners' assessment can be seen at the readability test as follows:

Readability Test

The readability test aimed to obtain an assessment or validation from the evaluation experts, measurement experts, vice principals, vocational skill programs teachers about the quality of the instrument, evaluation guide, evaluation procedure and effectiveness of the developed model. The readability test was done by 2 evaluation experts, 2 measurement experts, 4 vice principals, and 4 vocational skills programs teachers. The results of the readability test were analyzed using descriptive analysis and comparing with the criteria of product validity proposed by Sultan, et al. (2017). The product validation criteria table can be seen in Table 1.

Table 1

Criteria for Product Validity

Average Score	Category
3.26-4.00	Very feasible
2.51-3.25	Feasible
1.76-2.50	Feasible enough
1.00-1.75	Not feasible

Readability Test of the Questionnaire

The Readability test of the Questionnaire was performed to see how far the instruments are feasible to be used to evaluate vocational skills programs on the local content curriculum. There are seven aspects that are assessed in the instrument of questionnaire that is, 1) clarity of instructions to use the Instruments, 2) communicative language use, 3) the choice of words, 4) the structure of sentences, 5) multi-interpretation, 6) evaluation of local content curriculum at high schools in Riau Province, and 7) usability of the instruments. The result of an expert and practitioner assessment can be seen in Table 2.

Table 2

Readability Test of Questionnaire

Aspects assessed	Min	Max	Mean	Standard deviation	Category
Aspect 1	3.00	4.00	3.750	0.452	Very feasible
Aspect 2	3.00	4.00	3.667	0.492	Very feasible
Aspect 3	2.00	4.00	3.333	0.651	Very feasible
Aspect 4	3.00	4.00	3.417	0.515	Very feasible
Aspect 5	3.00	4.00	3.500	0.522	Very feasible
Aspect 6	3.00	4.00	3.667	0.492	Very feasible
Aspect 7	3.00	4.00	3.750	0.452	Very feasible

Based on Table 2 it can be concluded that according to experts and practitioners the developed instruments can be used to evaluate vocational skills programs on the local content curriculum of the senior high school in Riau Province because the lowest average score was 3.33 with a very feasible category.

Readability Test of the Interview Guidance

The readability test on interview guides was done to see how far the instruments of interview guidance can be used by education offices, principals, and teachers. The Readability test was also assessed by experts and practitioners. There are seven aspects that are assessed in the instrument of the interview guidance that are, 1) clarity of instructions to use the instruments, 2) communicative language, 3) the choice of words, 4) the structure of sentences, 5) multi-interpretation, 6) evaluation of the local content curriculum at high schools in Riau Province, and 7) usability of the instruments. The results of the expert and practitioner assessment can be seen in Table 3.

Table 3

Readability Test of Interview Guidance

Aspects assessed	Min	Max	Mean	Standard deviation	Category
Aspect 1	3.00	4.00	3.58	0.51	Very feasible
Aspect 2	2.00	4.00	3.25	0.75	Feasible
Aspect 3	3.00	4.00	3.58	0.51	Very feasible
Aspect 4	2.00	4.00	3.33	0.65	Very feasible
Aspect 5	2.00	4.00	3.50	0.67	Very feasible
Aspect 6	2.00	4.00	3.50	0.67	Very feasible
Aspect 7	2.00	4.00	2.92	0.79	Feasible

Based on Table 3 it can be concluded that according to experts and practitioners the developed instruments in the form of interview guidance can be used to evaluate vocational skills programs on the local content curriculum of the senior high schools in Riau Province, because the lowest average score was 2.92 with feasible category.

Readability Test of the Evaluation Guide

Assessment of evaluation guidance helped to see how far the developed evaluation guide can be comprehended and used to evaluate vocational skills programs on the local content curriculum. There are several important points to consider in terms of evaluation guide that are, 1) evaluation model of vocational skill program can be used to diagnose weakness or deficiency in vocational skills program, 2) user of this model is principal, vice-principal, and teacher, 3) model evaluation will evaluate important components in vocational program, 4) this model is used after school exams, 5) Instrument in this evaluation model uses likert scale and interview guide, 6) Students are required to fill in their names and class in available column, 7) student are asked to read the instructions in the instrument carefully, and 8) the time provided to answer the statement is one hour. The assessment was done by experts and practitioners. There are nine aspects assessed on the evaluation guide namely; 1) the instructions in the evaluation guide, 2) the steps in the evaluation, 3) how to use the evaluation model, 4) clarity of time to carry out an evaluation, 5) communicative language use, 6) the structure of sentences, 7) the choice of words, 8) multi-interpretation, and 9) ease of the evaluation guide. The results of the assessment can be seen in Table 4.

Table 4

Results of Assessment of Expert &Practitioners

Aspects assessed	Min	Max	Mean	Standard deviation	Category
Aspect 1	2.00	4.00	3.333	0.778	Very feasible
Aspect 2	2.00	4.00	2.833	0.835	Feasible
Aspect 3	2.00	4.00	3.083	0.793	Feasible
Aspect 4	2.00	4.00	2.917	0.669	Feasible
Aspect 5	2.00	4.00	3.000	0.603	Feasible
Aspect 6	2.00	4.00	3.083	0.900	Feasible
Aspect 7	2.00	4.00	3.000	0.953	Feasible
Aspect 8	2.00	4.00	3.250	0.754	Feasible
Aspect 9	2.00	4.00	3.250	0.754	Feasible

Based on the assessment of experts and practitioners it can be concluded that the evaluation guide is feasible to be used to evaluate vocational skill programs on the local content curriculum at high schools in Riau Province because all the aspects in the guide were in feasible and very feasible category.

The Readability Test of the Evaluation Procedure

Assessment of the evaluation procedure revealed how far the developed procedure can be carried out practically without any difficulties. There are five procedures that must be performed by user evaluation that are, 1) Evaluation is done after 1st semester, 2) Vocational skills program teachers provide tests to see the competence of the students with the instrument that has been designed by the teacher 3) the score of the test results are calculated and compared with the predetermined success criteria, 4) the teacher reports the results of the vocational skills program test subjects to the students, 5) the teacher collects students outside the lesson hours and provides the evaluation instrument of the vocational program to evaluate the overall vocational skills program which has been running for one semester. Assessment of evaluation procedures is done by experts and practitioners. There are seven aspects assessed on evaluation procedure, namely; 1) clarity of evaluation procedures, 2) Effective sentence use, 3) correct use of spelling and punctuation, 4) practicality of the evaluation procedure (easy to follow), 5) Efficiency in terms of time, 6) efficiency in terms of cost, and 7) efficiency in terms of energy. The results of the assessment of experts and practitioners can be seen in Table 5.

Table 5

The Results of the Assessment of Experts and Practitioners

Aspects assessed	Min	Max	Mean	Standard deviation	Category
Aspect 1	2.00	4.00	3.417	0.669	Very feasible
Aspect 2	2.00	4.00	3.500	0.798	Very feasible
Aspect 3	2.00	4.00	2.667	0.778	Feasible
Aspect 4	2.00	4.00	3.000	0.953	Feasible
Aspect 5	2.00	4.00	2.750	0.866	Feasible
Aspect 6	2.00	4.00	2.917	0.669	Feasible
Aspect 7	2.00	4.00	3.083	0.900	Feasible

Based on the assessment of experts and practitioners it can be concluded that the developed evaluation procedure was feasible to be used to evaluate vocational skills program on the local content curriculum because all aspects were considered to be in the feasible and very feasible category.

Assessment of the Effectiveness of the Developed Model

Testing the effectiveness of the model helped to see whether the developed model could be implemented easily and effectively to evaluate vocational skill programs on the local content curriculum. The test of the effectiveness of the model was held by experts and practitioners. There are six aspects assessed on the effectiveness of the

developed model, namely; 1) comprehensiveness of the developed model, 2) the accuracy of the model to evaluate local content curriculum, 3) The accuracy of the instrument type, 4) guidance in the implementation of the locals, 5) Ease of use of the instrument, and 6) the evaluation report. The effectiveness test results can be seen in Table 6.

Table 6

The Results of Effectiveness Test

Aspects assessed	Min	Max	Mean	Standard deviation	Category
Aspect 1	2.00	4.00	3.417	0.669	Very Good
Aspect 2	2.00	4.00	3.083	0.669	Good
Aspect 3	2.00	4.00	2.917	0.793	Good
Aspect 4	2.00	4.00	2.667	0.778	Good
Aspect 5	2.00	4.00	2.583	0.669	Good
Aspect 6	2.00	4.00	3.000	0.739	Good

Based on the results of effectiveness test assessed by experts and practitioners, it can be concluded that evaluation model developed was effective and easy to use to evaluate vocational skills programs on the local content curriculum. All aspects were considered by experts and practitioners in the good and very good category.

Trial Phases

Small-Scale Trial

The small-scale trial was analyzed using exploratory factor analysis with varimax rotation method. The analysis results obtained by Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was 0.907 (KMO> 0.5) and Bartlett's Test of Sphericity was 0.00 (α <0.05). The value indicates that the sample used was feasible to analyze using factor analysis. Furthermore, we can see that the loading factor was obtained by varimax rotation method. The result of exploratory factor analysis with varimax rotation method obtained four factors. The result of the analysis can be seen in Table 7.

Table 7

The Results of Exploratory Factor Analysis With Varimax Rotation Method

Item	Component			
	Factor 1	Factor 2	Factor 3	Factor 4
Learning Facilities 1				0.850
Learning Facilities 2				0.786
Learning Facilities 3				0.683
Learning Facilities 4				0.775
Learning Facilities 5				0.622
Learning Facilities 6				0.773
Teacher Preparation 1			0.918	
Teacher Preparation 2			0.870	
Teacher Preparation 3			0.844	
Teacher Preparation 4			0.837	
Teacher Preparation 5			0.908	
Curriculum Implementation 1	0.880			
Curriculum Implementation 2	0.872			
Curriculum implementation 3	0.866			
curriculum Implementation 4	0.782			
curriculum Implementation 5	0.849			
curriculum Implementation 6	0.790			
curriculum Implementation 7	0.860			
Assessment 1		0.911		
Assessment 2		0.844		
Assessment 3		0.855		
Assessment 4		0.739		
Assessment 5		0.831		
Assessment 6		0.847		
Total of Eigen Values	10.303	3.418	3.040	1.811
Percentage of variance explained	42.930	14.241	12.666	7.548
KMO				.907
Bartlett's test of sphericity*				4179.312
df				276
sig				0.000

a. Rotation converged in 5 iterations.

Based on Table 6, it can be concluded that 24 items that have been rotated by the varimax rotation formed into 4 factors. The first factor was the implementation of the curriculum, the second factor was the assessment, the third factor was the preparation of teachers, and the fourth factor was learning facilities. The four factors formed had a loading factor value of more than 0.5, so it can be concluded that the items that have been developed were valid and fit to be used to evaluate vocational skills programs on the local content curriculum. (Hair Jr, William, Babin& Anderson, 2014).

Reliability of Instrument

Reliability was tested using Cronbach's Alpha. This result indicated that the instrument was reliable because it has a value of Cronbach's Alfa 0.940. The result can be seen in Table 8.

Table 8

Reliability Statistics

Cronbach's Alpha	N of Items
0.940	24

Large-Scale Trial

The large-scale trial was utilized to find the validity and reliability of constructs obtained by the evaluation component that became the basis to evaluate the vocational skill programs on the local content curriculum. This construct should be empirically tested and analyzed using statistical confirmatory factor analysis. The purpose of this analysis was to determine the validity and reliability of constructs obtained by initial investigations with qualitative research approaches. The results of the analysis can be seen in Table 8.

Table 9

Validity and Reliability of Construct

Factor	Indicator	λ	Error	CR
Factor 1	Curriculum Implementation	0.856	0.267	0,884
Factor 2	Assessment	0.793	0.371	
Factor 3	Teacher Preparation	0.830	0.311	
Factor 4	Learning Facilities	0.756	0.428	

The results of confirmatory factor analysis showed that four factors had a factor load value of more than 0.5, so it can be concluded that four factors had validity and reliability of the constructs in good categories. The results of confirmatory factor

analysis showed that the four-factor met the goodness of fit statistics because they had Chi-Square value <2df, P-value> 0.05, and RMSEA <0.08, GFI> 0.9 and AGFI> 0.9. In other words, the model developed in accordance with the data that has been obtained in the field.

Discussion, Conclusion, and Recommendations

The readability test shows that the developed instruments in the form of questionnaires and interview guides are in a category worthy of use to evaluate vocational skills programs in local content curricula. All the aspects used to assess the instruments are in a decent and highly feasible category. The readability test of the instruments is an important part to consider in developing the model because the instruments will be used by education offices, principals, and teachers who have never used the instruments before to evaluate vocational programs. Education offices, principals, and teachers who have understood the contents of developed instruments can provide explanations to students as subject and object evaluation models when they implement evaluation models. Instruments have validated by experts can give good data or information about the effectiveness of local content curriculum (Andrian, Kartowagiran, & Hadi, 2018). The evaluation guide is a guide to the education office, principals, and teachers as users of the evaluation model that has been developed. The guide for evaluating vocational skills programs in the local curriculum should be thoroughly understood by the users of the vocational skills program evaluation model. The evaluation guide can provide an ease to the users in implementing the evaluation model directly in the field. High understanding of the evaluation guide on the program evaluation model can avoid errors in the process of data collection.

Evaluation procedures are the steps that the user of the vocational skills program evaluation model should follow. The steps in the evaluation need to be done carefully so that the users of the evaluation model of the vocational skill program are not confused in the implementation of the evaluation model that has been developed. The users must read the procedure in order not to make mistakes in implementing the evaluation model of vocational skill program. Procedures can be a compass for education offices, principals and teachers and evaluation agencies in evaluating vocational programs. Assessment of model effectiveness is an important part of developing an evaluation model of vocational skills programs in the local content curriculum. An effective evaluation model will make it easier for users to evaluate vocational skills programs in terms of fund use, time use and physical capacity utilization, as good evaluation models do not incriminate users to evaluate vocational programs. An effective evaluation model can make it easier for users to get data in the field and process the data into a conclusion that can improve vocational skill program that runs in Riau Province.

Based on the research results, it can be concluded that the evaluation model for evaluating vocational skill programs on local content curriculum that has been developed is supported by a valid and reliable instrument. These results suggest that

evaluation models can help improve the implementation of vocational skills programs in the local content curriculum. Accurate information obtained from valid and reliable instruments can give you an idea of what needs to be fixed. Ungar & Santos (2003) state that the instruments having content and construct validity can acquire valuable information. Valid and reliable instruments are important for evaluating educational programs including the curriculum. Loiacono, Watson & Goodhue (2007) state that instruments that have the power of validity can acquire deep and wide information. Information about the strengths and weaknesses of educational programs can be detected. Validity and reliability are required in instrument development. Valid and reliable instruments can accurately represent the results of research (Burton & Mazerolle, 2011). Valid and reliable instruments can describe the quality of a measurement (Bayraktar, Tatoglu & Zaim, 2008). Trials need to be taken seriously, in order to obtain accurate data.

Conclusion

The evaluation model developed consists of; evaluation instrument, evaluation procedure, evaluation guide, and evaluation effectiveness model sheet. The results of the assessment indicated that the instrument, evaluation guide, and evaluation procedure were feasible to be used to evaluate vocational skills programs on the local content curriculum. According to experts and practitioners, the evaluation model is very effectively used for evaluating vocational skills programs on the local content curriculum. The results of the small-scale trial obtained a value of the loading factor greater than 0.5 or valid, and Cronbach's Alpha value was greater than 0.7 or reliable. The construct instruments that have been developed have validity, reliability, and model fit that can be accepted. All developed factors have a loading factor value greater than 0.5 and a reliability of construct coefficient above 0.7. Value of Chi-Square <2 df, P-Value value greater than 0.5, RSMEA value <0.08, GFI value greater than 0.9. and the AGFI value is greater than 0.9.

Recommendations

Evaluation of vocational skills programs on local content curriculum should be carried out continuously or periodically so that weaknesses and deficiencies can be detected immediately. Local governments, principals, and teachers should collaborate in implementing an evaluation model so that the goal of preserving regional uniqueness and local culture through vocational skills learning can be achieved maximally.

References

Andrian, D., Kartowagiran, B., &Hadi, S. (2018). The instrument development to evaluate local curriculum in Indonesia. *International Journal of Instruction*, 11(4), 921-934. https://doi.org/10.12973/iji.2018.11458a

Adin-Surkis, A. (2016). Teachers evaluate the new curriculum in English: Views regarding evaluation and evaluation tools. *Research in Education*, (93). https://doi.org/http://dx.doi.org/10.7227/RIE.0009

Akker, J.V.D., & Verloop, N. (1994). Evaluation approaches and results in curriculum research and development in The Netherlands. Studies in Educational Evaluation, 20(4), 421–436. doi:10.1016/0191-491x(94)00035-f

Akshay, N., Deepu, S., Rahul, E. S., Ranjith, R., Jose, J., Unnikrishnan, R., & Bhavani, R. R. (2013). Design and evaluation of a haptic simulator for vocational skill Training and Assessment. *IECON Proceedings (Industrial Electronics Conference)*, 6108–6113. https://doi.org/10.1109/IECON.2013.6700139

Bayraktar, E., Tatoglu, E., & Zaim, S. (2008). An instrument for measuring the critical factors of TQM in Turkish higher education. *Total Quality Management & Business Excellence,* 19(6), 551–574. doi:10.1080/14783360802023921

Bharvad, A. J. (2010). Curriculum Evaluation, I(12), 2009–2011.

Borg, W R & Gall, M. D. (1983). *Educational Research Forth Edition*. New York: Longman.

Brown, M., Sack, F., & Piper Rodd, C. (2013). Student voice in "skills for sustainability": A missing component from the demand side of Australian vocational education and training. *International Journal of Training Research*, 11(3), 213–224. doi:10.5172/ijtr.2013.11.3.213

Burton, L. J., & Mazerolle, S. M. (2011). Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research. *Journal of Athletic Training Education*, 6(1), 27–35.

Council, N. R. (2004). On evaluating curricular effectiveness judging the quality. Washington: National Academies Press.

Depdiknas. (2006). Regulation of the Minister of National Education of the Republic of Indonesia Number 22 of 2006. *Appendix About Standard Content*, 1-48

Eğmir, E., Erdem, C., &Koçyiğit, M. (2017). Trends in educational research: A content analysis of the studies published in international journal of instruction. *International Journal of Instruction*, 10(3), 277–294. https://doi.org/10.12973/iji.2017.10318a

Figueiredo, C., Leite, C., & Fernandes, P. (2016). The curriculum in school external evaluation frameworks in Portugal and England. *Research in Comparative and International Education*, 11(3), 282–297. https://doi.org/10.1177/1745499916661933

Fink, C.R., & Crunkilton, J. (1999). *Curriculum development in vocational and technical education. Planning, content, and implementation*. Boston: Allyn and Bacon Inc.

Haghparast, N., Sedghizadeh, P. P., Shuler, C. F., Ferati, D., &Christersson, C. (2007). Evaluation of student and faculty perceptions of the PBL curriculum at two dental schools from a student perspective: a cross-sectional survey. *European Journal of Dental Education*, 11(1), 14–22. https://doi.org/10.1111/j.1600-0579.2007.00423.x

Hair Jr, J. F., William, C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. Upper Saddle River: Pearson

Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: development, reliability and validity. *Procedia - Social and Behavioral Sciences*, 15, 592–599. doi:10.1016/j.sbspro.2011.03.146

Hassall, T., Joyce, J., Arquero Montaño, J. L., & Donoso Anes, J. A. (2005). Priorities for the development of vocational skills in management accountants: *A European perspective. Accounting Forum*, 29(4), 379–394. doi:10.1016/j.accfor.2005.03.002

Howell, K.W.,&Nolet, V. (2000). *Curriculum based evaluation: teaching and decision making*. Belmont: Wadswortth Thomson Learning.

Hussain, A., Dogar, A. H., Azeem, M., & Shakoor, A. (2011). Evaluation of Curriculum Development Process. *International Journal of Humanities and Social* Science, 1(14), 263–271.

Leathwood, C., & Phillips, D. (2000). Developing curriculum evaluation research in higher education : Procesa, politics and practicalities. *Higher Education*, 40(3), 313–330. https://doi.org/10.1023/a:1004183527173

Loiacono, E., Watson, R., & Goodhue, D. (2007). WebQual: An Instrument for Consumer Evaluation of Web Sites. *International Journal of Electronic Commerce*, 11(3), 51–87. https://doi.org/10.2753/JEC1086-4415110302

Merritt, B. K., Blake, A. I., McIntyre, A. H., & Packer, T. L. (2012). Curriculum evaluation: Linking curriculum objectives to essential competencies. *Canadian Journal of Occupational Therapy*, 79(3), 175–180. https://doi.org/10.2182/cjot.2012.79.3.7

Miles, M. B., & Huberman, M. a. (1994). Qualitative data analysis: An expanded sourcebook. *Evaluation and Program Planning*. https://doi.org/10.1016/0149-7189(96)88232-2

Montano, J. L. A., Donoso, J. A., Hassall, T., & Joyce, J. (2001). Vocational skills in the accounting professional profile: the Chartered Institute of Management Accountants (CIMA) employers' opinion. *Accounting Education*, 10(3), 299–313. https://doi.org/10.1080/09639280210122339

Powell, J. J. W., Bernhard, N., & Graf, L. (2012). The emergent European model in skill formation: Comparing higher education and vocational training in the

Bologna and Copenhagen processes. *Sociology of Education*, 85(3), 240–258. https://doi.org/10.1177/0038040711427313

Prastiwi, Y. (2013). Transmitting Local Cultural Knowledge through English as Foreign Language (EFL) Learning as a Means of Fostering "Unity in Diversity." *Academic Journal of Interdisciplinary Studies*, 2(3), 507–514. https://doi.org/10.5901/ajis.2013.v2n3p507

Saputra, G. A. S. (2013). Enhancing Local Wisdom Through Local Content Of Elementary School in Java, Indonesia. *Proceeding Of the Glbal Summit on Education*, 2013(March), 11–12.

Sorenson, R., Goldsmith, Ll., Mendez, Z., & Maxwell, K. (1979*). Curriculum development in vocational and technical education*. London: SAGE Company.

Sultan, Rofiuddin, A., Nurhadi, & Priyatni, E. T. (2017). The development of a critical reading learning model to promote university students' critical awareness. *New Educational Review*, 48(2004), 76–86. https://doi.org/10.15804/tner.2017.48.2.06

Suryadi, B., Ekayanti, F., & Euis, A. (2017). An integrated curriculum at an islamic university: perceptions of students and lecturers, *Eurasian Journal of Educational Research*, 67(January), 161–182. https://doi.org/10.14689/ejer.2018.74.2

Syomwene, A., Kitainge, K., &Mwaka, M. (2013). Psychological influences in the curriculum decision making process. *Journal of Education and Practice*, 4(8), 173–181. https://doi.org/ISSN 2222-1735.

The National Research Council (2004). *On evaluating curricular effectiveness judging the quality of k-12 mathematics evaluations*. Washington: The National Academies Press.

Ungar, W. J., & Santos, M. T. (2003). The pediatric quality appraisal questionnaire: An instrument for evaluation of the pediatric health economics literature. *Value in Health*, 6(5), 584–594. https://doi.org/10.1046/j.1524-4733.2003.65253.x

Uys, L. R., & Gwele, N. S. (2005). *Curriculum development in nursing: process and innovation*, South Africa: Psychology Press.

Van den Akker, J., & Verloop, N. (1994). Evaluation approaches and results in curriculum research and development in The Netherlands. *Studies in Educational Evaluation,* 20(4), 421–436. doi:10.1016/0191-491x(94)00035-f

Yeung, S. S. (2010). Using School Evaluation Policy to Effect Curriculum Change? A Reflection on the SSE and ESR Exercise in Hong Kong, *Educational Research Journal* 25(2), 187-209

Eurasian Journal of Educational Research 82 (2019) 63-80

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

Using Rank-order Judgments Scaling to Determine Students' Evaluation Preferences*

Durmus OZBASI[1]

ARTICLE INFO

Article History:
Received: 11 Mar. 2019
Received in revised form: 28 May. 2019
Accepted: 08 Jul.2019
DOI: 10.14689/ejer.2019.82.4

Keywords
Rank-order judgments scaling, Evaluation preferences, Assessment preferences, Evaluation tool

ABSTRACT

Purpose: This study sought to determine university students' evaluation preferences and then scaled them based on their rank-order judgments. **Research Methods**: The survey model was used in this study. This study was conducted with a total of 376 university students of varying grade levels enrolled in different departments of the faculty of education of two separate state universities in Turkey during the 2017-2018 academic year. Data were collected using a 13-item survey designed specifically for this study that solicited answers regarding students' evaluation preferences in measuring their academic performance. Students first ranked to evaluation types from most to least preferred and then assigned a single number for each stimulus. The data attained from the study were then scaled based on rank-order judgments. **Findings**: The study findings revealed that students most preferred to be assessed using oral exams and least preferred tests composed of multiple-choice questions. **Implications for Research and Practice**: This study was restricted to university students enrolled in the faculty of education of two state universities in Turkey. By conducting a similar study with students enrolled in other faculties in the same or different higher education institutions results and potential differences between faculties may be compared.

*This study was partly presented at the 2rd International Symposium on Social Sciences Education in Çanakkale, 30 October – 01 November, 2018.

[1] Corresponding Author, Çanakkale Onsekiz Mart University, TURKEY, e-mail: dozbasi@comu.edu.tr , ORCID ID: https://orcid.org/0000-0001-5078-477X

Introduction

Current technology has developed at such remarkable speeds that access to information has been made profoundly easy, and this has forced educational systems to adjust themselves to fit the current age. Educational systems change in line with this and develop appropriate strategies conforming to prevailing pedagogical methods and techniques while also seeking to cultivate in students the type of human strengths needed in the current age. Consequently, education systems play a critical role within the greater education process in raising qualified individuals equipped with the knowledge and skills necessary to succeed in life.

One of the fundamental objectives of the educational process is, undoubtedly, to instill a set of desired behaviors in students and to have them act intentionally. In addition to this objective, education also seeks to determine deficiencies in students' learning, to correct any mistakes in their knowledge, and to provide effective feedback to students. Accordingly, students have at their disposal practical information to support their learning like what goals and proficiencies they have completed or acquired, and what skills they still need to attain. It is important that students' goals and learning outcomes are met in order to carry out the education process in a qualified way. Undoubtedly, one of the important components in this process is evaluation. The quality of the education system, its functioning and whether the process is proceeding effectively can only be determined by evaluation. Although evaluation is carried out at the end of the process and is seen as independent of teaching, it provides information on many variables such as how students prepare for lessons, what they experience with the learning process and how they prepare for the exam. Evaluations are used by educational systems not only to determine students' vocational or creative inclinations but also to increase teaching quality (Turgut & Baykul, 2010). Evaluations are needed for a variety of purposes in education systems, including diagnostic, formative and summative types.

Diagnostic evaluations are conducted to determine whether potential students have the necessary behaviors and/or cognitive skills required to attend a program, course, or activity (Tekin, 2005). This type of evaluation seeks to measure students' pedagogical characteristics as opposed to psychological ones (Özcelik, 2010). Formative-type evaluations are conducted after every unit or topic to determine students' deficiencies and to provide feedback related to these deficiencies. Consequently, formative-type evaluations are considered a part of the greater instruction process. Instead of gauging students' academic performance, this type of evaluation is given during the course of the instruction process to determine what aspects of a unit or topic students could improve upon and what their gaps in knowledge are. Summative evaluations are yet another type of evaluation that encompasses more than one proficiency and are given either during or after the instruction process. This type of evaluation seeks to assign a grade to students, to determine whether they have a sufficient foundation to continue to the subsequent step in their course or program, and serve as a predictor of future performance (Tekin, 2005). Accordingly, the degree to which students have attained the goals and

proficiencies expected of them and how prepared they are to proceed to the subsequent stage in their education may be measured.

Evaluations should not be considered independent from or as a tool administered solely at the end of instruction. Evaluations present information on different variables like how prepared students are for tests, what difficulties they encountered during instruction, and how well they have studied their lessons (Birenbaum, 1997; Struyven, Dochy & Janssens, 2005). According to Biggs (2003), evaluations make significant contributions to students' learning progress. For this reason, a quality evaluation method should not only constitute an integral part of students' learning processes but also guide students on how to make the most of their learning during this process (Gulbahar & Buyukozturk, 2008).

Evaluation preferences are defined as scoring procedures used to measure students' academic performance (Birenbaum, 1994). Both traditional and supplementary evaluations are frequently used while performing in-class evaluations to measure students' academic performance. Traditional evaluation approaches generally measure academic performance outside the normal flow of instruction, are based on output, and frequently use such items as multiple-choice, short-answer, true-false, matching, and fill-in-the-blank type items. Seeking primarily to measure the level of students' cognitive abilities to recall and conceptualize, this type of evaluation is preferred by teachers who adopt a constructive approach. In fact, most teachers consider themselves proficient in this type of evaluation approach (Gelbal & Kelecioglu, 2007). Supplementary evaluation approaches like performance-based tasks, portfolios, and rubrics seek to measure students' higher-level cognitive abilities (e.g., critical thinking, researching, creativity) are used to assess educational goals (Kutlu, Dogan & Karakaya, 2008). One of the most important differences between supplementary and traditional evaluation approaches is that the former allows one to assess educational goals in addition to measuring individual differences and multiple intelligence areas.

The evaluation method that students prefer to during instruction may influence their academic performance (Biggs, 2003). The literature contains several studies (Biggs, 2003; Birenbaum, 2003; Ekinci, 2009; Struyven et al., 2005) discussing how students' preferences regarding instructional and evaluation methods impact their academic performance, their perceptions related to learning, and how they participate in learning processes and then, how all of these work together to impact students' evaluation preferences. Previous studies have found that there are significant differences in students' academic performance and preferred type of evaluations based on their learning approaches (Mayya, Rao & Ramnarayan, 2004; McManus, Richards & Winder, 1999; Sambell et al., 1997; Scouller, 2000). Just as students have different intelligence areas, they also have different learning approaches. While some students are familiar with practice-based learning methods, others may prefer teaching methods based entirely on direct instruction. Students' learning habits may cause them to adopt specific test preparation habits, which, in turn, may cause them to prefer specific evaluation types over others. A review of the literature, however, reveals that the majority of studies focus more on statistical significance and

relationship with regard to students' evaluation preferences and that there are only a few studies (Altun & Gelbal, 2014; Birgin & Gürbüz, 2008; Gelbal & Kelecioglu, 2007; Gijbels & Dochy, 2006; Isnac, 2018; Struyven, Dochy & Janssens, 2005; Struyven, Dochy & Janssens, 2005; Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) seeking to identify what type of evaluations students like to take or which types they prefer over others. Of these studies, only in Isnac (2018) and Sahin, Ozturk Boztunc and Teker Tasdelen (2015) are scaling methods used to assess students' evaluation preferences. In their study, Sahin, Ozturk Boztunc and Teker Tasdelen (2015) found that pre-service teachers preferred true-false tests the most and performance-based tasks the least. Isnac (2018) conducted a study with middle school students in which she scaled students' evaluation preferences according to their learning approaches. Since in scaling approaches, psychological and emotional characteristics can be scaled and a shared point regarding individuals' preferences can be reached (Anil & Guler, 2006; Kan, 2008; Ozkan Ozer & Guvendir Acar, 2011), one method to determine which type of evaluation students prefer is to scale their preferences after having asked them directly. An examination of the literature reveals that only a single study exists investigating university students' evaluation preferences using rank-order judgment scaling. Accordingly, the current study sought to identify which types of evaluations university students preferred, to scale their preferences according to rank-order judgments, and to determine scale values for their preferences. To accomplish this objective, responses to the following questions were solicited:

1. What evaluation methods do university students prefer to measure their academic performance?
2. Using rank-order scaling, what are the item scale values for university students' evaluation preferences?

Method

Research Design

The survey model used in this study aims to describe a situation that has existed or still exists (Cresswell, 2003; Karasar, 2014). Accordingly, the current study follows a survey model since university students' evaluation preferences were sought.

Research Sample

The study group was described in this research. The study was conducted with a total of 376 students enrolled in the faculty of education in two different universities in Turkey (i.e., Canakkale Onsekiz Mart University and Trakya University). Approximately 31% (n=116) of the students participating in the study were male and 69% (n=260) were female. Table 1 presents student distributions by university and department.

Table 1

Student Distribution by Department

Department	f	%
German Language Education	11	2.9
Computer and Instructional Technologies	38	10.1
English Language Education	64	17.0
Japanese Language Education	24	6.4
Psychological Counseling and Guidance	136	36.2
Art Education	16	4.3
Social Sciences Education	35	9.3
Turkish Language Education	52	13.8
Total	376	100.0

Of the participating students, 2.9% were enrolled in the German Language Education Department, 10.1% in the Computer and Instructional Technologies Department, 17% in the English Language Education Department, 36.2% in the Psychological Counseling and Guidance Department, 4.3% in the Art Education Department, 9.3% in the Social Sciences Education Department, and 13.8% in the Turkish Language Education Department. Table 2 presents student distributions by grade level.

Table 2

Student Distribution by Grade

Grade level	f	%
1.00	56	14.9
2.00	109	29.0
3.00	203	54.0
4.00	8	2.1
Total	376	100.0

As seen in Table 2, 14.9% of the participating students were in their first year, 29% in their second year, 54% in their third year and 2.1% in their fourth year of undergraduate education.

Research Instruments and Procedures

The data collection tool was developed by the researcher for this study. During the development of the data collection tool, the researcher performed a review of the literature and determined potential tools that may be used to assess students' academic performance. A semi-structured interview form was also used to solicit

responses regarding students' evaluation preferences. A total of 80 students enrolled in different grade levels within the faculty of education completed this interview form in which they were asked to rank their evaluation preferences in writing. After examining students' responses to the interview form, similar responses were aggregated to form a total of 16 items. The items were subsequently presented to three measurement and evaluation experts whose opinions were used to construct a 13-item tool measuring students' evaluation preferences. The 13 items included in the measurement tool were: (i) preparing individual presentations, (ii) taking frequent quizzes, (iii) doing a project as homework instead of a test, (iv) taking an open-book and notes test (no time limit), (v) answering open-ended questions (closed book and notes, with time limit), (vi) doing group homework instead of a test, (vii) taking an oral exam, (viii) doing individual homework, (ix) answering short-answer and open-ended questions, (x) taking a multiple-choice test, (xi) taking a test including different types of questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching), (xii) doing individual homework, and (xiii) taking an open-book and notes test (with time limit). The researcher implemented the measurement tool with the students, which took approximately 5 minutes. The participants were provided with all necessary explanations regarding the measurement tool and how to complete it prior to its administration.

Data Analysis

Rank-order judgments based on Thurstone's (1927) law of comparative judgment (Turgut & Baykul, 1992) were used to analyze the data collected for the study. The rank-order judgment scaling method is itself based on one of two fundamental scaling approaches (i.e., the judgment and reaction). In the judgment approach, stimuli are scaled according to observers' judgments in a predetermined dimension. This approach seeks to define the degree of stimulation of K number of stimuli for each of the N number of observers. In this approach, the observer's duty is to rank each stimulus in the scaling dimension in reference to the other stimuli. Accordingly, the average of observer judgments for any stimulus constitutes its scale value. The reaction approach, however, determines individuals' reactions by executing K number of stimuli to a group composed of N number of individuals. In this approach, the individuals reacting are not objective experts. They only determine the position of the stimulus examined in reference to the other stimuli on the same scale (Anil & Guler, 2006; Guilford, 1954; Tezbasaran, 2004; Torgerson, 1958). In the this study, rank-order judgment scaling based on the judgmental decisions approach was used.

Since ranking is based on the size difference between stimuli, it resembles a pair-wise comparative approach. Moreover, since participants experience fewer contradictions in rank-order judgment scaling, the results are potentially more consistent (Guilford, 1954; Turgut & Baykul, 1992). All stimuli are given to observers in rank-order judgment scaling. Observers rank stimuli by assigning each stimulus a rank number. The basic premise in this scaling approach is that each stimulus is given a rank number. This way, the entire stimuli group is transformed into a uniform standard to which each individual stimulus is compared. Scale values are attained by comparing the ratio attained from the rank-order judgments assigned to stimuli with the uniform standard. Subsequently, the pair-wise comparison is analyzed as in the fifth equation of Thurstone's (1927) law of comparative judgment theory (Anil & Inal,

2018; Guilford, 1954). In this study, a rank-order frequencies matrix was constructed showing how many times and in which order students placed each of their preferred type of evaluation. A ratios matrix was constructed using the rank-order judgments that students assigned to their evaluation preferences. The unit normal deviation matrix was constructed by calculating the z-values corresponding to the ratio's matrix elements. The total of each column's values was written at the bottom row of the unit normal deviation matrix. Scale values were then attained by calculating the average of each z-value appearing in this row for each separate column.

Results

This section presents students' rankings for evaluation preferences in regard to scaling procedure steps and explanations. In order to answer the first secondary objective of this study, a rank-order frequencies matrix was obtained using the rankings of 376 students' preferences regarding 13 stimuli.

Table 3
Frequencies Matrix

Rank-Order Frequencies Matrix for Stimuli

R	r	A	B	C	D	E	F	G	H	I	J	K	L	M	Total
1	13	17	55	9	55	28	28	82	3	5	5	56	7	26	**376**
2	12	16	58	14	30	41	16	62	11	20	11	32	19	46	**376**
3	11	19	35	15	29	51	25	47	14	30	19	41	15	36	**376**
4	10	22	41	30	19	51	26	27	24	28	16	33	26	33	**376**
5	9	22	30	15	27	39	29	42	28	34	15	32	27	36	**376**
6	8	30	27	27	26	34	24	32	31	35	23	27	33	27	**376**
7	7	22	17	35	19	26	28	25	34	43	37	33	28	29	**376**
8	6	38	23	36	26	27	38	10	30	36	28	24	35	25	**376**
9	5	45	20	41	20	19	30	18	42	24	41	24	23	29	**376**
10	4	29	23	39	24	19	31	14	49	37	28	33	40	10	**376**
11	3	32	18	43	19	20	39	11	38	38	50	22	23	23	**376**
12	2	38	14	42	22	16	30	4	42	31	52	10	32	43	**376**
13	1	46	15	30	60	5	32	2	30	15	51	9	68	13	**376**
	Total	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	**376**	

Table 3 consists of a frequencies table depicting the rank numbers assigned by 376 judgments for each of the 13 total stimuli. The total number of observers for each row and column is 376. Following this procedure, n(sjk>ski) tables for all stimuli were prepared and a pair-wise frequencies matrix for all stimuli was constructed.

The ratios matrix in Table 4 was constructed by dividing the column totals in the frequency matrix by the square number of total participants ($n^2=376^2$).

Table 4

Ratios Matrix for Evaluation Preferences

	P Matrix												
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
A		0.30	0.51	0.42	0.30	0.45	0.21	0.52	0.44	0.56	0.32	0.52	0.37
B	0.70		0.66	0.55	0.47	0.60	0.37	0.67	0.61	0.70	0.48	0.66	0.53
C	0.49	0.34		0.41	0.28	0.44	0.19	0.51	0.42	0.56	0.30	0.51	0.36
D	0.58	0.45	0.59		0.42	0.54	0.33	0.60	0.54	0.63	0.42	0.60	0.47
E	0.70	0.53	0.72	0.58		0.65	0.38	0.73	0.66	0.76	0.50	0.71	0.56
F	0.55	0.40	0.56	0.46	0.35		0.26	0.57	0.49	0.61	0.36	0.57	0.42
G	0.79	0.63	0.81	0.67	0.62	0.26		0.82	0.76	0.84	0.62	0.80	0.67
H	0.48	0.33	0.49	0.40	0.27	0.43	0.18		0.42	0.55	0.29	0.50	0.34
I	0.56	0.39	0.58	0.46	0.34	0.51	0.24	0.58		0.63	0.36	0.58	0.41
J	0.44	0.30	0.44	0.37	0.24	0.39	0.16	0.45	0.37		0.26	0.46	0.31
K	0.68	0.52	0.70	0.58	0.50	0.64	0.38	0.71	0.64	0.74		0.70	0.55
L	0.48	0.34	0.49	0.40	0.29	0.43	0.20	0.50	0.42	0.54	0.30		0.35
M	0.63	0.47	0.64	0.53	0.44	0.58	0.33	0.66	0.59	0.69	0.45	0.65	

In addition to the ratios matrix presented in Table 4, the analysis continued using scaling methods based on a pair-wise comparative approach. In ratios matrices, the most significant point diagonals 1 was given to the most important point in the ratios matrix. Upon examination of Table 4, it is observed that the sum of diagonals for the ratios matrix was equal to 1. After confirming this, the analysis continued using the fifth equation like in the pair-wise comparative scaling method. As a result, a unit normal deviation matrix (z) was constructed and presented in Table 5.

Table 5
Unit Normal Deviation Matrix for Evaluation Preferences

	z Matrix (unit normal deviation matrix)												
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
A		-0.53	0.02	-0.21	-0.52	-0.12	-0.79	0.04	-0.16	0.15	-0.48	0.05	-0.34
B	0.53		0.42	0.12	-0.07	0.26	-0.33	0.45	0.28	0.53	-0.06	0.42	0.07
C	-0.02	-0.42		-0.23	-0.57	-0.14	-0.86	0.02	-0.19	0.14	-0.52	0.03	-0.37
D	0.21	-0.12	0.23		-0.20	0.10	-0.44	0.25	0.10	0.34	-0.19	0.25	-0.08
E	0.52	0.07	0.57	0.20		0.38	-0.32	0.62	0.41	0.70	0.01	0.57	0.15
F	0.12	-0.26	0.14	-0.10	-0.38		-0.65	0.17	-0.03	0.27	-0.35	0.17	-0.21
G	0.79	0.33	0.86	0.44	0.32	-0.65		0.92	0.71	0.99	0.30	0.84	0.44
H	-0.04	-0.45	-0.02	-0.25	-0.62	-0.17	-0.92		-0.21	0.13	-0.56	0.01	-0.40
I	0.16	-0.28	0.19	-0.10	-0.41	0.03	-0.71	0.21		0.33	-0.36	0.20	-0.22
J	-0.15	-0.53	-0.14	-0.34	-0.70	-0.27	-0.99	-0.13	-0.33		-0.65	-0.10	-0.49
K	0.48	0.06	0.52	0.19	-0.01	0.35	-0.30	0.56	0.36	0.65		0.52	0.13
L	-0.05	-0.42	-0.03	-0.25	-0.57	-0.17	-0.84	-0.01	-0.20	0.10	-0.52		-0.39
M	0.34	-0.07	0.37	0.08	-0.15	0.21	-0.44	0.40	0.22	0.49	-0.13	0.39	
Total	2.89	-2.63	3.14	-0.44	-3.87	-0.18	-7.59	3.50	0.95	4.81	-3.53	3.36	-1.71
Mean	0.24	-0.22	0.26	-0.04	-0.32	-0.02	-0.63	0.29	0.08	0.40	-0.29	0.28	-0.14
SJ	0.87	0.41	0.89	0.60	0.31	0.62	0.00	0.92	0.71	1.03	0.34	0.91	0.49

In order to determine university students' evaluation preferences, the data collected by the measurement tools were scaled based on rank-order judgments. The smallest value on the unit normal distribution matrix was -7.59 for evaluation preference G. By taking this value as the ranking criteria, an absolute value of 7.59 was attained and is shown on a number line in Figure 1.

Figure 1. Scale values placed on a number line

Table 6

Evaluation Preferences and Scale Values

Stimulus Rank Number	Scale Values	Evaluation Preference
9	0.87	Preparing Individual presentations.
4	0.41	Taking frequent quizzes.
10	0.89	Doing a project as homework instead of a test.
6	0.60	Taking an open-book and notes test (no time limit).
2	0.31	Answering open-ended questions (closed book and notes, with a time limit).
7	0.62	Doing group homework instead of a test.
1	0.00	Taking an oral exam.
12	0.92	Doing individual homework.
8	0.71	Answering short-answer and open-ended questions.
13	1.03	Taking a multiple-choice test.
3	0.34	Using a given individual article to write as homework.
11	0.91	Taking a test, including different types of questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching).
5	0.49	Taking an open-book and notes test (with a time limit).

As seen in Table 6, the most preferred evaluation type by students was taking an oral exam whereas the least preferred was taking a multiple-choice test. Students' evaluation preferences from most to least preferred were (1) taking an oral exam, (2) answering open-ended questions (closed book and notes, with time limit), (3) using a given individual article to write as homework, (4) taking frequent quizzes, (5) taking an open-book and notes test (with time limit), (6) taking an open-book and notes test (no time limit), (7) doing group homework instead of a test, (8) answering short-answer and open-ended questions, (9) preparing Individual presentations, (10) Doing a project as homework instead of a test, (11) taking a test including different types of

questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching), (12) doing individual homework, and (13) taking a multiple-choice test.

Discussion, Conclusion, and Recommendations

Using rank-order judgments-based scaling (one of the scaling methods based on judgment decisions), this study has sought to determine what evaluation types that university students enrolled in different departments preferred to be used to measure their academic performance.

This study found that students most preferred oral exams to measure their academic performance. Oral exams are defined as a type of evaluation in which questions are generally asked and responses are given orally (Tekin, 2005). Oral exams have historically been frequently used in educational environments. Although students are less likely to suffer from nervousness and unnecessary fear while taking oral exams, the existence of subjectivity in the scoring system, the interaction between the examiner and examinee, the preparation of individual questions for each examinee, and the fact that they need to be conducted individually render them considerably difficult to administer (Turgut & Baykul, 2010). Other hurdles that render the administration of oral exams even more difficult include the need to record answers given during the test and the need to ensure proper oversight of the exam in order to ensure impartiality and to avert any potential legally questionable situations (Sezer & Bilgin, 2009).

Students' second most preferred evaluation type was answering open-ended questions which they could use their book and notes to respond and for which a time limit was set. Although open-ended questions are appropriate to measure high-level cognitive skills, there are some difficulties in scoring them (e.g., scoring not being objective, low content validity) (Atilgan, Kan & Dogan, 2009). In their study examining the appropriateness of open-ended questions used in tests prepared by middle school teachers, Incecam, Demir and Demir (2018) found that teachers made extensive use of open-ended questions. The results of a study by Sahin, Ozturk Boztunc and Teker Tasdelen (2015) reveal that open-ended questions were the fourth most preferred method of evaluation by pre-service teachers. Scouller (1998) found that the majority of students who adopt a deep learning approach performed better on exams that included written open-ended item format.

The findings of the current study revealed that the third most preferred evaluation type by university students was using a given individual article to write as homework. Individual learning and wanting to complete the homework given to oneself are considered general characteristics of students espousing a deep learning approach (Minbashian, Huon & Bird, 2004; Prosser & Trigwell, 1999; Ramsden, 1991). In his study conducted with university students enrolled in a faculty of education, Scouller (1998) found that while those students embracing a deep learning approach performed poorly on multiple-choice item format, they received higher scores on written tests. Accordingly, students' preferences in being given an article to write as homework may be considered a sign that they have adopted a deep learning approach. As such, the findings of the current study are consistent with those of Scouller's (1998).

Examination of students' three least preferred evaluation types revealed that they preferred stimulus to be assessed by multiple-choice type questions. Despite nearly all of the large-scale tests students have taken from elementary school to university are multiple-choice tests, the fact that this type of test was the least preferred by students was an important finding of this study. This specific finding contradicts those obtained by Struyven, Dochy, and Janssens (2005) study, in which students stated that they experienced less anxiety and had higher expectations for success when responding to multiple-choice questions that did not require them to construct their own answers.

Students' second least preferred type of evaluation was individual homework. Students stated either that they did not like being assessed by individual homework or that they made efforts to avoid it altogether. Similarly, Ozer-Ozkan and Acar-Guvendir (2013) conducted a scaling study with students enrolled in a measurement and evaluation class in which they attempted to identify students' preferred teaching styles. In their study, they found that students' least preferred evaluation types were narratives and individual work. In a similar vein, Cross (1981) asserted that individual work could cause deficiencies in communication between student and instructor. Furthermore, students' fears of being subjected to a subjective grading scale may cause them to dislike being assessed by individual homework.

Students' third least preferred type of evaluation were those composed of different types of questions (e.g., multiple-choice, short-answer, true-false, matching). The nation-wide tests that students take (e.g., Council of Higher Education Exam [YKS], High School Entrance Exam [LYS]) are composed strictly of multiple-choice questions and include no other types of questions. The fact that students are most familiar with tests composed of a single type of question may explain their dislike of being assessed by tests containing a multitude of question types.

Several recommendations may be made based on the findings of this study. The first recommendation is that student anxiety may be slightly reduced by explaining them how the evaluation will be structured and by providing them a grading rubric when they are to be given open-ended questions or homework. As stated by the majority of students, a single type of question may be used on tests or students can be briefed on other question types prior to being given a test composed of more than one question type in order to reduce their prejudices toward those with which they are less familiar. Recommendations for subsequent studies include asking students to provide qualitative explanations as to why they ranked their preferred question types in evaluations in the order they did, as doing so will shed light on the rationale underlying their choices. Since this study was restricted to university students enrolled in the faculty of education of two universities, other researchers may conduct similar studies with students enrolled in other faculties in the same or different higher education institutions, and compare the results between them.

References

Altun, A. & Gelbal, S. (2014). Öğretmenlerinin kullandıkları ölçme ve değerlendirme yöntem veya araçlarının ikili karşılaştırma yöntemiyle belirlenmesi [Determining teachers' measurement tools or techniques via pair-wise comparison method]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 1-11.

Anil, D. & Guler, N. (2006). İkili karşılaştırma yöntemi ile ölçekleme çalışmasına bir örnek [An example of the scaling study by pair-wise comparison method]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 30, 30-36.

Anil, D. & Inal, H. (2017). *Psikofizikte ölçekleme uygulamaları*. Ankara: PegemAkademi.

Atilgan, H., Kan, A. & Dogan, N. (2007). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.

Biggs, J. (2003). *Teaching for quality learning at university* (2nd edition). United Kingdom: The Society for Research into Higher Education and Open University Pres.

Birenbaum, M. (1994). Toward adaptive evaluation - the student's angle. *Studies in Educational Evaluation*, 20, 239-255.

Birenbaum, M. (1997). Evaluation preferences and their relationship to learning strategies and orientations. *Higher Education*, 33, 71-84.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for evaluation. Segers, M., Dochy, F. ve Cascallar, E. (eds.). *Optimizing New Methods of Evaluation: In Search of Qualities and Standards* (13-36). Dordrecht, The Netherlands: Kluwer.

Birgin, O., & Gürbüz, R. (2008). Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki bilgi düzeylerinin incelenmesi [Investigation of pre-service primary teachers' knowledge level about measurement and evaluation]. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20, 163-179.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Harcourt Brace Jovanovich College Publishers: Philadelphia.

Cross, K. P. (1981). *Adults as learners.* San Francisco: Jossey-Bass.

Ekinci, N. (2008). *Üniversite öğrencilerinin öğrenme yaklaşımlarının belirlenmesi ve öğretme öğrenme süreci değişkenleri ile ilişkisi*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Gelbal, S. & Kellecioglu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33, 135-145* .

Gijbels, D. & Dochy, F. (2006). Students' evaluation preferences and approaches to learning: can formative evaluation make a difference?. *Educational Studies*, 32(4), 399-409.

Guilford, J. P. (1954). *Psychometrics methods*. New York: Mc Graw-Hill Book Co.

Gulbahar, Y. & Buyukozturk, Ş. (2008). Değerlendirme tercihleri ölçeğinin Türkçeye uyarlaması [Adaptation of evaluation preferences inventory to Turkish]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 148-161.

Incecam, B., Demir, E. & Demir, E. (2018). Ortaokul öğretmenlerinin sınıf içi ölçme ve değerlendirmelerde yazılı yoklamalarda kullandıkları açık uçlu maddeleri hazırlama yeterlikleri [Competencies of middle school teachers to prepare open-ended items used in open-ended test for in-classroom evaluation]. *Elementary Education Online*, 17(4): 1912-1927

Isnac, F. (2018). *Ortaöğretim öğrencilerinin öğrenme yaklaşımlarına göre değerlendirme tercihlerinin ölçeklenmesi* [Scaling evaluation preferences of secondary school students according to learning approaches]. Master Thesis, Ankara University, Educational Sciences Institute, Ankara.

Kan, A. (2008). Psikolojik değişkenleri ölçmek için kullanılan ölçekleme yaklaşımları üzerine bir karşılaştırma [A comparison between scaling procedures for measurıing psychological variables]. *Eğitimde Kuram ve Uygulama*, 4(1), 2-18.

Karasar, N. (2014). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.

Kutlu, O., Dogan, C. D. & Karakaya, I. (2008*). Öğrenci başarısının belirlenmesi performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi yayıncılık.

Mayya, S., Rao, A.K. & Ramnarayan, K. (2004). Learning approaches, learning difficulties and academic performance of undergraduate students ofphysiotherapy. *The Internet Journal of Allied Health Sciences and Practice*, 2(4), 1-6.

McManus, I.C., Richards, P. & Winder, B.C. (1999). Intercalated degrees, learning styles, and career preferences: Prospective longitudinal study of UK medical students. *British Medical Journal International edition*, 319, 542-546.

Minbashian, A., Huon, G. F. & Bird, K. D. (2004). Approaches to studying and academic performance in short-essay exams. *Higher Education*, 47,161-176.

Ozcelik, D. A. (2011). *Ölçme ve değerlendirme*. Ankara: Pegem Akademi

Ozkan Ozer, Y. & Guvendir Acar, M. (2013). Öğrencilerin ölçme ve değerlendirme dersinin sunulmasında tercih ettikleri öğretim yöntemleri [Students' teaching methods preferences for measurement and evaluation course]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* 4(1), 1-14.

Prosser, M. & Trigwell, K. (1999). Relational perspectives on higher education teaching and learning in the science. *Studies in Secience Education*, 33(1), 31-60.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire, *Studies in Higher Education*, 16:2, 129-150.

Sambell, K., McDowell, L. & Brown, S. (1997). 'But is it fair?': An exploratory study of student perceptions of the consequential validity of evaluation. *Studies in Educational Evaluation,* 23(4), 349-371.

Scouller, K. (2000). The influence of evaluation method on student learning. *Australian Association for Research in Education*, 35(4), 453-472.

Scouller, K. (1998). The influence of evaluation method on students' learningapproaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453–472.

Sezer, Y. & Bilgin, H. (2009). Sözlü sınavların yargısal denetimi [Judicial review of oral exams]. *TBB Dergisi*, 86, 168-187.

Struyven, K., Dochy, F. & Janssens, S. (2005). Students' perceptions about evaluation and evaluation in higher education: A review. *Evaluation & Evaluation in Higher Education*, 30(4), 325-341.

Sahin, M. G., Ozturk Boztunc N. & Teker Tasdelen, G.(2015). Öğretmen adaylarının başarılarının değerlendirilmesinde tercih ettikleri ölçme araçlarının belirlenmesi [Determining the pre-service teachers' measurement tool preferences for evaluation of their achievement]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6 (1), 95-106.

Tekin, H. (2004). *Eğitimde ölçme ve değerlendirme.* Ankara: Yargı yayın evi.

Tezbasaran A. (2004). Likert tipi ölçeklere madde seçmede geleneksel madde analizi tekniklerinin karşılaştırılması. *Türk Psikoloji Dergisi*, 19 (54), 77-90.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Rewiev*, 34, 273-286.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley & Son.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi

Turgut, F., M., & Baykul, Y. (1992). *Ölçekleme teknikleri.* Ankara: ÖSYM Yayınları

Öğrencilerin Değerlendirme Tercihlerinin Sıralama Yargılarına Dayalı Ölçeklenmesi

Atıf:

Ozbasi, D. (2019). Using rank-order judgments scaling to determine students' evaluation preferences. *Eurasian Journal of Educational Research*, 82, 63-80, DOI: 10.14689/ejer.2019.82.4

Özet

Problem Durumu: Eğitim sürecinin önemli ögelerinden biri değerlendirmedir. Eğitim sisteminin, işleyişi ve sürecin etkili bir şekilde ilerleyip ilerlemediğinin kontrolü ancak

değerlendirme ile belirlenebilmektedir. Ölçme sonuçlarının belirlenen ölçüt veya ölçütler takımıyla karşılaştırılarak bir karara varma süreci olan değerlendirme, öğrencilerin mesleki veya beceri açıdan eğilimlerinin belirlenmesine yardımcı olmasının yanısıra öğretimin niteliği hakkında ilgili paydaşlara önemli bilgiler sağlar (Turgut & Baykul, 2010). Değerlendirme, sadece öğrenme sürecinin sonunda gerçekleştirilen ve öğretimden bağımsız bir durum olarak ele alınamaz. Aynı zamanda öğrenci başarısı, hedef ve kazanımlara ulaşma düzeyleri hakkında da bilgi sağlar. Ayrıca, öğrencilerin sınavlara nasıl hazırlandıklarını, öğrenme sürecinde karşılaştıkları zorlukları, derslerine nasıl çalıştıkları gibi farklı değişkenler hakkında da bilgi elde edilmesine yardımcı olur (Birenbaum, 1997; Struyven, Dochy & Janssens, 2005). Biggs (2003)'e göre, değerlendirme öğrenmenin gelişmesine önemli katkı sunmaktadır. Bu nedenle de, nitelikli bir değerlendirme yöntemi, öğrencilerin öğrenme süreçlerinin bir parçası olmalı ve öğrencilere öğrenme sürecinde nasıl öğrenmeleri gerektiği konusunda da rehberlik etmelidir (Gulbahar & Büyüköztürk, 2008).

Öğrencilerin öğretimsel tercihleri ve değerlendirme yöntemlerine ilişkin tercihlerinin, akademik başarı, öğrenmeye ilişkin algıları ve öğrencilerin öğrenme sürecini nasıl gerçekleştirdiklerini ve de tüm bunların değerlendirme tercihine nasıl bir etkiye sahip olduğuna ilişkin alan yazında (Biggs, 2003; Brenbaum, 2003; Struyven & Digerleri, 2005; Ekinci, 2009) yapılmış araştırmalar bulunmaktadır. Yapılan araştırmalarda (Mayya, Rao ve Ramnarayan, 2004; McManus, Richards & Winder 1999; Sambell, McDowell & Brown, 1997; Scouller, 2000) öğrenme yaklaşımlarının öğrencilerin akademik başarıları ile değerlendirme tercihlerine anlamlı düzeyde farklılık oluşturduğu tespit edilmiştir. Ancak alan yazın incelendiğinde, yapılan araştırmaların çoğunluğunun, öğrencilerin değerlendirme tercihlerine ilişkin daha çok istatistiksel anlamlılık ve ilişki üzerine olduğu tespit edilmiş (Gelbal & Kelecioglu, 2007; Birgin & Gurbuz, 2008; Struyven, Dochy & Janssens, 2005; Gijbels & Dochy, 2006); öğrencilerin ne tür değerlendirme türlerini tercih ettiklerini ortaya çıkaran az sayıda (Altun ve Gelbal, 2014; Isnac, 2018; Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) araştırma bulunmaktadır. Bu araştırmalardan sadece Isnac (2018) ve Sahin, Ozturk Boztunc & Teker Tasdelen (2015) tarafından yapılan araştırmalarda, öğrencilerin değerlendirme tercihleri ölçekleme yöntemleriyle belirlenmiştir. Alan yazın incelendiğinde, üniversite öğrencilerinin değerlendirme tercihlerinin sıralama yargılarına dayalı ölçekleme yoluyla incelendiği sadece bir çalışmaya (Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) rastlanmıştır. Öğrencilerin başarılarının değerlendirilmesinde tercih ettikleri değerlendirme türlerinin belirlenmesi, öğrencilerin çalışma alışkanlıkları ve sınavlara nasıl hazırlandıkları hakkında önemli bilgiler ortaya koyacağı düşünülmektedir. Bu bağlamda araştırmanın problemi, üniversite öğrencilerinin değerlendirme tercihlerine ilişkin ölçek değerlerinin belirlenmesidir.

Araştırmanın Amacı: Araştırmanın amacı, üniversite öğrencilerinin değerlendirme tercihlerinin belirlenmesi ve değerlendirme tercihlerinin sıralama yargılarına göre ölçeklenmesidir.

Araştırmanın Yöntemi: Araştırma tarama modeli ile tasarlanmıştır. Araştırma Canakkale Onsekiz Mart ve Trakya Üniversitesi eğitim fakültelerinde öğrenim görmekte olan toplam 376 üniversite öğrencisi ile gerçekleştirilmiştir. Araştırmada veri toplama aracı araştırmacı tarafından geliştirilmiştir. Veri toplama aracının

geliştirme aşamasında, alan yazın taranmış ve öğrencilerin akademik başarılarını değerlendirmede kullanılabilecek ölçme araçları belirlenmiştir. Ayrıca araştırma kapsamında öğrencilere değerlendirme tercihlerinin sorulduğu yarı yapılandırılmış bir görüşme formu kullanılmıştır. Bu görüşme formu, eğitim fakültesinde çeşitli sınıflarda öğrenim görmekte olan 80 öğrenciye uygulanmış ve değerlendirme tercihlerini sırasız olarak yazmaları istenmiştir. Öğrencilerin vermiş oldukları yanıtlar incelenerek ortak olanlar (toplam 16 madde) bir araya getirilmiştir. Daha sonra bu maddeler uzmanların (üç ölçme ve değerlendirme uzmanı) görüşüne sunulmuş ve uzman görüşleri doğrultusunda 13 madde ölçme aracının maddelerini oluşturmuştur.

Araştırmanın Bulguları: Bulgularda öncelikle toplam 13 uyarıcı için 376 yargıcının her bir uyarıcı için vermiş oldukları sıra numaralarına ilişkin frekans tablosu oluşturulmuştur. Satır ve sütunların toplam gözlem sayısı 376 'dır. Bu işlemden sonra ikili olarak tüm uyarıcılar arasındaki $n(s_{jk}>s_{ki})$ tabloları hazırlanmış ve tüm uyarıcılar arasında ikili sıra frekanslar matrisi elde edilmiştir. Daha sonra oranlar matrisi hesaplanmıştır. Oranlar matrisinde en önemli nokta köşegenleri toplamının 1'i vermesidir. Buna göre, oranlar matrisindeki köşegenler toplamının 1'e eşit olup olmadığı incelenmiş ve köşegenler toplamının 1'e eşit olduğu tespit edilmiştir. Bu kontrolün ardından, V. Hal denklemi kullanılarak, ikili karşılaştırma yöntemi ile analize devam edilmiştir. Daha sonra birim normal sapmalar matris değerleri hesaplanmıştır.

Sonuç ve Öneriler: Araştırmadan elde edilen bulgulara göre, öğrenciler en çok sözlü sınavı tercih etmişlerdir. Sözlü sınav her ne kadar, öğrenciyi heyecandan ve sınav ile ilgili gereksiz korkulardan kurtarsa da, puanlanmasındaki öznellik, sınav yapan ile yapılan arasındaki etkileşim ve sınav sorularının her birey için önceden hazırlama gerekliliği gibi sebeplerden dolayı kalabalık gruplarda gerçekleştirilmesi oldukça zor olabilmektedir. (Turgut & Baykul, 2010).

Öğrencilerin değerlendirme türü olarak en çok tercih ettiği ikinci uyarıcı ise "sınırlı süre içerisinde, kitap ve defter açık bir şekilde açık uçlu sorularla değerlendirme" yapılmasıdır. Açık uçlu sorular ölçtüğü özellik açısından üst düzey bilişsel beceriler ölçmek için uygun olsa da, soruların değerlendirilmesi ve puanlanmasındaki bazı sorunlar (puanlamanın objektif olmaması, kapsam geçerliğinin düşük olması gibi) taşımaktadır (Atılgan, Kan & Dogan, 2009). Scouller (1998) tarafından yapılan araştırma sonuçlarına göre, derin öğrenme yaklaşımına sahip öğrencilerin çoğunlukla yazılı (açık uçlu) yoklama soru türündeki sınavlarda daha başarılı oldukları belirtilmiştir.

Araştırma sonuçlarına göre, öğrencilerinin tercih ettiği üçüncü değerlendirme türü ise "Bireysel makale ödevi verilerek" yapılan değerlendirmedir. Bireysel öğrenme ve verilen ödevi istekli olarak yerine getirme davranışı, çoğunlukla derin öğrenme yaklaşımını benimseyen öğrencilerin genel özelliklerinden biridir (Minbashian, Huon & Bird, 2004; Prosser & Trigwell, 1999; Ramsden, 1991). Bu bağlamda, araştırma grubunda yer alan ve derin öğrenme özelliğine sahip öğrenciler, bireysel makale ödevini değerlendirme türü olarak tercih etmiş olabilirler.

Öğrencilerin en az tercih etttikleri değerlendirme türü çoktan seçmeli madde türüdür. Öğrencilerin ilkokul seviyesinden üniversite düzeyine kadar girmiş olduğu geniş ölçekli sınavların neredeyse tümünün çoktan seçmeli sınav olmasına rağmen,

değerlendirme türü olarak en az tercih etmeleri, araştırmadan elde edilen önemli sonuçlardan biridir. Öğrencilerin en sondan ikinci olarak tercih ettikleri değerlendirme türü bireysel ev ödevidir. Öğrencilerin en sondan üçüncü sıraya yerleştirdikleri değerlendirme türü ise farklı soru türlerinin (çoktan seçmeli, kısa yanıtlı, doğru-yanlış, eşleştirme) aynı anda kullanılmasıyla gerçekleştirilen değerlendirmedir.

Bu araştırmanın sonuçlarına dayalı olarak yapılabilecek önerilerden biri, öğrencilerin birçoğunun belirtmiş olduğu gibi değerlendirme aşamasında tek tür soru formatı kullanılmasıdır. Araştırma bulgularında öğrencilerin, aynı anda birçok soru formatının kullanılmasını pek tercih etmedikleri belirlenmiştir. Bu nedenle, farklı soru formatları kullanıldığı durumlarda, bu soru formatları ile ilgili yönerge veya açıklamalara yer verilmesi, bu konudaki korku ve isteksizliği azaltacağı düşünülmektedir. Bu araştırma sadece eğitim fakültesinde öğrenim görmekte olan üniversite öğrencileri ile yürütülmüştür. Farklı fakülte veya yüksekokullarda öğrenim görmekte olan üniversite öğrencileri ile böyle bir çalışma yapılması alan yazına katkı sağlayabilir.

Anahtar Kavramlar: sıralama yargılarına dayalı ölçekleme, Değerlendirme tercihleri, Değerlendirme araçları

Eurasian Journal of Educational Research 82 (2019) 81-100

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

Gratitude, Hope, Optimism and Life Satisfaction as Predictors of Psychological Well-Being*

Ferhat KARDAS[1], Zekeriya CAM[2], Mustafa ESKISU[3], Sedat GELIBOLU [4]

ARTICLE INFO

Article History:
Received: 24 Sept. 2018
Received in revised form: 20 Jun. 2019
Accepted: 7 Jul. 2019
DOI: 10.14689/ejer.2019.82.5

Keywords
well-being, gratitude, hope, optimism, life satisfaction

ABSTRACT

Purpose: The mainstream psychology has mostly focused on the negative factors that threaten human mental health. With the emergence of positive psychology approach, the human strengths have started to be studied and the number of studies on the issue of happiness and well-being has increased. The aim of this research is to reveal the level of prediction for psychological well-being by the variables of gratitude, hope, optimism and life satisfaction. **Method:** Research sample consisted of 510 students from various faculties and departments of four different universities. The study group consisted of 359 female (70.4%) and 151 male (29.6%) university students. The age range of participants varied between 17 and 30.

Findings: All the variables discussed in the study were positively related to each other, and the variables of gratitude, optimism, hope, and life satisfaction all together accounted for about 51% of the variance of psychological well-being. Gratitude was determined as the most predictive variable for well-being and it was followed by the variables; hope, optimism, and life satisfaction, respectively. Accordingly; gratitude predicted 35.4% of the variance in well-being alone; gratitude and hope predicted 45.4% together; gratitude, hope, and optimism predicted 48.8% together; while gratitude, hope, optimism, and life satisfaction predicted 50.7% percent altogether.

Implications for Research and Practice: These findings are thought to be important in terms of revealing variables that predict well-being with a high level of variance. It is thought that these findings can be used in the field of mental health especially by the researchers and practitioners using intervention studies. Especially, since the gratitude level accounts for 35.4% of the variance in well-being alone, gratitude interventions can effectively be used to improve well-being level.

*This study was partly presented at the 14rd International Psychological Counseling and Guidance Congress in Antalya, 15-17 December, 2017

[1]Corresponding Author, Van Yuzuncu Yıl University, Faculty of Education, E-mail: kardas-90@hotmail.com, **Orcid:** 0000-0003-3386-3956

[2]Mus Alparslan University, Faculty of Education,E-mail: zekeriyacam@gmail.com, **Orcid:** 0000-0002-2965-2390

[3]Erzincan Binali YildirimUniversity, Faculty of Education, E-mail: meskisu@gmail.com, **Orcid:** 0000-0002-7992-653X

[4]Mus Alparslan University, Faculty of Education, E-mail: sgelibolu@gmail.com, **Orcid:** 0000-0002-9451-7435

Introduction

What makes a good and virtuous life and how to live it is one of the oldest debates in human history. This debate has continued for centuries within the framework of different worldviews, especially in the field of religion and philosophy. Psychology, a younger field of science than philosophy, has also benefited considerably from discussions and accumulations in philosophy while dealing with the subjects of happiness and good life. The history of psychology is based on observing some human behaviors experimentally in the laboratory environment. In the period from the beginning of the 19th century to the Second World War, it seems that an approach which focused on the negative aspects of life, concentrated on diagnosis and treatment, and dealt with the problematic parts of individuals rather than the strengths had been prevailing in the field of psychology. Especially after the Second World War, the influence of war, the spread of the understanding of humanism, and the need to embrace the positive aspects of life have forced a change in the dominant paradigm in psychology. The influence of the work of pioneers such as Carl Rogers and Abraham Maslow, in this framework, has begun to increase the emphasis on the strengths of people. The positive psychology approach has brought significant criticism to this pathology-oriented view and has suggested that psychology should embrace the positive aspects of life and aim to increase well-being by focusing on the strengths of individuals. In this context; Martin Seligman's speech (1999) as president of the American Psychological Association (APA) is considered as a milestone of positive psychology. In this speech, Seligman argued that psychology studies have spent half a century by examining negative mental effects of isolation, trauma, abuse, physical illness, war; but this negative focus has left psychology blind to the many instances of growth and mastery. For this reason, he proposed a new positive paradigm of psychology which focuses on positive sides of human and human strengths. This speech has prepared a basis for the formation of a positive psychology approach.

The positive psychology approach is mainly concerned with understanding issues such as happiness and well-being, positive features and activities that require intense attention, and meaningful positive relationships (Carr, 2013). In this framework, it focuses on the well-being of individuals with a paradigm of embracing positive sides of life. One of the issues that being widely discussed in this approach is the concept of well-being. While well-being is handled in the context of various conceptualizations and theoretical foundations, nowadays it is widely discussed in the distinction of subjective and psychological well-being. The basis of this distinction is also based on the discussion of *Eudaimonia-Hedonia* in the field of philosophy. In the positive psychology approach, the *Eudaimonic* tradition is studied in the works of Alan Waterman and Carol Ryff, and the hedonic approach is discussed in the works of Kahneman, Diener and Schwarz (Carr, 2013). Current debate continues in the field of positive psychology in terms of different contexts (Sheldon, Corcoran&Prentice, 2019).

Subjective well-being means individuals' avoidance from pain, having more positive experiences and more positive emotions. According to this, subjective well-

being refers to life satisfaction as a cognitive evaluation of life, less experience of negative feelings and more frequent experience of positive feelings (Diener, 2000). In this respect, subjective well-being is an approach parallel to the views of the hedonic approach which focuses on avoiding pain and reaching the pleasure. On the other hand, the concept of psychological well-being is more about being able to realize the potentials of the individual as a concept related to the "*Eudaimonia*", and it also refers to a meaningful and fulfilling life. According to this approach, fulfilling one's potential is at the heart of well-being (Carr, 2013; Hefferon&Boniwell, 2011).

One of the main issues that positive psychology concentrates on is positive feelings. Seligman (2004) classifies positive emotions as emotions related to the past, future, and present. According to this; optimism, hope, and confidence express positive feelings about the future; while satisfaction, gladness, and tranquility express about past; and physical pleasures like momentary pleasures, and persistent pleasures like joy, comfort, merriness and enthusiasm express feelings about present. In this context; gratitude, hope, and optimism are among the positive feelings that are widely discussed in the literature.

In recent years, the concept of gratitude is considered as one of the treatment factors in the field of mental health in both Turkey and abroad (Kardaş&Yalçın, 2018). Gratitude is conceptualized in various forms such as emotion, attitude, moral virtue, habit, personality trait and a coping reaction in the positive psychology literature (Emmons, 2009), and various definitions have been made about gratitude in this framework. According to this, gratitude is defined in various forms like; being grateful to the person who gives help and does a favor (Roberts, 2004), thanking for a goodness and being in a sense of gratitude (Watkins et al., 2003), a sense of gladness towards a benefit seen (Emmons, 2004), and as part of the general tendency to recognize and appreciate positive things in life (Wood, Froh&Geraghty, 2010). In various studies conducted in Turkey, Western Countries and Eastern Societies it has been demonstrated that gratitude has significant effects on; improving the mental health of individuals, increasing their well-being, strengthening social support perceptions and self-esteem, changing resilience levels positively, increasing optimism, and being a protective factor against stress and depression (Emmons & McCullough, 2003; Fredrickson, 2004; Kardaş, 2017; Lin, 2015; Lyubomirsky, 2008).

Another variable that is considered as one of the factors that affect well-being is optimism. Optimism basically means being in positive expectations about the future and it is considered in relation with emotion, perseverance, problem-solving; academic, sportive, military, vocational and political achievement, being healthy, and conceptualized differently in various studies(Peterson, 2000). Theoretical approaches like personality traits, attributional styles, expectancy effect (Scheier& Carver, 1987; Seligman, 2006) are determinative in the conceptualization of well-being. Research findings reveal that optimism is significantly related to subjective resilience (Terzi, 2008), life satisfaction (Sapmaz&Doğan, 2012), depressive symptoms (Chang &Sanna, 2001), and resistance to negative experiences (Segerstrom, 2007).

Hope, as a concept related to optimism, is conceptualized in the context of ability to plan paths to achieve the goal that the individual desires in spite of obstacles and a motivation source to utilize these paths (Snyder, 2000). In this framework, hope is defined as the determination to achieve goals and belief that there can be many other ways (Hefferon&Boniwell, 2011). According to this, hope is most powerful when it depends on the valuable targets that have moderate reaching possibility and derive from challenging but not impassable obstacles. In other words, the individual does not need to hope when he is sure he will achieve his goals, and he gets desperate when he thinks he will never succeed (Carr, 2013). The approach that deals with optimism in the context of attributional styles and the theory of hope in which Snyder conceptualized hope as the energy of fighting for the important goals suggest that individuals can significantly change their levels of optimism and positive expectations. Accordingly, individuals can remain optimistic and sustain their hopes through goal setting and having positive opinions about reaching those goals, maintaining motivation to make plans to reach these goals, and feeling of agency in these processes. This will make it easier for someone to reach their goals. Research conducted suggests that hope is also a factor closely related to the level of well-being of the individuals (Dursun, 2012).

In the literature, life satisfaction is expressed as a cognitive evaluation process in which the individual's life is assessed according to certain criteria (Shin &Johnson, 1978; Diener, Emmons, Larsen &Griffin, 1985). In this framework, life satisfaction is considered as one of the components of subjective well-being. According to this, life satisfaction is one of the important concepts expressing the well-being of individuals together with positive-negative affectivity as a cognitive evaluation of life.

The aim of this research is to reveal the degree of prediction for psychological well-being by different variables like gratitude, hope, optimism and life satisfaction in the university students, according to the order of relative importance, with a regression model.

The variables discussed in the study were determined as a result of a comprehensive literature review. When the studies are examined, it is seen that the studies that predict subjective well-being are more common (Ilhan&Ozbay, 2016).In some studies, optimism's predicting the level of well-being is examined (Eryilmaz&Atak, 2011), and in some other studies relationship between psychological well-being and some demographic variables such as university department and gender is studied (Gundogdu&Yavuzer, 2012). In this respect, no study has been found that takes the variables discussed as a whole like in the present study. On the other hand, it has been found that studies related to optimism, hope and life satisfaction can be found frequently but studies related to gratitude and well-being are limited. As mentioned earlier, gratitude includes some aspects of optimism, hope and life satisfaction, and the feeling of gratitude is a concept closely related to these characteristics of the individual. In this direction, in a study which intends to predict well-being, including the variable gratitude allows us to make some comparisons between variables. On the other hand, although the findings regarding the relationship between mentioned variables and well-being have been reported,

which variable contributes to the prediction of well-being comparatively has not been examined adequately. The current study aimed to make this comparison. Considering the situations above, it is thought to be important to discuss gratitude as a predictive variable, to examine predictive levels of variables discussed comparatively on the prediction of well-being, and to reveal the relations between these variables.

Taking all these issues into consideration, it is considered that this study is important in terms of contributing to positive psychology research in the literature, providing evidence-based data for teachers and practitioners working in the field, and determining variables predicting psychological well-being. On the other hand, research findings are expected to contribute to the issue of positive youth development which has become an important subject of research in the field of mental health and foster the development of limited studies conducted in Turkey.

The following questions will be sought within the framework of the research purpose.

1. Are there any significant relationships among variables of gratitude, hope, optimism, life satisfaction, and well-being?
2. Are gratitude, hope, optimism, and life satisfaction variables significant predictors of well-being, and what is their relative order of importance in terms of predicting participants' well-being level?

Method

Research Design

This research was carried out with the quantitative research model. These studies, which describe the relationships between variables, are considered as correlational survey models in the literature (Karasar, 2011).

Research Sample

The study group, determined by convenience sampling method, consisted of 510 students from various faculties and departments of four different public universities in different regions of Turkey. In terms of gender, 70.4% (359) of the participants were female and 29.6% (n = 151) were male. The age range of the participants varied between 17 and 30, with a mean age of 21.5 (SD= 2.05).

Research Instruments and Procedures

Research data were collected from four different public universities located in different regions of Turkey. In this context, first, the data collection tools to be used in the research were determined and necessary permissions were obtained from the researchers. Afterward, necessary permissions for the application of the measurement tools were obtained from the related faculties. The aim and process of the research were explained to the participants, and it was stated that the

participation inthe research was on a voluntary basis. It took about 15 minutes for each participant to complete the data collection tools.

The Flourishing Scale (Telef, 2013), Satisfaction with Life Scale (Durak, Senol-Durak&Gencoz, 2010), Gratitude Scale (Kardaş&Yalçın, 2019), Life Orientation Test (Aydın &Tezer, 1991), and Hope Scale (Akman&Korkut, 1993) were used for data collection in the study.

The Flourishing Scale: It is a measurement tool developed by Diener et al. (2010) to measure individuals' level of well-being. The scale consists of eight items answered on a 1–7 scale that ranges from Strongly Agree to Strongly Disagree, and scores from the scale range from 8-56 (e.g. *"I lead a purposeful and meaningful life"*). The high score indicates a higher level of well-being. It is adapted to Turkish by Telef (2013) and the Cronbach alpha coefficient of the scale is .80, and test-retest reliability score is .86. The Cronbach alpha internal consistency coefficient of the psychological well-being scale in this study was found to be .84. We performed the confirmatory factor analysis to determine the technical adequacies of The Flourishing Scale. Results from confirmatory factor analysis indicated that it had good model-data fit statistics ($\chi2$=70.532, df=19, $\chi2$/df=3.71, p=.000, RMSEA=.073 [90% confidence interval (CI) = .055-.091], SRMR= .039, GFI=.97, CFI=.96, NFI=.94, IFI=.96) and factor loadings (range from .57 to .65).

Satisfaction with Life Scale (SWLS): It is a measurement tool developed by Diener et al. (1985) to determine the life satisfaction of individuals. The scale consists of five items answered on seven points Likert type survey (e.g. "*In most ways, my life is close to my ideal*"). High scores indicate that the perception of life satisfaction is high. Various adaptation studies have been conducted in Turkish. Durak, Senol-Durak and Gencoz (2010) conducted the adaptation study of the form used in this research. Confirmatory factor analysis indicated that SWLS had good model-data fit statistics ($\chi2$=13.274, df=4, $\chi2$/df=3.31, p=.000, RMSEA=.067 [90% confidence interval (CI) = .030-.109], SRMR= .019, GFI=.99, CFI=.99, NFI=.99, IFI=.99) and factor loadings (range from .65 to .75).The Cronbach alpha internal consistency coefficient for the satisfaction with life scale in this study was found to be .83.

Gratitude Scale: It is developed by Kardas and Yalçın (2019) to measure gratitude levels of individuals. The scale consists of 25 items answered on five points Likert type survey, and higher scores indicate individuals' higher level of feeling gratitude. Scale consists of six sub-dimensions that are positive social comparison (e.g. *"When I look at the world, I see much to be grateful for."*), focusing on positive (e.g. "*I believe life is full of beauty.*"), appreciating contributions of family and others(e.g. "*I'm grateful to my parents for what they did for me.*"), feelings of abundance (e.g. "*I think I'm a lucky person in life.*"), gratitude for simple things (e.g. *"Even a small incident can increase my gratitude."*) and expressing gratitude (e.g. "*I have no trouble thanking for a favor.*"). The internal consistency coefficients are .88 for overall scale, .82 for positive social comparison, .73 for focusing on positive, .72 for appreciating contributions of family and others, .70 for feelings of abundance, .76 for gratitude for simple things and .72 for expressing gratitude. In the present study, the Cronbach alpha internal consistency coefficient was found to be .90. Results of the second-order confirmatory

factor analysis showed that Gratitude Scale had acceptable model-data fit statistics (χ2=948.04, df=269, χ2/df=3.52, p=.000, RMSEA=.070 [90% confidence interval (CI) =.066-.075], SRMR= .076, GFI=.84, CFI=.95, NFI=.93, IFI=.95) and factor loadings (range from .43 to .78).

Life Orientation Test: It was used to measure individuals' optimism levels in the study. Developed by Scheier and Carver (1985) to measure the life orientations of individuals, this scale is a five point-Likert-type measuring instrument consisting of 12 items(e.g., *"I'm always optimistic about my future"*). Scores on the scale range from 0 to 32 and high scores indicate that individuals have a positive view of life. Aydin and Tezer (1991) conducted the adaptation study of the scale to Turkish. Cronbach alpha coefficient was found to be .72 in the adaptation study. In this study, the Cronbach alpha internal consistency coefficient was found to be .71. We performed a confirmatory factor analysis for The Life Orientation Test. Results from confirmatory factor analysis indicated that it had good model-data fit statistics (χ2=78.088, df=17, χ2/df=4.59, p=.000, RMSEA=.084 [90% confidence interval (CI) = .066-.103], SRMR= .066, GFI=.97, CFI=.92, NFI=.90, IFI=.92) and factor loadings (range from .15 to .69).

Hope Scale: It was developed by Snyder et al. (1991) to assess hope levels of individuals and adapted to Turkish by Akman and Korkut (1993). Hope Scale consists of 8 items and two sub-dimensions as pathway (e.g., "*I can think of many ways to get out of a jam.*") and agency (e.g., "*I energetically pursue my goals.*"). Cronbach alpha coefficient is found to be .75 in the original form. In Turkish form, this coefficient is .65, and test-retest reliability score is .66. In this study, the form adapted by Akman and Korkut (1993) was used. Confirmatory factor analysis was performed primarily on the scores obtained from the scale and inequivalence with the original form, and it was found to be having two dimensions (Atik&Atik, 2017). We also conducted the second-order confirmatory factor analysis to determine the construct validity of The Hope Scale. Results from confirmatory factor analysis revealed that Hope Scale ha good model-data fit statistics (χ2=49.450, df=17, χ2/df=2.91, p=.000, RMSEA=.061 [90% confidence interval (CI) = .042-.081], SRMR= .036, GFI=.98, CFI=.98, NFI=.96, IFI=.98) and factor loadings (range from .47 to .75).In light of these results, the theoretical structure in the original form of the scale was confirmed. In this study, the Cronbach alpha internal consistency coefficient of the Hope Scale was found to be .84.

Data Analysis

Multiple linear regression analysis was used to analyze the data in this study. Before data analysis, alongside the assumptions of normality, linearity and homogeneity, multivariate normal distribution, linear relationship between predictor variable and dependent variable, absence of multiple-connection problem (multicollinearity) between independent variables, variance of error terms to be constant and mean of error terms equal to zero which are all assumptions of multiple linear regression were examined (Küçüksille, 2010). After the assumptions were met, basic analyzes were carried out. SPSS 21.0 software was used for statistical analysis. Stepwise regression analysis in which order of entry of variables depends on

statistical criteria was performed. Because the study aimed to determine the most significant predictive variables of well-being in order of relative importance, stepwise regression analysis, which is a model-building rather than model-testing, might be more beneficial to define variables in order to build structural models in future researches (Tabachnick&Fidell, 2013).

Results

In a preliminary analysis, descriptive statistics of psychological well-being, life satisfaction, gratitude, optimism, and hope variables were calculated, and then difference according to gender, the relationships between these variables, and generated regression model were examined, and the findings were presented under four headings.

Descriptive statistics of the variables examined in the study

Descriptive statistics were calculated for the mean averages related to the variables covered in the present study. In this way, how the data set had a distribution was examined. Descriptive statistics calculated for variables are presented in Table 1.

Table 1

Descriptive Statistics

Variables	N	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
Life satisfaction	510	5,00	35,00	20,16	6,77	-,102	-,641
Hope	510	11,00	32,00	25,34	4,10	-,601	,401
Gratitude	510	42,00	125,00	96,63	13,51	-,658	,795
Optimism	510	11,00	38,00	25,85	5,02	-,270	,031
Psychological Well-being	510	15,00	56,00	42,14	7,51	-,772	,493

When the values in Table 1 are examined, it is seen that the skewness and kurtosis values of the variables are within ± 1 values. It is stated in the literature that when the values of kurtosis and skewness are between ± 1 values, the dataset has a normal distribution (Büyüköztürk, 2013). Based on these values, it can be said that the mean scores of the variables handled within the scope of the research have a normal distribution.

Findings about relationships between variables

The relationship between the variables in the study was examined by Pearson's Product-Moment Correlation Coefficient and the results are given in Table 2.

Table 2

Relationships Between Variables

Variables	1	2	3	4	5
1.Psychological well-being	-	.596**	.501**	.516**	.527**
2. Gratitude		-	.468**	.486**	.394**
3. Life satisfaction			-	.464**	.402**
4. Optimism				-	.391**
5. Hope					-

** = *p*<.01

As shown in Table 2, significant positive relationships were found at moderate and high levels between psychological well-being and all variables. According to this, the relationship coefficients between well-being and other variables found to be; .596 (p<.01) for gratitude, .501 (p<.01) for life satisfaction, .516 (p<.01) for optimism, and .527 (p<.01) for hope.

After determining the relationships between the variables by correlation analysis, stepwise multiple regression analysis was performed to determine the predictability between the variables. However, there are some assumptions that this analysis requires. These are the absence of multicollinearity and autocorrelation between variables. Whether these requirements are met or not is examined with Durbin-Watson, Condition Index, Variance Inflation Factor/VIF, and Tolerance Value. Calculated Durbin-Watson value of this study is 1.925 while Condition Index is 24.099, VIF value is 1.494, and Tolerance Value is .677. Based on these values, it can be said that there is not a multicollinearity or autocorrelation problem among these variables. In order to determine the predictability between the variables in terms of the obtained values, stepwise multiple regression analysis was applied. The regression model for the prediction of the psychological well-being of the predictive variables in the study is given in Table 3.

Table 3

Stepwise Multiple Regression Analysis Result

Model	R	R^2	SE	ΔR^2	F	p
1.	,596[a]	0,355	6,04	0,355	279,474	,000
2.	,675[b]	0,456	5,55	0,101	94,077	,000
3.	,700[c]	0,491	5,38	0,035	34,481	,000
4.	,712d	0,507	5,30	0,016	16,896	.000

a. Dependent Variable: Psychological well-being

b. Predictors: (Constant), Gratitude

c. Predictors: (Constant), Gratitude, Hope

d. Predictors: (Constant), Gratitude, Hope, Optimism

e. Predictors: (Constant), Gratitude, Hope, Optimism and Life Satisfaction

It was determined that all added predictor variables to the generated model significantly predicted well-being, and according to the stepwise multiple regression analysis results, all generated models were significant. According to this, the variables that were most predictive of psychological well-being were gratitude, hope, optimism and life satisfaction, respectively. According to the results of the stepwise regression analysis, in the first model, gratitude alone predicted 35% of the well-being (R^2 = .355, p<.05). Adding the hope variable to the model, these two variables together accounted for 45 percent of the variance in well-being (R^2= .456, p<.05); when optimism was added, these three variables together accounted for49 percent of the variance in well-being (R^2= .491, p<.05); and at last when life satisfaction variable was added, four variables together accounted for 50.7 percent of the variance in well-being(R^2= .507, p<.05). This indicates that all variables are significant predictors but gratitude accounts for a large part of the variance alone for well-being.

Discussion, Conclusion, and Recommendations

In this study, the relationships between psychological well-being, life satisfaction, gratitude, optimism and hope variables in university students studying in different cities and universities were examined, and the created regression model was tested. The results showed that all the variables in the study had significant positive correlations with each other. Various studies have shown a positive relationship between these variables. In this context, in various studies, a positive relationship was found between gratitude and life satisfaction (Robustelli&Whisman, 2018; Yang, Zhang & Kou, 2016) and positive relationship among gratitude, hope, and happiness (Witvliet, Richie, Root Luna, & Van Tongeren, 2019). In another study, Guse and Shaw (2018) explored the relationship between hope, meaning in life, and well-being among a group of emerging adults who were enrolled at university. The results indicated that hope might lead to the setting of meaningful goals as well as spiritually meaningful experiences; thus, increasing meaning in life and eventually, well-being.

According to the results of the stepwise regression analysis of the variables that predicted psychological well-being, it was determined that the variables that were most predictive of psychological well-being were gratitude, hope, optimism, and life satisfaction, respectively; and these findings were discussed in the context of the literature.

The results showed that the most powerful predictor of psychological well-being in the study was gratitude. According to this, gratitude alone predicted about 36 percent of the variance in well-being. Gratitude is considered as one of the treatment factors in the field of mental health, and the number of efforts to improve the well-being of individuals through gratitude interventions is increasing day by day(Lyubomirsky 2008; Nelson &Lyubomirsky 2016; Young &Hutchinson, 2012). Gratitude is accepted as having a strong relationship with various measures of well-being, and experimental studies suggest that gratitude actually causes increases in happiness. In this context, it is suggested by Watkins, McLaughlin and Parker (2019)

that grateful responding can be enhanced by training in noticing the good in one's life, and by encouraging interpretations and appraisals that have been found to promote gratitude.

Gratitude is conceptualized and measured on the basis of such perspectives as making positive social comparisons, satisfying from simple things, focusing on the positive aspects of life, expressing gratitude, and recognizing the contributions of the family and the community. This suggests that all of the features mentioned are good predictors of well-being at the same time. Accordingly, it can be said that individuals having these characteristics are more satisfied with their lives and they are happier than the people whose gratitude levels are lower. On the other hand, the fact that gratitude predicts well-being stronger than life satisfaction, hope and optimism variables, can be considered as one of the important findings of this research. In line with the findings of this research, it was revealed that gratitude is a significant predictor of well-being just like in the studies investigating the relationship between gratitude and well-being held in Turkey, gratitude intervention activities improve well-being, and gratitude predicts well-being through various variables (Eryılmaz, 2014, Kardaş, 2017, Oğuz Duran & Tan, 2013). A number of studies abroad have also shown that gratitude is a significant and powerful predictor of well-being (Chan, 2013; Emmons & McCullough, 2003; Lin, 2015), and in various studies a positive relationship was found between gratitude and life satisfaction (Robustelli&Whisman, 2018; Yang, Zhang & Kou, 2016).

In the study, hope was found to be the second most predictive variable after gratitude in predicting psychological well-being. Hope basically means that individuals have motivation about the goals to be reached and that they can make plans to achieve their goals. In this sense, it is expected that having a mission to be accomplished, having the motivation to reach these goals, and making plans that can help to reach these goals will make a meaningful contribution to the life of the individual; and thus, increase well-being. In this research, when hope entered into the regression model, it contributes 10% to the explained variance alone. This indicates that hope is one of the important variables closely related to well-being. Relationship between hope and well-being is one of the well-studied subjects in the literature. In this context, in McCullough's (2002) conceptualization of the relationship between hope and gratitude, it is argued that people having more hope and gratitude generally appreciate their past as positive and pursue more meaningful goals for future. Besides, gratitude as appreciating others' positive contributions may facilitate the felling of hope. In another research on graduate students conducted by Uzun Ozer and Tezer (2008), the results have shown that there is a significant difference in the positive effect levels between individuals whose level of hope is high and low. Similarly, in a study on undergraduate students conducted by Dursun (2012); it has been stated that hope is a significant predictor of individuals' subjective well-being levels.

Another variable associated with well-being is optimism. Optimism alone was a significant predictor in the model and when it was included in the model after gratitude and hope variables, it contributed about 3% to the explained variance.

Optimism basically means that individuals are in a positive expectation about the future. It is expected that people with positive expectations will be more flexible than those with negative thoughts, be more open to alternative thinking, and develop a more positive attitude towards the difficulties. Therefore, it is expected that optimism will be a significant predictor of well-being. In parallel with the findings in this study, various previous studies have indicated that optimism has a significant relation with well-being and it is a significant predictor of well-being. Accordingly, in a study on university students, Sapmaz and Dogan (2012)found that optimism revealed 59% of the variance in happiness and 60% of the variance in life satisfaction. The fact that the predictive rate was much higher than this research finding might be due to the difference in approach to measure happiness or the effect of possible mediating variables. Thus, when optimism was added to this study along with other variables, the effect on the prediction remained low. This indicates that there may be mediating or moderating variables in the relationship between optimism and well-being. In this framework, in a study on university students conducted by Kardas (2017), it is found that optimism is one of the significant predictors of well-being, but social support, self-esteem, and psychological need satisfaction variables play a mediating role in the tested model. This suggests that optimism can have a greater effect on well-being with some mediating variables.

Another variable that is significantly related to psychological well-being is life satisfaction. According to the findings, life satisfaction accounts for a relatively small portion of the variance in well-being in comparison with gratitude, hope, and optimism. In this research, it was revealed that gratitude, optimism, and hope as positive emotions had a significantly higher effect on the prediction of well-being. This suggests that emotional components contribute more to well-being and happiness than a general cognitive appraisal of life. This proves the general assertion of the positive psychology approach that positive emotions can be used as an important therapeutic power in well-being and happiness interventions.

An approach regarding the predictive level and order of the variables of gratitude, optimism, and hope as positive emotions, and life satisfaction as a general cognitive appraisal of life in the prediction of well-being was followed. The results show that positive emotions, especially gratitude as a current concept of psychology (Kardas&Yalcin, 2018), are powerful predictors of psychological well-being. This situation confirms the positive psychology approach's criticism to the pathology-oriented approach of mainstream psychology. According to this, positive emotions can be counted as effective sources in increasing the happiness and well-being of the individuals.

In the light of the current findings, we suggest researchers design similar investigations by varying positive emotions about the prediction of well-being, put forward new model proposals by concentrating on the mediating relationships between the variables discussed in this study, and repeat this study on different age and occupational groups. The research also has several implications for practitioners and educators. According to this; it is important to investigate positive psychology-based topics in the field of education to conduct research and applications on

character strengths in education, and study the variables that increase the well-being of students. These efforts will directly contribute to the well-being of students and indirectly increase their academic success. Besides, this study revealed that gratitude explained about one-third of the variance in well-being level. Accordingly, effective gratitude interventions in school settings can also be used in Turkey.

References

Akman, Y., & Korkut, F. (1993). Umut ölcegi uzerine bir calişma [A Study on Hope Scale] *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *9*, 193-202. Retrieved from http://www.efdergi.hacettepe.edu.tr/shw_artcl-1370.html

Atik, G.,& Atik, Z. E. (2017). Predicting hope levels of high school students: The role of academic self-efficacy and problem-solving. *Education and Science*, *42*(190), 157-169. https://doi.org/10.15390/EB.2017.5348

Aydin, G., &Tezer, E. (1991). İyimserlik, sağlık sorunları ve akademik başarı ilişkisi [The relationship between optimism, health problems and academic achievement]. *TürkPsikolojiDergisi, 26*(7), 2-9.

Buyukozturk, S. (2013). *Sosyal bilimler için veri analizi elkitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* [Handbook of data analysis for social sciences: Statistics, research design, SPSS applications and interpretation]. (18. Baskı). Ankara: Pegem Akademi Yayınları.

Carr, A. (2013). *Positive psychology: The science of happiness and human strengths*. Routledge.

Chan, D. W. (2013). Subjective well-being of Hong Kong Chinese teachers: The contribution of gratitude, forgiveness, and the orientations to happiness. *Teaching and Teacher Education*, *32*, 22-30. Retrieved from https://www.sciencedirect.com/science/article/pii/S0742051X12001837

Chang, E. C.,&Sanna, L. J. (2001). Optimism, pessimism, and positive and negative affectivity in middle-aged adults: A test of a cognitive-affective model of psychological adjustment. *Psychology and Aging*, *16*(3), 524-531. https://doi.org/10.1037/0882-7974.16.3.524

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment, 49(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E. (2000). Subjective well-Being: The science of happiness and a proposal for a National Index. *American Psychologist*, *55*, 1, 34-43. Retrievedfrom https://psycnet.apa.org/journals/amp/55/1/34/

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S.,& Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143-156. https://doi.org/10.1007/s11205-009-9493-y

Durak, M., Senol-Durak, E., & Gencoz, T. (2010). Psychometric properties of the satisfaction with life scale among Turkish university students, correctional officers, and elderly adults. *Social Indicators Research, 99*(3), 413-429. https://doi.org/10.1007/s11205-010-9589-4

Dursun, P. (2012). *The role of meaning in life, optimism, hope, and coping styles in subjective well-being*. (Unpublished Doctoral Dissertation). Middle East Technical University, Graduate School of Social Sciences, Ankara.

Emmons, R. A. (2009). Gratitude. In *The Encyclopedia of Positive Psychology. (pp. 442-447).* Willey-Blackwell.

Emmons, R. A. (2004). The psychology of gratitude: An introduction. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude*(pp. 3–16). New York, NY: Oxford University Press.

Emmons, R. A., &McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, *84*, 377–389.

Eryilmaz, A., & Atak, H. (2001). The Relationships between adolescent subjective wellbeing, self-esteem and optimism. Electronic Journal of Social Sciences, 10(37), 170-181. Retrieved from

https://dergipark.org.tr/esosder/issue/6151/82621

Eryilmaz, A.(2014). Üniversite öğrencileri için geliştirilen öznel iyi oluşu artırma programının etkililiğinin incelenmesi [Examining the effectiveness of the subjective well-being increasing program developed for university students]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, *1*(31), 111-128. Retrieved from http://dergipark.gov.tr/maeuefd/article/206247

Fredrickson, B. L. (2004). Gratitude, like other positive emotions, broadens and builds. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 145-166). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.003.0008

Guse, T.,&Shaw, M. (2018). Hope, Meaning in Life and Well-Being Among a Group of Young Adults. In *Hopefor a Good Life* (pp. 63-77). Springer, Cham. https://doi.org/10.1007/978-3-319-78470-0_3

Gündoğdu, R., &Yavuzer, Y. (2012). Eğitim fakültesi öğrencilerinin öznel iyi oluş ve psikolojik ihtiyaçlarının demografik değişkenlere göre incelenmesi [Examining subjective well-being and psychological needs of students at the educational faculty according to demographic variables]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 12*(23), 115 – 131. Retrieved from http://acikerisim.aksaray.edu.tr/xmlui/handle/20.500.12451/1085

Hefferon, K.,&Boniwell, I. (2011). *Positivepsychology: Theory, researchandapplications*. McGraw-HillEducation.

Ilhan, T., & Ozbay, Y. (2016). Yaşam amaclarinin ve psikolojik ihtiyac doyumunun oznel iyi olus uzerindeki yordayici rolu [The Predictive role of life goals and

psychological need satisfaction on subjective well-being]. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, *4*(34), 109-118. Retrieved from http://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=12eef767-91e1-4a49-be01-e4ffbee235d7%40sdc-v-sessmgr02

Karasar, N. (2011). Bilimsel arastirma yontemi [Scientific research methodology]. *Ankara: Nobel*.

Kardas, F. (2017). Universite ogrencilerinde bir model sinamasi: pozitif duyguların ve kisisel kaynaklarin etkilesimi [Testing a model on thecollege students: interaction of positive feelings and personal resources]. (Unpublished Doctoral Dissertation). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Kardas, F., & Yalcin, İ. (2018). Sukran: Ruh Sağliği Alaninda Guncel Bir Kavram. *Psikiyatride Güncel Yaklasimlar - CurrentApproaches in Psychiatry*, *10*(1), 1–18. https://doi.org/10.18863/pgy.334494

Kardas, F., & Yalcin, İ. (2019). Sükran olcegi: Gecerlik ve guvenirlik calismasi (Gratitude scale: Validity and reliability study). *Electronic Journal of Social Sciences. 18*(69): 13-31. https://doi.org/10.17755/esosder.406306

Kucuksille, E. (2010). *SPSS uygulamaları, cok degiskenli istatistik teknikleri* [SPSS applications and multivariate statistical techniques]. (In: Ş. Kalaycı (ed). Ankara: AsilYayınDağıtım.

Lin, C. C. (2015). Gratitude and depression in young adults: The mediating role of self-esteem and well-being. *Personality and Individual Differences*, *87*, 30-34. https://doi.org/10.1016/j.paid.2015.07.017

Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting life you want.* New York: The Penguin Press.

McCullough, M. E. (2002). Savoring life, past and present: Explaining what hope and gratitude share in common. *Psychological Inquiry, 13*(4), 302–304. Retrieved from http://www.jstor.org/stable/1448873

Nelson, S.K., &Lyubomirsky, S. (2016). Gratitude. In: H. S. Friedman (Ed), *Encyclopedia of Mental Health (2. Bas).* (pp. 277-280). Waltham, MA: Academic Press.

Oguz Duran, N., & Tan, S. (2013). Minnettarlik ve yasam amaclari yazma calismalarinin oznel iyi olusa etkisi [The Effects of gratitude and life goals writing tasks on subjective well-being]. Türk Psikolojik Danışma ve Rehberlik Dergisi, *5*(40), 154-166. Retrieved from
Retrieved from https://dergipark.org.tr/tpdrd/issue/21460/229894

Peterson, C. (2000). The future of optimism. *American Psychologist, 55,* 44–55. https://doi.org/10.1037/0003-066X.55.1.44

Robustelli, B. L., &Whisman, M. A. (2018). Gratitude and life satisfaction in the United States and Japan. *Journal of Happiness Studies*, *19*(1), 41-55. https://doi.org/10.1007/s10902-016-9802-5

Roberts, R.C. (2004). The Blessings of gratitude. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 58–78).). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.003.0004

Sapmaz, F., & Dogan, T. (2012). Iyimserligin mutluluk ve yasam doyumuna etkisinin incelenmesi [Examination of the Effect of Optimism on Happiness and Life Satisfaction]. *Mersin Universitesi Egitim Fakültesi Dergisi*, *8*(3), 63-69. Retrieved from https://dergipark.org.tr/mersinefd/issue/17381/181487

Scheier, M. F.,& Carver, C. S. (1985). Optimism, coping and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4*,(3) 219- 247. Retrieved from https://psycnet.apa.org/record/1986-19862-001

Scheier, M. E., &Carver, C. S. (1987). Dispositional optimism and physical well-being: The influence of generalized outcome expectancies on health. *Journal of personality*, *55*(2), 169-210. https://doi.org/10.1111/j.1467-6494.1987.tb00434.x

Segerstrom, S. C. (2007). Optimism and resources: Effects on each other and on health over 10 years. *Journal of Research in Personality, 41*, 772- 786. Retrieved from https://www.sciencedirect.com/science/article/pii/S0092656606001127

Seligman, M. E. P. (1999). Positive social science. *Journal of Positive Behavior Interventions, 1*(3), 181-182. https://doi.org/10.1177/109830079900100306

Seligman, M. E. P. (2004). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Simonand Schuster.

Seligman, M. E. P. (2006). *Learned optimism: How to change your mind and your life*. Vintage.

Sheldon, K. M.,Corcoran, M., &Prentice, M. (2019). Pursuing eudaimonic functioning versus pursuing hedonic well-being: The first goal succeeds in its aim, whereas the second does not. *Journal of Happiness Studies*, *20*(3), 919–933. https://doi.org/10.1007/s10902-018-9980-4.

Shin, D. C., &Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social Indicators Research*, 5, 475-492. Retrievedfrom https://link.springer.com/article/10.1007/BF00352944

Snyder, C. R. (2000). Handbook of hope: Theory, measures, and applications. In *Handbook of hope: Theory, measures, and applications*.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., ... & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality And Social Psychology*, *60*(4), 570.https://doi.org/10.1037/0022-3514.60.4.570

Tabachnick, B. G., &Fidell, L. S. (2013). *Using multivariate statistics (6th ed.)*. Pearson.

Telef, B. B. (2013). Psikolojik iyi olus olceği: Türkceye uyarlama, gecerlik ve guvenirlik calisması [The Adaptation of Psychological Well-Being into

Turkish: A Validity and Reliability Study]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *28*(28-3). Retrieved from http://dergipark.gov.tr/hunefd/issue/7791/101929

Terzi, S. (2008). Universite ögrencilerinde kendini toparlama gucunun icsel koruyucu faktorlerle iliskisi [The Relationships between resilience and internal protective factors in university students]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *35*, 297-306. Retrieved from http://dergipark.gov.tr/hunefd/issue/7803/102295

Uzun Ozer, B.,& Tezer, E. (2008). Umut ve olumlu-olumsuz duygular arasindaki ilişkiler [Relationship between hope and positive-negativeaffect]. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, *23*, 81-86. Retrievedfrom http://dergipark.gov.tr/deubefd/issue/25428/268280

Watkins, P. C., McLaughlin, T., & Parker, J. P. (2019). Gratitude and subjective well-being: Cultivating gratitude for a harvest of happiness. In *Scientific Concepts Behind Happiness, Kindness, and Empathy in Contemporary Society* (pp. 20-42). IGI Global.

Watkins, P. C.,Woodward, K., Stone, T.,&Kolts, R. L. (2003). Gratitude and happiness: Development of a measure of gratitude, and relationships with subjective well-being. *Social Behavior and Personality, 31*(5), 431–452. https://doi.org/10.2224/sbp.2003.31.5.431

Witvliet, C. van O., Richie, F. J., Root Luna, L. M., & Van Tongeren, D. R. (2019). Gratitude predicts hope and happiness: A two-study assessment of traits and states. *Journal of Positive Psychology*, *14*(3), 271–282. https://doi.org/10.1080/17439760.2018.1424924

Wood, A. M., Froh, J.J., & Geraghty, A.W.A. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review*, *30(7),* 890–905. https://doi.org/10.1016/j.cpr.2010.03.005

Yang, Y.,Zhang, M. &Kou, Y. (2016). Self-compassionand life satisfaction: Themediating role of hope,*PersonalityandIndividual Differences,98*, 91–95. https://doi.org/10.1016/j.paid.2016.03.086

Young, M. E., & Hutchinson, T.S. (2012). The Rediscovery of gratitude: Implications for counseling practice. *Journal of Humanistic Counseling, 51*, 99-113. https://doi.org/10.1002/j.2161-1939.2012.00008.x

Psikolojik İyi Oluşun Yordayıcıları Olarak Şükran, Umut, İyimserlik ve Yaşam Doyumu

Atıf:

Kardas, F., Cam, Z., Eskisu, M., & Gelibolu, S. (2019). Gratitude, hope, optimism and life satisfaction as predictors of psychological well-being. *Eurasian Journal of Educational Research,* 82, 81-100, DOI: 10.14689/ejer.2019.82.5

Özet

Problem Durumu: 19. yüzyılın sonlarında başlayıp İkinci Dünya Savaşı'na kadar devam eden süreçte, psikoloji alanında genel anlamda hayatın negatif yönlerine odaklanan, sorunları teşhis ve tedaviye yoğunlaşan ve bireylerin güçlü yönleri yerine sorunlu yönlerine odaklanan bir yaklaşımın egemen olduğu görülmektedir. Özellikle İkinci Dünya Savaşı'ndan sonra savaşın etkisi, hümanizm anlayışının yaygınlaşması, hayatın pozitif yönlerini kucaklamaya yönelik ihtiyaç psikolojide hakim paradigmada farklılaşmalara neden olmuştur. Bu çerçevede Carl Rogers, Abraham Maslow gibi öncülerin çalışmalarının etkisiyle insanların güçlü yönlerine olan vurgu artmaya başlamıştır. Pozitif psikoloji yaklaşımı patoloji (hastalık) odaklı bu bakış açısına önemli eleştiriler getirmiş ve psikolojinin hayatın pozitif yönlerini kucaklaması gerektiğini ve bireylerin güçlü yönlerine odaklanarak onların iyi oluşlarını arttırmayı hedeflemesi gerektiğini ileri sürmüştür. Bu çerçevede ortaya çıkan ortaya çıkan çalışmalar 2000'li yılların başında Pozitif Psikoloji alanının sistematik olarak ortaya çıkmasına zemin hazırlamıştır. Pozitif psikoloji yaklaşımının yaygınlaşmasıyla birlikte mutluluk, iyi oluş, yaşam doyumu gibi konular sıkılıkla araştırılmaya başlanmıştır. Bu bağlamda bireylerin yaşam doyumlarını ve iyi oluş düzeylerini yordayan değişkenlerin ortaya konulması ve bu değişkenlere ilişkin çeşitli kuramsal ve uygulamalı çalışmaların gerçekleştirilmesi önem kazanmıştır. Bu bağlamda şükran, umut, iyimserlik ve yaşam doyumu değişkenlerinin birbirileriyle olan ilişkilerinin ortaya konulması ve bu değişkenlerin iyi oluşu yordama düzeylerinin ortaya konulmasının yararlı olacağı düşünülmektedir.

Araştırmanın Amacı: Yukarıda ifade edilenler bağlamında, bu araştırmanın amacı, üniversite öğrencilerinde şükran, yaşam doyumu, iyimserlik ve umut düzeylerinin psikolojik iyi oluş durumunu, göreli önem sırasına göre, hangi düzeyde yordadığını bir regresyon modeli ile ortaya koymaktır. Bu amaç çerçevesinde sözü edilen değişkenlerin arasındaki ilişkilerin ortaya konulması ve bu değişkenlerin iyi oluşu yordama düzeylerinin önem sırasına göre belirlenmesi amaçlanmaktadır. Diğer yandan araştırmada ele alınan değişkenlerin cinsiyete göre anlamlı farklılaşma gösterip göstermediği incelenmektedir.

Araştırmanın Yöntemi: Araştırma tarama modeli ile gerçekleştirilen nicel bir çalışmadır. Araştırmanın çalışma grubu üniversite öğrencilerinden oluşmaktadır. Çalışmaya Van Yüzüncü Yıl Üniversitesi, Muş Alparslan Üniversitesi, Erzincan

Binali Yıldırım Üniversitesi ve Ankara Üniversitesinde, 2017-2018 eğitim öğretim yılında çeşitli fakülte ve bölümlerde öğrenim görmekte olan 510 öğrenci katılmıştır. Araştırmaya katılanların % 70,4'ünü (359) kadın, % 29,6'sını (*n*=151) erkekler oluşturmaktadır. Katılımcıların yaş aralığı 17-30 aralığında değişmekte olup, yaş ortalaması 21,5'tir (S_x=2,05). Araştırmada veri toplama amacıyla Psikolojik İyi Oluş Ölçeği Yaşam Doyumu Ölçeği,, Şükran Ölçeği, Yaşam Yönelimi ve Umut Ölçeği kullanılmıştır. Veri analizinde çoklu doğrusal regresyon analizinden yararlanılmıştır. Veri analizi öncesinde normallik, doğrusallık ve homojenlik varsayımlarının yanı sıra, çoklu doğrusal regresyon modelinin sayıltıları olan çok değişkenli normal dağılım, doğrusallık, yordayıcı değişkenlerle bağımlı değişken arasında doğrusal bir ilişki, bağımsız değişkenler arasında çoklu bağlantı (multicollinearity) probleminin olmaması, hata terimlerinin varyansının sabit ve hata terimlerinin ortalamasının sıfıra eşit olması varsayımları incelenmiştir. Varsayımların karşılandığı anlaşıldıktan sonra temel analizler yapılmıştır. Analizlerde SPSS paket programı kullanılmıştır. Veri analizi sürecinde betimsel istatistikler, iki aşamalı çoklu doğrusal regresyon analizi, bağımsız örneklem t testi ve Pearson Momentler Çarpımı Korelasyon Katsayısı yöntemlerinden yararlanılmıştır.

Araştırmanın Bulguları*:* Araştırmada sonucunda yaşam doyumu, şükran, iyimserlik ve umut değişkenlerinden sadece yaşam doyumu düzeyinin cinsiyete göre anlamlı şekilde farklılaştığı, diğer değişkenlerde bir farklılaşma olmadığı belirlenmiştir. Diğer yandan araştırmadaki bütün değişkenlerin birbirileriyle anlamlı pozitif ilişkilere sahip oldukları sonucuna ulaşılmıştır. Psikolojik iyi oluşu yordayan değişkenlere ilişkin aşamalı regresyon analizi sonuçları psikolojik iyi oluşu en fazla yordayan değişkenlerin sırasıyla şükran, umut, iyimserlik ve yaşam doyumu olduğu belirlenmiştir.Araştırmada psikolojik iyi oluşun en güçlü yordayıcısının şükran duyma olduğu belirlenmiştir. Buna göre şükran duyma tek başına iyi oluştaki değişimin yaklaşık yüzde 36'sını yordamaktadır.

Araştırmanın Sonuçları ve Önerileri*:*Bu araştırmada farklı şehirlerde ve farklı üniversitelerde öğrenim görmekte olan üniversite öğrencilerinde psikolojik iyi oluş, yaşam doyumu, şükran, iyimserlik ve umut değişkenleri arasındaki ilişkiler incelenmiş ve oluşturulan regresyon modeli test edilmiştir. Araştırma sonucunda ele alınan değişkenlerin birbirileriyle çeşitli düzeylerde anlamlı pozitif ilişkiler içinde olduğu ortaya konulmuştur. Araştırma bulguları çerçevesinde araştırmacılara iyi oluşu yordamaya ilişkin pozitif duyguların çeşitlenerek benzer araştırmaların tasarlanması, bu araştırmada ele alınan değişkenler arasındaki aracılık ilişkilerine yoğunlaşarak yeni model önerilerinin ortaya konulması, çalışmanın farklı yaş ve meslek gruplarıyla tekrarlanması önerilmektedir. Uygulayıcıların bu araştırmanın bulgularından hareketle şükranı arttırmaya ve şükran aracılığıyla bireylerin iyi oluşlarına katkı sunmaya, iyimserlik ve umut çalışmaları aracılığıyla iyi oluşu arttırmaya ve bireylerin yaşam doyumlarını yükseltmeye ilişkin çalışmalar yapmalarının önemli olduğu düşünülmektedir.

Anahtar Kelimeler: İyi oluş, şükran, umut, iyimserlik, yaşam doyumu

Eurasian Journal of Educational Research 82 (2019) 101-120

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

Undergraduate Counseling Trainees' Perceptions and Experiences Related to Structured Peer Group Supervision: A Mixed Method Study*

Gokhan ATIK[1] , Zeynep ERKAN ATIK[2]

ARTICLE INFO

Article History:
Received: 23 Mar.2019
Received in revised form: 15 May. 2019
Accepted: 14 Jun. 2019
DOI: 10.14689/ejer.2019.82.6

Keywords
counseling supervision, counselor education, structured peer group supervision, mixed method research design, linear growth model

ABSTRACT

Purpose: The aim of this study was to investigate the perceptions of undergraduate counseling trainees in regards to the use of the structured peer group supervision format and its effect on their counseling self-efficacy beliefs.

Method: This study utilized a mixed method research design incorporating both qualitative and quantitative methodologies. To obtain the opinions of counseling trainees in regards to the use of the structured peer group supervision format, qualitative data was collected via a demographic information form and survey developed specifically for this research. While, the quantitative data for the study was collected via a Counselor Activity Self-Efficacy Scale (CASES) utilized to measure the counseling self-efficacy levels of the participating counseling trainees. The research participant group was comprised of senior-level university students studying in the Counseling and Guidance program at a state university located in Ankara, Turkey. The participant group for the qualitative portion of the study included 16 undergraduate-level counseling trainees, while the participant group for the quantitative portion included 21 undergraduate-level counseling trainees.

Findings: The qualitative findings indicated that the trainees' perceptions were clustered into four main themes: (1) contributions of the model, (2) strengths of the model compared to individual supervision, (3) aspects of the model that need improvement, and (4) applicability of the model for undergraduate-level counseling students. The quantitative findings revealed that the structured peer group supervision process positively affected the growth of the trainees' counseling self-efficacy beliefs.

Implications for Research and Practice: The results suggested that the structured peer group supervision format could be a useful and practical means of providing appropriate and effective supervision for undergraduate-level counseling trainees in Turkey.

*This study was partly presented at the XIII. National Psychological Counseling and Guidance Congress in Mersin, Turkey; and at the XI. International Interdisciplinary Conference on Clinical Supervision, Adelphi University, Garden City, New York, USA.

[1] Corresponding Author, Ankara University, Faculty of Educational Sciences, Guidance and Counseling, Ankara, TURKEY. E-mail: gokhanatik@gmail.com, ORCID: https://orcid.org/0000-0001-5386-6770
[2] Başkent University, Faculty of Education, Guidance and Counseling, Ankara, TURKEY. E-mail: zeyneperkanatik@gmail.com, ORCID: https://orcid.org/0000-0002-5310-2532

Introduction

Historically, the development of the counseling and guidance field within Turkish schools began in the 1950s (Tan, 1986). During this time, in order to develop disciplines such as psychology, counseling, and education within the Turkish educational system, scholars in the field of education visited Turkey from the United States and Europe in order to exchange their ideas with Turkish educators and scholars. In addition, Turkish educators and scholars reciprocated in the exchange of ideas and knowledge by visiting their counterparts in the United States and Europe. The exchange of ideas with the international academic counseling and guidance community was the impetus for the development of counseling and guidance programs within Turkey, leading to the first counseling and guidance undergraduate-level university program established in Turkey at Ankara University in 1965, and the first graduate-level university counseling and guidance program in Turkey established at Hacettepe University in 1967. Over the past 69 years, the number of counseling and guidance programs in Turkey has grown exponentially, and as a result, according to data published in 2018 posted on the Turkish Council of Higher Education official website, 66 state and 18 private foundation universities had students registered in undergraduate-level counseling and guidance education programs (Council of Higher Education, 2018). At the same time, 24 doctoral and 39 master degree programs in counseling and guidance had been established throughout Turkey (Z. Atik, 2017). The increase in the number of counseling and guidance programs throughout Turkey is an indication of the need of counselors to serve individuals, families, and groups connected with schools as well as those needed within the community. Although the increase in counseling and guidance opportunities within schools is a positive for all involved, there continue to be on-going challenges in Turkey regarding the standardization and accreditation of these programs as well as further issues connected to professional identity and specialization within the counseling field (Aladağ & Kemer, 2016a; Stockton & Yerin Güneri, 2011). As a result, to improve the quality of counselors and counseling programs throughout Turkey, it is critical that counselor education programs be standardized as well as there be an effective supervision system put in place as part of the counseling program curriculum. Importantly, the lack of standardization within counselor education programs reflects negatively on supervision practices. In particular, it is necessary to adjust and/or restructure the supervision of individual counseling sessions within the scope of individual counseling practicum in order for the individual development of counseling students to be closely followed throughout the educational practicum process (Bernard & Goodyear, 2009; Borders & Brown, 2005). During the supervision process, counseling trainees are provided guidance and support in regards to determining the appropriate counseling approaches for client-specific needs as well as developing their own counseling process skills as a way of becoming competent counseling practitioners (Borders, 2009). Also, creating a realistic perception of competence in counselor education is as important as developing the actual counseling skills (Larson & Daniels, 1998; Loganbill, Hardy, & Delworth, 1982), and the supervision process can serve in fostering and developing these perceptions within trainees. For example, counseling trainees being evaluated and receiving constructive feedback during the supervision process in regards to their performance

and use of techniques within the learning and developmental process can ultimately lead to positive effects towards the improvement of trainees' self-efficacy beliefs (Daniels & Larson, 2001; Fernando & Hulse-Killacky, 2005; Ladany, Ellis, & Friedlander, 1999; Meydan, 2015; Pamukçu, 2011; Satıcı & Türküm, 2015).

Over the past 15 years, supervision has emerged as a pressing issue within counselor education throughout Turkey (Aladağ & Kemer, 2016b; G. Atik, Çelik, Güç, & Tutal, 2016; Z. Atik, 2017; Erkan Atik & Yıldırım, 2017; Erkan Atik, Arıcı, & Ergene, 2014; Kemer & Aladağ, 2013; Koçyiğit Özyiğit & İşleyen, 2016; Meydan, 2015; Siviş Çetinkaya & Karaırmak, 2012), however, for a very long time this issue has not been systematically addressed within counselor education programs (Ersever, 1993). In order to become an effective and professional counselor, supervision plays an important role in the personal and professional development of counseling trainees (Borders, 2009). In the United States, where the practice of supervision in counselor education is highly developed, the Council for Accreditation of Counseling and Related Educational Programs (CACREP) (2016) requires supervision in a variety of modalities as a mandatory requirement of counselor education.

In Turkey, since there are a high number of students enrolled in undergraduate counseling programs, there is a growing need for the use of models that focus on group work within the supervision process. As evidence of an increase in the use of group-oriented supervision models, the findings of two national surveys in Turkey conducted with supervisors and counseling trainees, Aladağ and Kemer (2016b) and Z. Atik (2017) found that group supervision is in effect the most commonly used approach for dealing with the high number of counseling trainees. Examples of some of the commonly utilized group supervision methodologies presented in past studies were (Aladağ & Kemer, 2016b; Z. Atik, 2017) authoritative group supervision, supervisory interactive group supervision (Proctor & Inskipp, 2001), and although limited, peer group supervision (Borders, 1991). The group supervision model that counseling trainees in past studies stated benefitting the most from was the use of peer group supervision (Z. Atik, 2017). As a result, peer group supervision models are considered appropriate models of counseling supervision for use with undergraduate counseling trainees in Turkey. There are a variety of peer group supervision models presented in the counseling supervision literature, even though, many of them differ in terms of their types, degree of structuring, and focus (Borders, 2012). For example, some of the supervision models are designed for use with students, while others are primarily utilized for use with practitioners. One peer group supervision model, in particular, the structured peer group supervision format, was introduced by Borders (1991) and is considered to be an appropriate model for use in Turkey because it is a supervision format that supports the professional development of counselors at all levels. In the structured peer group supervision format, a small three to six participant group is supervised by one qualified trainer for $1^{1/2}$ to 3 hour-long weekly or biweekly meetings. At the initial structured peer group supervision meeting, group members identify their learning goals and then during subsequent meetings, the group members take turns presenting recordings of their counseling sessions for review and critique by their peers. Importantly, during these sessions, the peer group members choose or

are assigned specific roles, take on certain perspectives, and/or carry out tasks as part of the recording segment review process. The tasks carried out by the group members may include (a) observing counselor or client nonverbal behavior or a particular counseling skill, (b) assuming the role of the counselor, client, or parent, spouse, friend, teacher, or other significant person in the client's life, (c) viewing the session from a particular theoretical perspective, and (d) creating metaphor for the client, counselor, and counseling process (Borders, 1991).

The structured peer group supervision format not only focuses on developing counseling skills but also provides support for counseling trainees to evaluate their clients from different perspectives as well as to become competent counselors by increasing their mastery of the requisite abilities, knowledge, and interpersonal skills needed in this career pursuit (Borders, 1991). The structured group supervision format has specific strengths that make it an effective and useful system of providing feedback and direction to counseling trainees, for example, with the group-meeting format all group members are provided an opportunity to participate within the sessions and provide their perspective through instructive feedback for their peers. Importantly, the peer feedback provided should be objective, improve cognitive counseling skills, applicable for both experienced and novice counselors, teach self-monitoring, and provide a systematic procedure and framework for supervision (Borders, 1991). Another strength of the peer group supervision format is that its effectiveness is backed up by empirical evidence (Benshoff & Paisley, 1996; Crutchfield & Borders, 1997; Starling & Baker, 2000).

Considering the strengths of the structured peer group supervision format, the aim of this study was to examine the appropriateness and usefulness of the structured peer group supervision format, frequently utilized in university graduate-level counseling programs in the United States, for the instruction and development of undergraduate counseling students within Turkey. In addition, another intention of this study was to determine the usefulness of this supervision approach for increasing the self-efficacy beliefs of counseling trainees within the scope of the courses carried out under the supervision process. More specifically the research questions explored in this study are as follows:

1. What are the perceptions of undergraduate counseling trainees in regards to the structured peer group supervision format?

2. What is the level of improvement of counseling self-efficacy among undergraduate counseling trainees who received structured peer group supervision?

Method

Research Design

This study utilized a sequential exploratory mixed method research design. In this design, first, the qualitative data was collected and analyzed, and then in the subsequent step, the quantitative data was collected and analyzed. The integration of

the qualitative and quantitative data occurred later during the interpretation and discussion stage (Hanson, Creswell, Plano Clark, Petska, & Creswell, 2005). The qualitative aspect of this study was to examine the undergraduate-level counseling trainees' perceptions as they related to the structured peer group supervision format. The quantitative aspect of the study, which utilized a pre-experimental design, focused on investigating the effect of the structured peer group supervision format on the level of improvement in the counseling self-efficacy of the participating undergraduate counseling trainees.

Participants

The participant group for this study was comprised of senior-level university students from a Counseling and Guidance program at a state university located in Ankara, Turkey. In the qualitative portion of this study, the participant group included 16 undergraduate counseling trainees (13 female and 3 male) with ages ranging between 21 to 23 years old (M = 22.31, SD = .70). The quantitative portion of the study included a participant group of 21 undergraduate counseling trainees (14 female and 7 male) with ages ranging between 21 and 23 years old (M = 22.05, SD = .67). In addition, a convenient sampling strategy was employed in both portions of the study.

Measures

The qualitative data was collected via a *demographic information form* to acquire age and gender information, and a *survey form* developed for obtaining the counseling trainees' opinions in regards to their experiences with the structured peer group supervision format. The *survey form* developed for this study included the following six open-ended questions: (a) How did receiving supervision from your peers contribute to your development? (b) What kind of awareness did you gain through your peer group supervision session? (c) How did taking on the roles (i.e., counselor, clients, etc.) or perspectives (i.e., assessment of the client, evaluation of the process, etc.) for providing feedback to your peers contribute to your own development? (d) What are the strengths of structured peer group supervision as compared to individual supervision? (e) What aspects of the structured peer group supervision format do you think should be improved? (f) What do you think about the use of the structured peer group supervision format for the training of undergraduate-level counseling students? It is also important to point out that both the purpose of the study and the related supervision literature were carefully considered when preparing the study's survey questions.

In the quantitative portion of this study, the counseling self-efficacy levels of the trainees were measured through the *Counselor Activity Self-Efficacy Scales (CASES)* (Lent, Hill, & Hoffman, 2003), which was adapted for Turkish by Pamukçu and Demir (2013). The original version of the CASES consists of three sub-domains of counselor self-efficacy including; helping skills self-efficacy, session management self-efficacy, and counseling challenges self-efficacy. Each domain of the CASES consists of a unique number of items, for example, the helping skills self-efficacy domain includes 15 items that assess the helping-counseling skills, the session management self-efficacy domain

consists of 10 items to evaluate the management of specific counseling session tasks, and finally the counseling challenges self-efficacy domain consists of 16 items to assess any relationship conflicts and/or client distress. The CASES is a 41-item measurement tool that includes a description of competencies that counselors perceive for various counseling session tasks and utilizes a 10-point rating scale (0 = no confidence, 9 = complete confidence) (Lent, Hill, & Hoffman, 2003). Besides, for the original form of CASES, the overall Cronbach alpha coefficient was .97. While, the Cronbach alpha coefficients for the sub-domains ranged from .79 to .94 (Lent, Hill, & Hoffman, 2003), and the Cronbach alpha coefficients obtained from the Turkish adaptation study were .98 for the overall scale, which ranged between .92 and .95 for the sub-scales (Pamukçu & Demir, 2013).

Procedure

In the qualitative portion of the study, data were collected over three successive academic years including 2011-12 (n = 8), 2012-13 (n = 2), and 2013-14 (n = 6). The counseling trainees completed the survey forms during the final session of their structured peer group supervision sessions. The survey forms were completed over a period of approximately 30 minutes at a designated university-counseling center. Before collecting any data, the researchers explained the purpose of the study to the participants as well as the importance of their participation in the study. It was ensured that participation was voluntary and that they could discontinue participation in the study at any time without experiencing consequences. Furthermore, the privacy of the counseling trainees was ensured by not inquiring about any personal information on the measures.

The quantitative data for the study were collected during the spring semester of the 2014-15 academic years. A total of three measures relating to the counseling self-efficacy were queried from the counseling trainees. The first measure was conducted at the out-set of the first counselor trainee supervision session, the second measure took place at the out-set of the fourth supervision session, and finally, the last measure occurred at the close of the final supervision session. All of the counselor trainee supervision sessions, as well as the administration of all information and survey forms for this study, took place at a university-counseling center.

The structured peer group supervision sessions, which were performed during the four educational years of the counseling trainees, were carried out under the leadership of the primary author. In each educational year, the trainees were expected to complete at least six counseling sessions with an individual client. The supervision sessions were conducted in line with the steps explained regarding the structured peer group supervision model developed by Borders (1991). For example, during each course period, supervision groups with an average of six members were formed to carry out the structured peer group supervision process. During the first session the students were introduced to the structured peer group supervision format, the responsibilities of the trainees were explained, and the supervision objectives of each trainee were identified. Next, the structured peer group supervision sessions were organized in order to gain feedback from each individual trainee. Finally, a

termination session was held in which the supervision objectives of the trainees as well as the process as a whole were evaluated. Throughout this entire peer group supervision process a qualified supervisor took on the role of the moderator (Borders, 1991). In effect, the role of the moderator was to facilitate the session discussions, share in the roles of others, summarize session feedback, and provide other essential feedback when necessary. Importantly, in order to evaluate the effectiveness and usefulness of this model, only one supervision methodology was utilized throughout the entirety of the process. During the qualitative data collection period of this study, the counseling trainees received both structured peer group supervision as well as individual supervision in order for comparisons to be made across supervision modalities. However, during the period of quantitative data collection, the counseling trainees received only structured peer group supervision as a means of preventing any form of confounding effect that might occur from other supervision modalities.

Data Analysis

This study utilized a mixed research design in order to measure the research aim from a holistic perspective by conducting an in-depth examination of both qualitative and quantitative data (Creswell, 2003; Tashakkori & Creswell, 2007). In the qualitative portion of the study, trainee responses were analyzed through qualitative content analysis as well as an inductive approach was followed in order to reveal new constructs from the data without generating predetermined ideas. Since the phenomena in the scope of this study was being researched for the first in Turkey, the use of the previously mentioned approach was deemed to be the preferred methodology to pursue. Before the content analysis was carried out, the researchers utilizing the Windows Office-Word program transcribed the counselor trainees' written responses verbatim. This transcription process also provided the researchers with an opportunity to become acquainted with the content provided in the written survey forms. In order for the researchers to better organize the counseling trainees' transcribed responses, they were divided into two columns using a Windows Office-Word document, in which the right column was utilized for the categorizing and coding of data. Through researchers' consensus, the themes and categories utilized for the organization of data were identified from the study's raw data as well as from related literature. When the new categories emerged, all of the responses were reviewed and then grouped accordingly. Following the consensus reached by the researchers regarding the themes and categories for data analysis, they were again discussed with one doctoral student as well as one assistant professor from counselor education to reach further agreement consensus. While, in the quantitative portion of the study, any change in the trainees' level of counseling self-efficacy that may have occurred through the structured peer group supervision process was examined by utilizing random effect analysis. The use of the random effect analysis was conducted by performing a linear growth model in HLM7 (Raudenbush, Bryk, & Congdon, 2010). This model is especially useful for short-term measurements (Raudenbush & Bryk, 2002) because it is extremely effective at determining individual differences and developments in terms of self-efficacy levels.

Results

The data analysis results for this study are presented in the following section and separated into two parts, including the qualitative and quantitative analysis. The detailed integration and analysis of both data sets are presented later in the discussion section.

Results of Qualitative Part

The perceptions of the counseling trainees were clustered into four main themes: (1) contributions of the model, (2) strengths of the model compared to individual supervision, (3) aspects of the model that need improvement, and (4) applicability of the model for use with undergraduate-level counseling trainees. The categories and subcategories for the counseling trainees' perceptions as they relate to the use of and participation in the structured peer group supervision format are provided in the following (*see* Table 1).

Table 1
Trainees' Perceptions of Structured Peer Group Supervision

Themes	*Categories/Subcategories*
Contributions of the model	Increase in counseling competencies and self-awareness (*n* =15) Different perspectives (*n* = 15) Supportive, constructive, instructive, and self-esteem building environment (*n* = 13) Working on different cases and gaining experience (*n* = 12) Seeing similar experiences and normalization (*n* = 9) Finding the process effective and benefitting from the process (*n* = 10) Being active in the process (*n* = 7) Enjoying the process (*n* = 5)
Strengths of the model compared to individual supervision	Different perspectives and working on different cases (*n* = 13) Seeing similar experiences and normalization (*n* = 9) Receiving more feedback and gaining awareness (*n* = 7) A more comfortable process (*n* = 5)
Aspects of the model that need improvement	Recommendations for the model (*n* = 7) *Use of the roles in the process (n = 4)* *Giving feedback (n = 3)* *The role of supervisor (n = 2)* *Preparation for the supervision (n = 1)* Issues for the implementation of the model (*n* = 12) *Time and duration issues (n = 6)* *Not being able to receive more group supervision (n = 4)* *Place/setting problems (n = 2)* *Nonattendance (n = 1)* *Unable to collect sufficient data (n = 1)*

Table 1 continued

Trainees' Perceptions of Structured Peer Group Supervision

Themes	*Categories/Subcategories*
Applicability of the model for undergraduate counseling students	Definitely used (n = 16)

Contributions of the Model

The counseling trainees highlighted that use of the structured peer group supervision model contributed to increases in their counseling competencies and self-awareness. The trainees also reported that the group supervision process significantly contributed to improvements in their counseling skills (i.e., focusing, empathy, reflection of contents and feelings, asking questions, confrontation, etc.), case conceptualization skills, knowledge regarding counseling theories, techniques, and interventions as well as their self-awareness in relation to identifying their strengths and aspects of the process that they believe need improvement. For example, one trainee stated, *"I checked and was sure through group supervision whether or not I was applying the ABC model in the right way. I noticed that I could find appropriate metaphors for counselors or clients. I noticed how to focus on a client and handle his or her statements while applying a cognitive disputing technique. I also developed my skill of constructive criticism."*

The trainees emphasized that the process of role taking provided them with opportunities to look at the situation from differing perspectives as well as hear differing opinions from group peers. Also, the structured peer group supervision model appeared to make a significant contribution in terms of working with clients who expressed they had unique concerns. Furthermore, trainees stated structured peer group supervision provided an environment that was supportive, constructive, instructive, and self-esteem building. Also, the counseling trainees noted that peer group supervision required, as participants, to play an active part in the process. Finally, several of the trainees pointed out that parts of group supervision format were enjoyable, for example, taking on the metaphor role.

Strengths of the Model Compared to Individual Supervision

When the trainees were queried about the strengths of the model in comparison to individual supervision, they frequently reported that the structured peer group supervision model provided opportunities to gain different perspectives as well as allowing them to work on cases with differing clients. For example, one trainee stated, *"We were six people in the group supervision and this meant six different perspectives, six different solutions to the problem. This also meant six different pair of eyes given to us to see the things we cannot see."* In addition, most of the trainees stated that they felt they were not alone in the peer group supervision process, generally made the same mistakes, and had the same concerns regarding the assisting process. Also, they normalized their experiences and concerns in regards to the supervision. Another strength of the

structured peer group supervision model emphasized by some of the trainees was that they were able to receive a greater level of feedback as well as gain a deeper awareness of the situation. To substantiate this sentiment, one trainee stated, *"If we think about fish as gains and awareness, individual supervision is similar to fishing with a fishing rod; whereas, peer group supervision is similar to fishing with fishing nets."* Finally, several of the counseling trainees reported that structured peer group supervision, in comparison to individual supervision, was more comfortable in terms of the relationship dynamics that occurred through the counseling process.

Aspects of the Model that Need Improvement

In terms of the aspects of the model that might need improvement, the trainees provided some recommendations for improving the structured peer group supervision format as well as suggestions for improvements to the implementation process. In regards to improving the use of structured peer group supervision, several trainees recommended adding further roles to the model, and as a result, having counseling trainees focus on multiple roles instead of just one. Although no one experienced this problem, several of the trainees pointed out that providing peer feedback is a critical aspect of the process, and they believe if constructive feedback is not provided, there could be deleterious damage to group dynamics. Surprisingly, in terms of the supervisor's role in the process, one trainee commented that they felt the supervisor should act more as a director instead of providing the more democratic approach that they experienced in their supervision experience. Another counseling trainee reported that they felt it could be beneficial if they received a variety of forms of explanation about how to prepare for the session before attending and receiving counseling supervision.

As far as the implementation process of the group supervision model, the most frequent weaknesses that trainees emphasized were the limited amount of time allotted for peer group supervision, unavailability of the supervisors' time, and not receiving a greater amount of group supervision. The least frequently reported problems for the implementation process were problems with the place and/or setting, nonattendance, and being unable to collect a sufficient amount of data to provide feedback.

Applicability of the Model for Undergraduate Counseling Students

All of the trainees queried in this study provided favorable responses in regards to the use of the structured peer group supervision format as a means of educating undergraduate counseling trainees within Turkey. An example of this was provided by one trainee who expressed, "*Group supervision must be used in the education of counseling trainees because it saves the trainees from crawling into uncertainty and provides an opportunity for the trainees to move in the right direction.*"

Results of Quantitative Part

The results of the qualitative data analysis indicated that the average counseling self-efficacy score from the first round of testing was calculated at $(\hat{\beta}_{00})$ 132.364 logits, and the trainees gained an average $(\hat{\beta}_{10})$ 46.846 logits per measurement over the study

duration (*see* Table 2). The counseling trainees varied significantly in counseling self-efficacy scores at the beginning of the structured peer group supervision process (χ^2 = 82.993, *df* = 20, $p < .001$). Besides, there was a significant variation among the trainees' growth rates after the beginning of structured peer group supervision process (χ^2 = 55.186, *df* = 20, $p < .001$). Initial reliability coefficients indicated that there was a significant variation in the data regarding individual differences in both the initial status and the growth rates. When the correlation between individual change and initial status was examined, the progress of counseling trainees who exhibited low counseling self-efficacy scores at the beginning of structured peer group supervision process was higher than those who had high counseling self-efficacy scores at the beginning of the same process.

Table 2
Linear Model of Growth in the Counseling Self-Efficacy of Undergraduate Trainees

Fixed Effect	*Coefficient*	*SE*	*t ratio*	
Mean initial status, β_{00}	132.364	14.204	9.318	
Mean growth rate, β_{10}	46.846	5.570	8.410	
Random Effect	*Variance Component*	*df*	χ^2	*p-value*
Initial status, r_{0i}	3195.722	20	82.993	<0.001
Growth rate, r_{1i}	405.826	20	55.186	<0.001
Level-1 error, e_{ti}	418.284			
Reliability of OLS Regression Coefficient Estimate				
Initial status, π_{0i}	0.743			
Growth rate, π_{1i}	0.612			

Discussion, Conclusion, and Recommendations

In this study, the experiences of undergraduate counseling trainees related to the use of structured peer group supervision, and the effect of this supervision format on the growth of the counseling self-efficacy beliefs of trainees were examined. The qualitative portion of this study revealed that the structured peer group supervision format contributed significantly to the professional development of the trainees. The aspects of the group supervision format that contributed to the counseling trainees' development were an increase in their competencies and self-awareness; gaining differing perspectives; being in a constructive, instructive, and supportive environment that increased self-esteem; working on a variety of cases and gaining

experience; witnessing similar experiences and normalizing those experiences; taking an active role in the process; and finally, overall enjoyment and satisfaction. Consistent with the findings of this study, Borders et al. (2012) reported that group supervision provided multiple perspectives to trainees, helped them to normalize their experiences, and exposed them to multiple counselor styles. In addition, there were explanations regarding the fact that peer group supervision does have merits for group members in terms of providing a safe environment, encouraging process participation, providing goal-oriented and constructive feedback, developing cognitive counseling skills, and teaching group members to critique themselves (Borders, 1991). Gillam, Hayes, and Paisley (1997), pointed out that trainees benefit from peer group supervision by experiencing a variety of counseling and personal styles, learning from their peers' presentations, adapting what they have learned to their own practices, and learning to provide appropriate feedback.

Supporting the qualitative findings relating to the contributions of the structured group supervision model, the quantitative findings of this study indicated that structured peer group supervision positively affected the counseling self-efficacy beliefs of trainees, and the counseling self-efficacy levels of the trainees showed increases with each measurement. Counseling self-efficacy refers to a counselor's judgments about his or her ability to conduct effective counseling with his or her clients. These judgments are generally related to aspects of counseling such as having effective helping skills, conducting sessions in line with the purpose of each session, and dealing with challenging clients (Larson & Daniels, 1998). According to Bandura (1997), self-efficacy beliefs about performing tasks are fed by four sources: performance accomplishments, vicarious experiences, verbal persuasion, and physiological states or emotional arousal. To have positive counseling self-efficacy beliefs there is a need have these experiences, and the structured peer group supervision environment can provide an opportunity for trainees to experience performance accomplishments as well as provide possibilities for vicarious experiences provided through feedback from others. However, group member feedback following the watching and/or listening of peers' recorded counseling sessions can have a critical impact on trainee development (Borders, 1991; Borders & Brown, 2005), and as a result, the activation of verbal persuasion can reach a high level. The support of the physiological states or emotional arousal which makes up the final source of self-efficacy, through the supervision process the emotional state of trainees as well as how to cope with trainees' concerns are covered in the supervision process. When all the self-efficacy sources are evaluated, it is clear that the structured peer group supervision process does provide multidimensional support for the development of trainees. In agreement with this assertion, many studies describing the contributions of the structured peer group supervision format also posited that this supervision approach does lead to increases in the counseling self-efficacy of counseling trainees (Cashwell & Dooley, 2001; Crutchfield & Borders, 1997).

One of the remarkable findings of this study was that the trainees starting with lower counseling self-efficacy beliefs had greater increases in their end of study counseling self-efficacy beliefs than those trainees starting with higher counseling self-

efficacy beliefs. From previous studies it is known that novice-counseling trainees usually experience higher anxiety and lower perceived self-efficacy (Daniels & Larson, 2001; Larson & Daniels, 1998). In particular, it can be postulated that structured peer group supervision can be an important normalization and recovery process for trainees who do not have a realistic perception of self-efficacy due to environmental features. As a result, trainees with low counseling self-efficacy at the beginning of the supervision process may experience this normalization more actively in comparison to their peers with high counseling self-efficacy.

The trainees emphasized what they believed were the strengths of the structured peer group supervision format in comparison to the individual supervision approach, for example, the multiple perspectives gained and ability to work on different cases, witnessing similar experiences and the normalization of these experiences, receiving a higher level of feedback and gaining greater awareness, and overall feeling of being more comfortable throughout the supervision process. According to these findings, the structured peer group supervision model, in comparison to individual supervision, provided additional contributions to the development of the trainees' knowledge and skills related to counseling. In Borders et al. (2012), the strengths of structured peer group supervision compared to the individual supervision were in line with the findings of this study. This consistency indicates that the model has the same influence on both undergraduate and graduate counseling trainees. As a result, since individual supervision allows for one-on-one focus with each trainee and considering the trainee's dependence on the supervisor (Bernard & Goodyear, 2009; Stoltenberg & McNeill, 1997), the use of both supervision modalities, individual and structured peer group supervision, should be deemed both complementary and beneficial for the education of counseling trainees.

The trainees' opinions regarding the improvements that could be made to the group supervision model were limited in number. A few trainees who did recommend adding more roles, providing feedback in a more constructive way, recognizing the importance of the supervisor's role, and provided opportunities for trainee preparation prior to the supervision process. The trainees did report problems they believed were associated with the implementation process of the model. For example, in addition to the roles already offered in the group supervision model, the use of narrative approaches for increasing cognitive skills as well as found poetry (Waalkes, DeCino, & Borders, 2017) and metaphor can also provide important contributions to this supervision modality. Furthermore, during the supervision process, the supervisor adopted the role of moderator and provided feedback that the trainees believed was beneficial. It is believed that this scenario can contribute to the transfer of positive experiences related to the group supervision process, especially for those undergraduate counseling trainees who gain an opportunity to act more independently.

Overall, the results from this study suggested that the structured peer group supervision format is a very practical and useful approach for the education of undergraduate-level counseling students. All trainees queried in this study recommended that the use of the group supervision model for counselor education

programs in Turkey is an appropriate decision. Providing supervision to students enrolled in Turkish university undergraduate-level counseling programs faces a big challenge due to the overwhelming number of students enrolled in these programs. Therefore, the structured peer group supervision approach could be an effective means of providing supervision for all of the counseling trainees in an efficient and productive manner. In addition, results of this study reveal that it can be useful to carry out supervisor training for the dissemination of the structured peer group supervision method as well as to teach this counseling supervision approach as part of doctoral-level counseling supervision course.

Within the scope of this study, the limitations of using a pre-experimental research design were recognized. As a result, in order to control any deleterious factors that might affect the development of the counseling trainees' self-efficacy levels during the supervision process, it is recommended that future studies utilize a control group as well as randomly select and assign participants into experimental and control groups. Also, it can be useful to include variables related to the supervision process that can ultimately aid in the counseling self-efficacy development of trainees (i.e., supervisory working alliance, trainee developmental levels, and trainee perceptions regarding supervision styles). Finally, it is also believed that it is important to utilize larger sample sizes in future studies as a means of better generalizing research findings.

References

Aladağ, M., & Kemer, G. (2016a). Clinical supervision: An emerging counseling specialty in Turkey. *The Clinical Supervisor, 35*(2), 175–191. doi: 10.1080/07325223.2016.1223775

Aladağ, M., & Kemer, G. (2016b). *Psikolojik danışman eğitiminde bireyle psikolojik danışma uygulaması ve süpervizyonunun incelenmesi [Investigation of individual counseling practicum and supervision in counselor education]*. Ege University Scientific Research Project Report, Izmir, Turkey.

Atik, G., Çelik, E. G., Güç, E., & Tutal, N. (2016). Psikolojik danışman adaylarının yapılandırılmış akran grup süpervizyonu sürecindeki metafor kullanımına ilişkin görüşleri [Opinions of counseling students about the use of metaphor in the structured peer group supervision process]. *Ege Eğitim Dergisi, 17*(2), 597–619. doi: 10.12984/egeefd.280759

Atik, Z. (2017). *Psikolojik danışman adaylarının bireyle psikolojik danışma uygulaması ve süpervizyonuna ilişkin değerlendirmeleri [Counselor candidates' evaluation of individual counseling practicum and supervision]* (Unpublished doctoral dissertation). Hacettepe University, Ankara, Turkey.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. NY: W. H. Freeman and Company.

Benshoff, J. M., & Paisley, P. O. (1996). The structured peer consultation model for school counselors. *Journal of Counseling & Development, 74*, 314–318. doi: 10.1002/j.1556-6676.1996.tb01872.x

Bernard, J. M., & Goodyear, R. K. (2009). *Fundamentals of clinical supervision* (Fourth ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Borders, L. D. (1991). A systematic approach to peer group supervision. *Journal of Counseling & Development, 69*(3), 248–252. doi: 10.1002/j.1556-6676.1991.tb01497.x

Borders, L. D. (2009). A primer on counseling supervision. In F. Korkut Owen, R. Özyürek, & D. Owen (Eds.), *Gelişen psikolojik danışma ve rehberlik: Meslekleşme sürecinde ilerlemeler [Emerging psychological counseling and guidance: Advances in the process of professionalization]* (pp. 35-49). Ankara: Nobel Yayın Dağıtım.

Borders, L. D. (2012). Dyadic, triadic, and group models of peer supervision/consultation: What are their components, and is there evidence of their effectiveness. *Clinical Psychologist, 16*, 59–71. doi: 10.1111/j.1742-9552.2012.00046.x

Borders, L. D., & Brown, L. L. (2005). *The new handbook of counseling supervision* (Second ed.). NJ: Lawrence Erlbaum Association.

Borders, L. D., Welfare, L. E., Greason, P. B., Paladino, D. A., Mobley, A. K., Villalba, J. A., & Wester, K. L. (2012). Individual and triadic and group: Supervisee and supervisor perceptions of each modality. *Counselor Education & Supervision, 51*, 281–295. doi: 10.1002/j.1556-6978.2012.00021.x

Cashwell, T. H., & Dooley, K. (2001). The impact of supervision on counselor self-efficacy. *The Clinical Supervisor, 20*(1), 39–47. doi: 10.1300/j001v20n01_03

Council for Accreditation of Counseling and Related Educational Programs (CACREP). (2016). *CACREP accreditation manual*. Alexandria, VA: Author.

Council of Higher Education. (2018). *Lisans tercih sihirbazı [Undergraduate program preference wizard]*. Retrieved from https://yokatlas.yok.gov.tr/

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed method approaches* (Second ed.). USA: Sage Publications.

Crutchfield, L. B., & Borders, L. D. (1997). Impact of two clinical peer supervision models on practicing school counselors. *Journal of Counseling & Development, 75*, 219–230. doi: 10.1002/j.1556-6676.1997.tb02336.x

Daniels, J. A., & Larson, L. M. (2001). The impact of performance feedback on counseling self-efficacy and counselor anxiety. *Counselor Education & Supervision, 41*, 120–130. doi: 10.1002/j.1556-6978.2001.tb01276.x

Erkan Atik, Z., & Yıldırım, İ. (2017). Süpervizyonda Değerlendirme Süreci Envanteri Türkçe Formunun geçerlik ve güvenirlik çalışması [Validity and reliability of

Turkish Form of the Evaluation Process within Supervision Inventory]. *Ege Eğitim Dergisi, 1*(18), 146–173. doi: 10.12984/egeefd.328380

Erkan Atik, Z., Arıcı, F., & Ergene, T. (2014). Süpervizyon modelleri ve modellere ilişkin değerlendirmeler [Supervision models and assessment of these models]. *Türk Psikolojik Danışma ve Rehberlik Dergisi, 5*(42), 305–317. Retrieved from http://www.pdrdergisi.org/

Ersever, Ö. H. (1993). Psikolojik danışmada süpervizyon standartları [Supervision standards in counseling]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 2*(26), 609–619. Retrieved from http://dergipark.gov.tr/auebfd

Fernando, D. M., & Hulse-Killacky, D. (2005). The relationship of supervisory styles to satisfaction with supervision and the perceived self-efficacy of master's level counseling students. *Counselor Education & Supervision, 44*, 293–305. doi: 10.1002/j.1556-6978.2005.tb01757.x

Gillam, S. L., Hayes, R. L., & Paisley, P. O. (1997). Group work as a method of supervision. In S. Gladding (Ed.), *New developments in group counseling* (pp. 78–82). Greensboro, NC, ERIC/CASS Publications.

Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224–235. doi: 10.1037/0022-0167.52.2.224

Kemer, G., & Aladağ, M. (2013, September). *Psikolojik danışman eğitiminde bireyle psikolojik danışma uygulaması ve süpervizyonu: Ulusal bir tarama çalışması. [Counseling practicum and supervision in counselor education: A national survey]*. Paper presented at the World Congress of Psychological Counseling and Guidance, Istanbul, Turkey.

Koçyiğit Özyiğit, M., & İşleyen, F. (2016). Psikolojik danışmada süpervizör eğitimi [Supervisor training in counseling]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 16*(4), 1813–1831. Retrieved from http://dergipark.gov.tr/aibuefd/

Ladany, N., Ellis, M. V., & Friedlander, M. L. (1999). The supervisory working alliance, trainee self-efficacy and satisfaction. *Journal of Counseling & Development, 77*, 447–455. doi: 10.1002/j.1556-6676.1999.tb02472.x

Larson, L. M., & Daniels, J.A. (1998). Review of the counseling self-efficacy literature. *The Counseling Psychologist, 26*(2), 179–218. doi: 10.1177/0011000098262001

Lent, R. W., Hill, C. E., & Hoffman, M. A. (2003). Development and validation of the Counselor Activity Self-Efficacy Scales. *Journal of Counseling Psychology, 50*(1), 97–108. doi: 10.1037/0022-0167.50.1.97

Loganbill, C., Hardy, E., & Delworth, U. (1982). Supervision: A conceptual model. *The Counseling Psychologist, 10*, 3–42. doi: 10.1177/0011000082101002

Meydan, B. (2015). Bireyle psikolojik danışma uygulamasında mikro beceri süpervizyon modelinin etkililiğinin incelenmesi [Examining the effectiveness of microcounseling supervision model in individual counseling practice]. *Türk Psikolojik Danışma ve Rehberlik Dergisi, 5*(43), 55–68. Retrieved from http://www.pdrdergisi.org/

Pamukçu, B. (2011). *The investigation of counseling self-efficacy levels of counselor trainees* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.

Pamukçu, B., & Demir, A. (2013). Psikolojik Danışma Öz-Yeterlik Ölçeği Türkçe Formu'nun geçerlik ve güvenirlik çalışması [The validity and reliability study of the Turkish version of Counseling Self-Efficacy Scale]. *Türk Psikolojik Danışma ve Rehberlik Dergisi, 5*(40), 212–221. Retrieved from http://www.pdrdergisi.org/

Proctor, B., & Inskipp, F. (2001). Group supervision. In J. Scaife (Ed.), *Supervision in mental health professions. A practitioner's guide* (pp. 99–121). London: Brunner-Routledge.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Second ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S., Bryk, T., & Congdon, R. (2010). *HLM 7 hierarchical linear and nonlinear modeling* [Software]. Lincolnwood, IL: Scientific Software International.

Satıcı, B., & Türküm, A. S. (2015, October). *Psikolojik danışman adaylarının psikolojik danışma özyeterlik algılarının süpervizyon yaşantıları ve mizah tarzları açısından incelenmesi [Investigation of counseling self-efficacy beliefs of counselor trainees in terms of supervision experiences and humor styles]*. Paper presented at the XIII. National Psychological Counseling and Guidance Congress, Mersin, Turkey.

Siviş Çetinkaya, R., & Karaırmak, Ö. (2012). Psikolojik danışman eğitiminde süpervizyon [Supervision in counselor education]. *Türk Psikolojik Danışma ve Rehberlik Dergisi, 4*(37), 107–121. Retrieved from http://www.pdrdergisi.org/

Starling, P. V., & Baker, S. B. (2000). Structured peer group practicum supervision: Supervisees' perceptions of supervision theory. *Counselor Education and Supervision, 39*, 162–176. doi: 10.1002/j.1556-6978.2000.tb01229.x

Stockton, R., & Yerin Güneri, O. (2011). Counseling in Turkey: An evolving field. *Journal of Counseling and Development, 89*, 98–104. doi: 10.1002/j.1556-6678.2011.tb00065.x

Stoltenberg, C. D., & McNeill, B. W. (1997). Clinical supervision from a developmental perspective: Research and practice. In C. E. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 184–202). New York: Wiley.

Tan, H. (1986). *Psikolojik danışma ve rehberlik [Counseling and guidance]*. Istanbul: Milli Eğitim Basımevi.

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research, 1*, 3–7. doi: 10.1177/2345678906293042

Waalkes, P. L., DeCino, D. A., & Borders, L. D. (2017): The found poet: A new role for the structured peer group supervision model. *Journal of Poetry Therapy, 31*(1), 15–25. doi: 10.1080/08893675.2018.1396727

Psikolojik Danışman Adaylarının Yapılandırılmış Akran Grup Süpervizyonuna İlişkin Algıları ve Deneyimleri: Karma Desenli Bir Çalışma

Atıf:

Atik, G., & Erkan Atik, Z. (2019). Undergraduate counseling trainees' perceptions and experiences related to structured peer group supervision: A mixed method study. *Eurasian Journal of Educational Research, 82*, 101-120, DOI: 10.14689/ejer.2019.82.6

Özet

Problem Durumu: Türkiye'deki psikolojik danışman eğitiminin niteliğinin artırılması için lisans düzeyindeki psikolojik danışma ve rehberlik eğitim programlarının içerisine etkili süpervizyon uygulamalarının dâhil edilmesinin gerekliliği yaklaşık 15 yıllık bir süredir tartışılagelmektedir. Öte yandan eğitim programlarının standart bir uygulama deneyimi sunamaması, psikolojik danışman adaylarının farklı niteliklerde ve yeterlilikte yetişmesine yol açmaktadır. Bu durum aynı zamanda eğitim programları arasında, üzerinde uzlaşılan standart bir uygulama ve süpervizyon alt yapısının kurulmasını güçleştirmektedir. Psikolojik danışman adaylarının, danışanlarına özgü ihtiyaçları ve uygun yaklaşımları belirleyebilmelerine ve süreç becerilerini geliştirmelerine destek olmak, uygulamaya dayalı derslerin süpervizyon eşliğinde yürütülmesi ile mümkündür. Özellikle psikolojik danışman adaylarının süreç boyunca bireysel gelişimlerini yakından takip edebilme imkânı sağlayan, bireyle psikolojik danışma uygulamaları dersleri kapsamında yapılacak psikolojik danışma oturumlarına dönük süpervizyonun yapılandırılması gerekmektedir. Söz konusu yapılandırma, kullanılacak süpervizyon yöntemlerini, psikolojik danışman adaylarının ihtiyaçlarına ve programın özelliklerine göre biçimlendirmeyi kapsamaktadır. Amaçlı şekilde seçilecek olan yöntemler psikolojik danışman adaylarının gelişimsel ihtiyaçlarına uygun olarak kullanıldığında özyeterlik algısını güçlendirme üzerinde olumlu bir etkiye sahiptir.

Süpervizyon yöntemleri bireysel, üçlü ve grup olmak üzere farklı formatlarda kullanılabilmektedir. Özellikle Türkiye'de lisans eğitimi ile psikolojik danışman yetiştiren bir eğitim sistemine sahip olunduğu göz önünde bulundurulduğunda, hem öğrenci mevcutlarının fazlalığı ile baş etmek hem de nitelikli süpervizyon vermek

adına grup süpervizyonu formatları arasında yer alan yapılandırılmış akran grup süpervizyonu oldukça etkili bir yöntem olarak ön plana çıkmaktadır. Yapılandırılmış akran grup süpervizyonu süreci, psikolojik danışman adaylarının temel becerilerini ve kavramsallaştırma yeteneklerini geliştirmeye yönelik bir yöntem olmakla birlikte adayların sürece katılımlarını artıran, yapıcı geri bildirim vermelerine ve kendilerini gözlemlemelerine olanak tanıyan bir süreçtir. Yöntemin söz konusu güçlü yanları dikkate alındığında, Türkiye'deki uygulanabilirliği ve psikolojik danışman adaylarının gelişimlerine etkisi ile ilgili çalışmalara ihtiyaç duyulmaktadır.

Araştırmanın Amacı: Bu çalışmada, lisans düzeyindeki psikolojik danışman adaylarının yapılandırılmış akran grup süpervizyonu formatına ilişkin görüşlerinin ve bu süpervizyon formatının, adayların psikolojik danışma özyeterlik düzeyleri üzerindeki etkisinin incelenmesi amaçlanmıştır.

Araştırmanın Yöntemi: Karma araştırma modellerinden keşfedici ardışık desenin kullanıldığı bu çalışmada önce nitel veriler, ardından nicel veriler elde edilmiştir. Araştırmanın nitel bölümünde katılımcıların söz konusu süpervizyon formatına ilişkin görüşleri alınmış; nicel bölümünde ise, yarı deneysel araştırma deseni kullanılarak psikolojik danışman adaylarının özyeterlik düzeylerinin süpervizyon süreci boyunca gelişimi incelenmiştir. Araştırmaya Ankara'daki bir devlet üniversitesinde öğrenim gören psikolojik danışman adayları katılmıştır. Araştırmanın nitel bölümüne 2011-2012, 2012-2013, 2013-2014 Eğitim ve Öğretim yıllarında bireysel psikolojik danışma uygulaması dersini alan 16 öğrenci; nicel bölümüne ise 2014-2015 Eğitim ve Öğretim yılında aynı dersi alan toplam 21 öğrenci katılmıştır. Araştırmanın nitel bölümünde demografik bilgi formu ile birlikte altı açık uçlu sorunun yer aldığı soru formu kullanılmıştır. Soru formunda; akranlardan süpervizyon almanın katkıları, yapılandırılmış akran grup süpervizyonunda edinilen farkındalıklar ve kazanımlar, yapılandırılmış akran grup süpervizyonunda rollere girerek dönüt verme ve alma, yapılandırılmış akran grup süpervizyonunun bireysel süpervizyona göre güçlü yönleri, yöntemin yetersiz ve geliştirilmesi gereken yönleri ile bu yöntemin psikolojik danışman eğitiminde kullanılmasına yönelik katılımcı görüşlerini değerlendirmeye yönelik sorulara yer verilmiştir.

Araştırmanın nicel bölümünde ise psikolojik danışman adaylarının yapılandırılmış akran grup süpervizyonu süresince özyeterlik düzeylerindeki değişimi ölçmek üzere Psikolojik Danışma Özyeterlik Ölçeği kullanılmıştır. Ölçme aracı yardım becerileri özyeterliği, oturum yönetmeye ilişkin özyeterlik ve psikolojik danışma sürecindeki zorluklara ilişkin özyeterlik olarak adlandırılan üç faktörlü bir yapıya sahiptir. Toplam 41 maddeden oluşan ölçme aracı 10'lu derecelendirme üzerinden yanıtlanmaktadır. Ölçme aracından alınan yüksek puanlar yüksek özyeterlik algısına işaret etmektedir. Nitel verilerin analizi, içerik analizi yöntemi kullanılarak gerçekleştirilmiştir. Elde edilen bulgular tema ve kategoriler altında sunulmuştur. Psikolojik danışman adaylarının özyeterlik düzeylerinin yapılandırılmış akran grup süpervizyonu süresince değişimini ölçmek amacıyla, yapılandırılmış akran grup süpervizyonu süresince üç aşamalı ölçüm elde edilmiştir. Elde edilen veriler, HLM7 paket programı kullanılarak doğrusal gelişimsel model ile analiz edilmiştir.

Araştırmanın Bulguları: Araştırmanın nitel bulgularına göre, psikolojik danışman adaylarının yapılandırılmış akran süpervizyonuna ilişkin görüşleri dört tema altında toplanmıştır. Bu temalar; 1) modelin katkıları, 2) bireysel süpervizyona göre modelin güçlü yanları, 3) modelin geliştirilmesi gereken yanları ve 4) modelin lisans düzeyindeki psikolojik danışman adayları ile kullanımıdır. Modelin katkıları temasındaki öne çıkan kategoriler arasında psikolojik danışmaya ilişkin becerileri ve öz farkındalığı artırma, farklı bakış açılarından bakabilme, destekleyici, yapıcı, yönlendirici ve benlik algısını geliştirici bir ortam sağlama, farklı vakalar üzerinde çalışabilme ve deneyim elde edebilme yer almaktadır. Bireysel süpervizyona göre modelin güçlü yanları temasında, farklı perspektiflerden bakabilme ve farklı vakalar üzerinde çalışabilme ile akranlar arası benzer deneyimlerden geçme ve normalleştirme kategorileri öne çıkmaktadır. Modelin geliştirilmesi gereken yanları teması, modele ilişkin öneriler ve modelin uygulama koşullarına ilişkin konular kategorilerini içermekte ve bu kategoriler içerisinde süpervizyondaki roller, geri bildirim verme, süre ve mekâna ilişkin çeşitli alt kategoriler yer almaktadır. Son olarak, modelin lisans düzeyindeki psikolojik danışman adayları ile kullanımında, tüm katılımcıların ortak olarak ifade ettiği kesinlikle kullanılması gerektiğini kapsayan tek bir kategori yer almaktadır. Nicel bölüme ilişkin bulgular ise yapılandırılmış akran grup süpervizyonunun psikolojik danışman adaylarının psikolojik danışmaya ilişkin özyeterliklerini olumlu yönde etkilediğine işaret etmektedir. Doğrusal gelişimsel model çerçevesinde elde edilen bulgulara göre, katılımcılara ait özyeterlik başlangıç güvenirlik katsayıları, hem başlangıç aşaması hem de gelişim oranlarındaki bireysel farklılıklar açısından veri setinde anlamlı bir farklılığın olduğuna işaret etmektedir. Değişimin başlangıç aşaması ile olan ilişkisi (-.878) incelendiğinde, yapılandırılmış akran grup süpervizyonu süreci başlamadan önce psikolojik danışma özyeterlik düzeyleri düşük olan öğrencilerin ilerlemeleri, psikolojik danışma özyeterlik düzeyleri başlangıçta daha yüksek düzeyde olanlardan daha fazladır.

Araştırmanın Sonuçları ve Önerileri: Araştırmanın sonuçları yapılandırılmış akran grup süpervizyonunun uygulamada pratik ve lisans düzeyindeki psikolojik danışman eğitimi için faydalı olduğunu ortaya koymakta ve tüm katılımcılar, yöntemin lisans düzeyinde uygulanabilir olduğunu ifade etmişlerdir. Aynı zamanda yapılandırılmış akran grup süpervizyonu süreci psikolojik danışman adaylarının psikolojik danışma özyeterlik algılarını olumlu bir şekilde etkilemiş ve her bir ölçümde öğrencilerin psikolojik danışma özyeterlik düzeyleri artış göstermiştir. Türkiye'deki psikolojik danışman eğitimi programlarındaki öğrenci sayılarının yüksekliğine rağmen etkili, kullanışlı ve öğrencilerin özyeterliklerini geliştirmeye dönük katkı sağlayan bu yöntemin kullanımının yaygınlaştırılması için süpervizör eğitimlerinin gerçekleştirilmesi önemli bir adım olacaktır.

Anahtar Kelimeler: Psikolojik danışmada süpervizyon, psikolojik danışman eğitimi, yapılandırılmış akran grup süpervizyonu, karma araştırma deseni, doğrusal gelişimsel model

Eurasian Journal of Educational Research 82 (2019) 121-136

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

The Development of an Instrument to Assess Visuo-Semiotic Reasoning in Biology

Lindelani MNGUNI[1]

ARTICLE INFO

Article History:
Received: 16 Feb. 2019
Received in revised form: 17 Jun. 2019
Accepted: 15 Jul. 2019
DOI: 10.14689/ejer.2019.82.7

Keywords
Molecular biology; Visual literacy; Visuo-semiotic models; Visuo-semiotic reasoning

ABSTRACT

Purpose: The significance of visuo-semiotic models in biology education has increased. Students have to develop visuo-semiotic skills, which could enable them to learn biology effectively. However, a lack of a universal theory of visual literacy has made it challenging to develop and assess visualization skills, including visuo-semiotic skills. The aim of the present research, therefore, was to develop an instrument for assessing visuo-semiotic reasoning in biology (VSR-b) in the context of amino acid structures. The research question guiding the research was "how could an instrument for assessing visuo-semiotic reasoning in biology be developed?"

Methods: Guided by a theoretical framework, the VSR-b Test was developed using a mixed-methods approach, by first identifying VSR-b Skills through a panel of nine experts after which items were designed and validated through the same panel of experts and pilot participants (n=18). The VSR-b Test was then tested on a group of molecular biology students (n=30).

Findings: Results showed satisfactory reliability and inter-item correlation. However, further research is required to corroborate findings of the present research in other contexts, with particular emphasis on assessment and development of visuo-semiotic reasoning among students.

Implications for research & practice: The current research has shown that VSR-b can be understood and assessed within the context of the theoretical cognitive process of visualization. It provides teachers and researchers a starting point in understanding how learning occurs through visuo-semiotic models. Instructional and curriculum designers, therefore, can use findings of this research as a guide to support student development in biology.

[1] Department of Science & Technology Education, College of Education, University of South Africa, Pretoria, SOUTH AFRICA. mngunle@unisa.ac.za ORCID: http://orcid.org/0000-0002-0361-0002

Introduction

Researchers around the world acknowledge the significance of visuo-semiotic models (VSM) in teaching and learning biology. VSMs are defined as visual models that use discipline-specific semiotics to represent scientific phenomena for research, teaching, and learning. These include written language, static multidimensional images, animations, simulation, and symbols. VSMs are used in biology because some biological content knowledge and phenomena exist at multiple microscopic levels, which cannot be visualized with a naked eye.

Consequently, VSMs are then used to visually illustrate and represent these otherwise unobservable phenomena and related scientific theories (Arnerson & Offerdahl, 2018). Scholars argue that in the 21st century, the concept of scientific literacy must include the ability to interpret, construct, transform, and evaluate different scientific representations which include VSMs (Nitz, Ainsworth, Nerdel & Prechtl, 2014). For instance, students are now expected to develop 21st-century skills, content literacy, academic communication literacy, science literacy, and visual literacy (Arsad, Osman & Soh, 2011; Mnguni, 2014). This is because the ability to construct and interpret visual scientific information has become a significant skill in modern science.

While scientists agree on the significance of VSMs and related competencies, there remain debates related to the definition of visual literacy (Avgerinou & Pettersson, 2011). For example, Arnerson and Offerdahl (2018) suggest that a description for visual literacy should include disciplinary discourses such as decoding and interpreting visual representations, encoding and creating visual representations, as well as generating mental models. Offerdahl, Arneson and Byrne (2017), however, suggest that visual literacy means the ability to read and write visual representations by using a complex system of semiotics used to represent disciplinary ways of knowing. Such a complex system of semiotics includes written language, images, and symbols which represent discipline-specific knowledge. Avgerinou and Pettersson (2011), however, suggest that visual literacy is an interdisciplinary visual language, grammar, syntax, and vocabulary which make up a visual language. Linenberger and Holme (2015, p. 24) also suggest that "visual literacy is the ability to understand (read) and use (write) images and to think and learn in terms of images."

Similar to the varying perspectives on what visual literacy means, there is also a lack of a cohesive theory of visual literacy (Avgerinou & Pettersson, 2011). As a result, there is no general agreement regarding the pedagogical applications of visual literacy, including methods for developing, measuring, and assessing it. Therefore, it is not surprising that a wide range of techniques has been used to measure and assess visual literacy. For example, Arnerson and Offerdahl (2018) proposed a visualization blooming tool, Offerdahl et al. (2017, p. 6) proposed a "taxonomy for characterizing abstraction in instructional representations," while Linenberger and Holme (2015) proposed a Biomolecular Visualization Framework. The complexity and variation between these tools are indicative of the differences in the understanding of visual literacy amongst scholars. It further justifies the present researcher's argument that visual literacy cannot be understood outside specific disciplines.

A case for visuo-semiotic reasoning in biology

In line with the above argument, Avgerinou and Ericson (1997) suggest that visual literacy abilities include visual thinking, visualization, and other related cognitive abilities. This includes cognitive abilities for understanding, using, and thinking through VSMs (Avgerinou & Quinn Knight, 2005). However, these VSMs are generally context-specific. For this reason, the present researcher cautions that definitions that draw parallels to general verbal literacy by referencing the ability to read (make sense of) and write (draw or create) visual representations (e.g., Offerdahl et al., 2017) may be misplaced. This is because visual literacy is context-specific rather than generic.

For this reason, visual literacy is directly intertwined with subject-specific knowledge and semiotics, which may differ significantly within and between disciplines. For example, a student may be visually literate in mathematics, but be visually illiterate in biology. Therefore, the present researcher proposes that visual literacy should be defined within the context of related disciplines. This could include, bio-visual literacy, and mathematical-visual literacy, wherein the visual literacy is directly defined within the context of disciplinary discourse and related semiotic resources, which provide access to a discipline-specific epistemology and ontology. This view is in line with Airey and Linder (2009, p. 33) who suggest that "visual literacy can be defined in terms of discursive fluency, that is, when a student understands the various ways in which the discipline generally uses that mode to represent a particular way of knowing." Similarly, Avgerinou and Knight (2005) argue that mathematics teachers who are visually literate in the mathematics context could teach mathematics better than those who are not. This is to say; mathematical visual literacy could be useful for teaching mathematics, and not for other disciplines.

Weliweriya, Sayre and Zollman (2018) suggest that semiotic resources include symbolic tools such as language, diagrams, sketches, graphs, and signs which are used to construct mental models and knowledge. According to Van Leeuwen (2005), semiotic resources offer the capability to students and researchers to construct the meaning of content knowledge depending on how they are used. Consequently, semiotics provide access to disciplinary knowledge, which student would otherwise not have (Weliweriya et al., 2018). In line with this reasoning, therefore, the present researcher believes that there is a specific set of semiotic resources which could afford biology students access to relevant biology-specific content knowledge, which however varies with the various sub-disciplines of biology. It is for this reason that content knowledge and related semiotic resources are arranged and taught hierarchically (Khodor, Halme & Walker, 2004). Therefore, the present research explores visuo-semiotic reasoning in biology (VSR-b) concerning content knowledge and semiotics of amino acid structures.

Based on the above discourse, VSR-b is defined in the present research as the ability to internalize, conceptualize, and externalize biology content knowledge through the use of VSMs and discipline-specific semiotics representing biology content. A small letter 'b' is used in the above abbreviation because 'biology' knowledge explored in the present research does not reflect all existing Biology knowledge.

Aims of the research

In light of the above discourse, the present research aimed to develop an instrument for assessing VSR-b by asking the question "how could an instrument for assessing visuo-semiotic reasoning in biology be developed?". While biology as a broad field was chosen, the research focused primarily on amino acid structures as a context for studying visuo-semiotic reasoning.

Theoretical Framework framing the VSR-b test

The following theoretical perspectives informed the development of a VSR-b Test reported in this article:

a) The theoretical cognitive process of visualization (Mnguni, 2014);
b) The taxonomy for teaching, learning and assessing content knowledge (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, ... &, Wittrock 2001);
c) The model of factors determining students' ability to interpret external representations in Biochemistry (Schönborn & Anderson, 2009); and,
d) The taxonomy for characterizing abstraction in instructional representations (Offerdahl et al., 2017).

In the theoretical cognitive process of visualization, Mnguni (2014) relies on cognitivism, constructivism and the cognitive theory of multimedia learning to suggest that learning from VSMs involves i) inputting of information from the external world into the cognitive structures (Internalization of VSMs, IVM), ii) the cognitive development and processing of visual information (Conceptualization of VSMs, CVM), and iii) the externalization of (visual) information (as VSMs) (Externalization of VSMs, EVM). Consequently, in the present research, VSR-b included skills related to IVM, CVM, and EVM (Figure 1).

These skills were identified using the taxonomy for teaching, learning, and assessing, also known as Bloom's taxonomy (Anderson et al., 2001). This taxonomy classifies a list of learning objectives in the cognitive domain, which involve knowledge and the development of intellectual skills, including the recognition of specific facts, procedural patterns, and concepts that serve in the construction of knowledge as well as the development of mental abilities and skills (Bloom, Krathwohl & Masia 1956). Bloom's taxonomy was used in the present research because it includes skills that would be used in the process of visualizations (i.e., IVM, CVM, and EVM) as suggested by Mnguni (2014) (Figure 1). Mnguni (2014), however, indicates that IVM skills are pre-attentive and require a significantly low degree of cognitive effort and almost no content knowledge. In line with this view, Kawahara and Yokosawa (2001) also indicate that these pre-attentive skills include target detection, region tracking, and counting.

Schönborn and Anderson's (2009) model of factors determining students' ability to interpret external representations in molecular biology was used in the present research to identify additional factors that could affect VSR-b. Schönborn and Anderson (2009) suggest that students' ability to interpret VSMs is affected by i) their

existing conceptual understanding and prior conceptual knowledge (of relevance to the VSM in question), ii) their ability to reason with the VSM and with their conceptual knowledge of relevance to the VSM, and iii) the actual external nature of the VSM. As a result, the development of the VSR-b Test in the present research would include investigating these three factors (Figure 1). In particular, students' existing content knowledge and their ability to process visual information (i.e., CVM) cognitively were investigated. The external nature of the VSM was understood in the present research within Offerdahl et al.'s (2017) taxonomy for characterizing abstraction in instructional representations. Offerdahl et al. (2017) argue that abstract VSMs can be classified as a symbolic, schematic, graphic, cartoon, or realistic, and students require a specific set of skills to be able to learn from them.

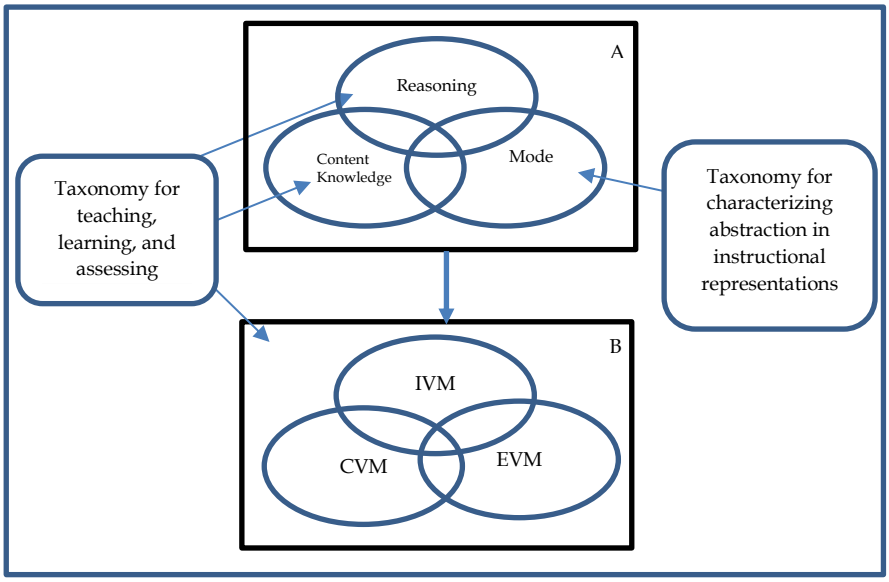


Figure 1. A theoretical framework used in the development of the VSR-b Test. In the framework, 'A' is Schönborn and Anderson's (2009) model of factors determining students' ability to interpret external representations while B is Mnguni's (2014) theoretical cognitive process of visualization.

Methods

As indicated earlier, the present research aimed to develop an instrument for assessing VSR-b using the theoretical framework discussed in Figure 1. To do this, a multiplistic realism research paradigm was adopted, which, according to Krauss (2005), allows for flexible integration of both qualitative and quantitative research methods. This enhances the validity and credibility of the findings. Consequently, the present research followed the explorative mixed-method research method approach for data collection and analysis.

Content knowledge and semiotics of amino acid structures were used as the context of the research. In particular, the researcher investigated knowledge and reasoning ability of students concerning the structures of α-amino carboxylic acid which contain an α-carbon, to which an amino, a carboxyl group, a hydrogen atom, and the R group are attached.

Due to the nature of this research, specific research methods (i.e., sampling, instruments, validity, and reliability as well as data analysis) are discussed within the different stages of the research as explained below:

a) Identification of VSR-b Skills

As mentioned above, Mnguni's (2014) theoretical cognitive process of visualization for science education and Bloom's taxonomy of learning objectives were used to identify visualization skills required for effective learning in biology. According to Mnguni (2014, p. 3), visualization occurs in three non-linear overlapping stages; namely, IVM, CVM, and EVM. It was, therefore, essential to identify the cognitive skills that are required to process VSMs in each of these stages.

Similarly, Schönborn and Anderson (2009) suggest that students' reasoning ability, students' understanding of the concepts of relevance to the ER, and the nature of the mode in which the ER represents the desired phenomenon are necessary for effective learning with VSMs. Of these factors, reasoning ability was identified by the present researcher as a cognitive function which utilized VSR-b skills. These skills were identified through an independent panel of nine experts which comprised of two biochemists (with PhDs in Molecular Biology), four secondary school biology teachers who were enrolled for Ph.D. studies in science education, and three postgraduate molecular biology students. The panel of experts was asked to independently study learning objectives related to Bloom's taxonomy and indicate the VSR-b skills they believe were utilized during the IVM, CVM, and EVM concerning learning about amino acid structures. Their responses were collected using an open-ended questionnaire. After that, a semi-structured questionnaire was used to verify the classification of VSR-b skills that emerged from the open-ended questionnaire. Responses of this semi-structured questionnaire were then used to calculate the content validity index (CVI) (Hyrkäs, Appelqvist-Schmidlechner & Oksa 2003) to determine the extent of agreement between the panel members. As suggested by Hyrkäs et al. (2003), only those skills that obtained a CVI of .69 and above were included in the development of the VSR-b Test. Results of this exercise are presented under 'VSR-b skills identified through the panel of experts' in the Results section.

b) Design and validation of the items

The researcher then developed items through which the VSR-b skills (listed in Table 1) could be tested (see examples in Appendix 1). In line with the Theoretical Framework (Figure 1), these items probed:

i) *Students' conceptual knowledge of amino acid structures.* Therefore, the VSR-b Test, probed content knowledge of amino acid structures which did not incorporate the use of VSMs. This knowledge was identified as a

prerequisite for students to answer the subsequent sections of the Test. This knowledge was taken from textbooks (e.g., McKee & McKee, 2017; Walsh, 2014) used by students in undergraduate biology.

ii) *Content-free VSR-b skills for IVM.* Performing these skills did not require biology-related content. Relevant items similarly probed generic visualization skills as the Senior Aptitude Test (Mnguni, 2018). They tested students' ability to recognize orientation through mental rotation, perceive spatial orientation in 2D and 3D, perceive and track the motion of an object on a 2D and 3D planes, perceive luminance, speed, texture, shapes, and patterns related to amino acid structures.

iii) *Content-specific VSR-b skills for CVM.* Items in this regard probed students' ability to reason with models by utilizing prior biology content knowledge (Schönborn & Anderson, 2009). This included remembering, understanding, analyzing, and evaluating visuo-semiotic models used to represent amino acid structures. In this instance, all items incorporated visuo-semiotic models of amino acid structures. These models were taken from textbooks used by students in undergraduate biology (e.g., McKee & McKee 2017; Walsh, 2014).

iv) *Content-specific VSR-b skills for EVM.* Items in this regard probed students' ability to externalize content knowledge through the use of visuo-semiotic models when applying, creating, and synthesizing biology content knowledge.

The independent panel of experts (mentioned earlier) was then tasked with validating the items. They were given a semi-structured validation questionnaire through which they agreed or disagreed that the items probed, i) students' knowledge of amino acid structures, ii) VSR-b skills for IVM, iii) VSR-b skills for CVM, and iv) VSR-b skills for EVM. This exercise sought to determine face and content validity of the items by asking the panel of experts to indicate amongst other things if the items probed what they ought to be; and, were suitable for their intended purpose. Results of this exercise are presented under 'Design and validation of the items for the VSR-b Test' in the Results section.

The items were also piloted on a group of 18 second-year molecular biology university students who had completed a course in amino acids and protein structures and functions. The aim in this regard was to further determine the face and procedural validity and reliability of the Test (Aboraya, France, Young, Curci & LePage, 2005).

c) Testing of the VSR-b Test

The instrument was then used to probe VSR-b of a group of 30 third-year Bachelor of Science students majoring in Molecular Biology at a South African University. These students were selected purposefully as they had completed a course in amino acids and protein structures and functions. This group was, however, different from the pilot group even though they were studying at the same institution. At the time of the research, the participants were all enrolled in a Protein Biochemistry class. Lecture materials and other study materials provided to the students as part of their molecular

biology courses included static and animated VSMs. They had all passed pre-requisite molecular biology modules in the first and second year, including other prerequisite subjects in biology, mathematics, physics, and chemistry. Results of this exercise are presented under 'Results from the application of the VSR-b Test' in the Results section.

Results

VSR-b skills identified through the panel of experts.

Twenty-five visualization skills were identified (Table 1). The panel of experts classified each of these into the three visualization stage as well as into Bloom's taxonomy of cognitive skills. The number of visualizations skills varied in each of the visualization stages. The panel of experts also indicated that IVM could have the most amount of visualization skills which were also classified as 'pre-cognitive' in that they may not require extensive content knowledge to be carried out. The panel of experts also indicated that some of the skills could be performed in more than one visualization stages. For example, 'outline' was classified in both the CVM and EVM stages.

Table 1

The List of VSR-b Skills and Related Stages of Visualization by the Panel of Experts (n = 9).

Visualization Stage	Bloom's stage	Visualization skills	Definition	Code	CVI
IVM	Pre-cognition	Recognize orientation through mental rotation	To move, arrange, operate, or control cognitively in a skillful manner for examination purposes and then to perceive multiple items with different orientation and shape to be the same if orientation and shape is rearranged	IVM01	.70
		Spatial perception 2D	To perceive spatial relationships and distances between objects, in 2-dimensions	IVM02	.77
		Spatial perception 3D	To perceive spatial relationships and distances between objects, in 3-dimensions	IVM03	.74
		Track (including find, and locate)	To come upon or discover by searching or making an effort; to discover or ascertain through observation, to determine or specify the position or limits of by searching, examining	IVM04	.94
		Ground perception	To detect or perceive the part of a scene (or picture) that lies behind objects in the foreground	IVM05	.78
		Perceive luminance /Identify colors	To detect or perceive a visual attribute of things that result from the light they emit or transmit or reflect	IVM06	.74
		Perceive motion	To recognize, discern, envision, or understand a change of position in space and assign meaning to	IVM07	.87
		Perceive speed	To recognize, discern, envision, or understand a rate of movement and meaning thereof	IVM08	.91
		Perceive texture	To recognize, discern, envision, or understand the characteristic visual and tactile quality of the surface and meaning of such	IVM09	.91
		Perceive shapes	To recognize, discern, envision, or understand the external form, or outline of a geometric figure.	IVM10	.87
		Perceive patterns	To recognize, discern, envision, or understand the arrangement or design found in objects	IVM11	.87

Table 1 Continue

CVM	Evaluating	Critique (including to judge)	To critically examine and judge something	CVM01	.76
	Analyzing	Outline	To give the main features or various aspects of; summarize	CVM02	.95
	Understanding	Interpret (including to analyze; Assess; Evaluate; Examine; Investigate)	To break down into components or essential features by making sense of or assigning meaning to or give an explanation and to examine and or assess carefully and observe or inquire into in detail by examining systematically to observe carefully or critically	CVM03	.89
		Compare (including discriminate)	To examine and note the similarities or differences of and bring into or link in logical or natural association and establish or demonstrate a connection between	CVM04	.70
		Classify (including to arrange, order, and organize)	To put into a specific order or relation through a methodical or systematic arrangement or to arrange in a coherent form or pattern based on specific features	CVM05	.91
		Describe (including to discuss, and explain)	To make understandable or comprehensible by adding details or to justify or offer reasons for or a cause and give a description of, by conveying an idea or impression in speech or writing; characterize	CVM06	.89
	Remembering/ comprehension	Recall/retrieve	To remember by retrieving information from memory	CVM07	.74
EVM	Creating/ synthesis	Complete	To make whole, with all necessary or standard elements or parts	EVM01	.71
		Imagine	To form a mental image of something that is not present or that is not given	EVM02	.77
	Applying	Illustrate (including to sketch)	To clarify, as by use of examples or comparisons and to use drawings to describe roughly or briefly or give the main points or summary of	EVM03	.82
		Outline	To give the main features or various aspects of; summarize	EVM04	.79
		Complete	To make whole, with all necessary or standard elements or parts	EVM05	.79
	Creating/ synthesis	Develop (including to formulate, devise, construct, create, produce, invent)	To cause to exist in a new or different form through artistic or imaginative effort	EVM06	.73
		Infer meaning	To conclude by reasoning; in logic or reason or establish by deduction or state, tell about, or make known in advance, based on specialized knowledge	EVM07	.76

Design and validation of the items for the VSR-b Test

During the validation of the VSR-b Test, expert validation of the instrument showed a significant inter-rater agreement measured through the inter-rater reliability (Table 2). The experts generally agreed with the design of all the items and their classification into the four sections, namely, Students' conceptual knowledge, Content-free VSR-b skills for IVM, Content-specific VSR-b skills for CVM, and Content-specific VSR-b skills for EVM. Results from the pilot research also showed that students in the pilot research were able to respond to all 45 items within 45 minutes where the reliability co-efficient (i.e., Cronbach Alpha) was .78. One point was allocated to each item, such that the total obtainable score was 45 points.

Table 2

Inter-Rater Reliability Within the Different Components of the VSR-b Test Items Computed from the Panel of Experts Responses.

Construct	Number of items	Inter-rater reliability
Content knowledge	20	.831
IVM	11	.793
CVM	7	.801
EVM	7	.768

The panel of experts also provided qualitative comments on how the instrument could be improved further. For example, experts commented that:

- *"The content of the test is taught at undergrad biology. Students should be able to answer all the questions without problems" (P6, a postgraduate student)*
- *I had seen the pictures [in books] before, even if they were not the same…so interpretation of the test pictures was easy". (P3, a postgraduate student)*
- *"The IVM section is interesting and assessing critical generic skills. Maybe students should take this test before enrolling for molecular biology modules" (P5, a qualified biochemist)*
- *interpretation of the test pictures was easy". (P2, a postgraduate student)*
- *"[The symbols] are large enough and spaced comfortably. I can work through the diagrams and notice differences" (e.g., P1, a postgraduate student)*
- *"Some questions are too complicated and need careful observation and consideration before answers can be developed" (P4, a Ph.D. student).*
- *"Students may not [necessarily] know what is meant in scientific terms" (P8, a Ph.D. Student).*
- *"The test is appropriate for 3rd-year molecular biology students, but not for lower levels as it was quite a challenging test". (P5, a qualified biochemist)*

These comments were then used by the researcher to refine the VSR-b Test further.

Results from the application of the VSR-b Test

Having satisfied the instrument design and validation by experts and pilot group, the instrument was then tested on Molecular Biology students for whom it was designed. Results in this instance showed that the majority of participating students scored above 50% in the content knowledge (M = 63.27. $S.D.$ = 12.39), CVM (M = 55.90 $S.D.$ = 10.59) and EVM (M = 56.90, $S.D.$ = 13.74) components of the test. IVM was however lower (M = 47.00, $S.D.$ = 9.12). The IVM score was significantly lower than the other scores (p < .001) while the content knowledge score was significantly higher than the IVM, CVM and EVM (p < .001). An analysis of students' performance in individual skills showed that the pass rates in individual skills were low (Figure 2).

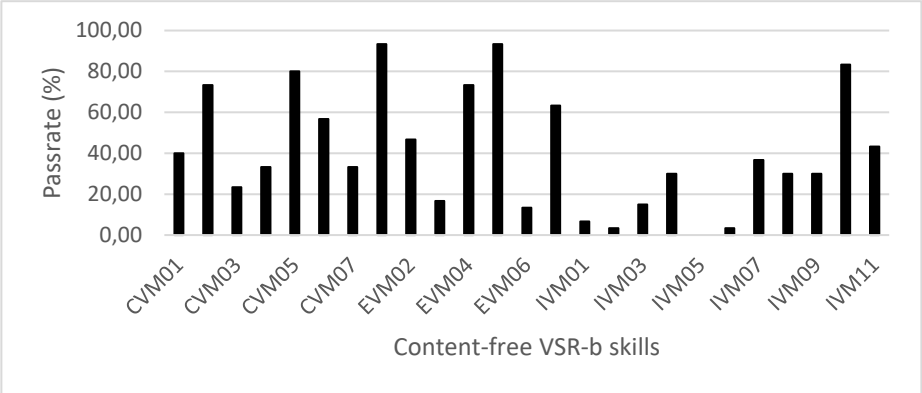


Figure 2. Students' performance in the IVM, CVM, and EVM

Results showed in this regard that in only eight of the 25 skills were half of the participants able to successfully perform the skills. Participants were not able to successfully perform most skills in the IVM. In this regard, the performance was worse in skills 1, 2, 5, and 6, where only very few students were able to perform the skills respectively successfully.

The average inter-item correlation was then calculated for the items that measured within each of the four components of the test. To this end, internal consistency reliability was generated within and between each of the four component of the test (i.e., content knowledge, IVM, CVM, and EVM) (Table 3 and 4). Results showed high internal consistency reliability in each of the four components of the test, except for CVM where Cronbach's Alpha was .597.

Table 3

Internal consistency reliability coefficients within the different components of the VSR-b Test.

Construct	Cronbach's Alpha	Cronbach's Alpha Based on standardized items
Content knowledge	.880	.883
IVM	.816	.857
CVM	.597	.660
EVM	.828	.840

The overall reliability of the test was relatively high (Cronbach Alpha = .843). Results also showed that there was a significant correlation between the different components of the test (Table 4). However, the correlation was not very strong between IVM and EVM as well as IVM and content knowledge.

Table 4

Correlation between content, IVM, CVM and EVM (n = 30)

		IVM	CVM	EVM
CVM	Correlation	.590**		
	Sig. (2-tailed)	0.001		
EVM	Correlation	.467**	.790**	
	Sig. (2-tailed)	0.009	0.001	
Content	Correlation	.443*	.600**	.596**
	Sig. (2-tailed)	0.014	0.001	0.001

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Discussion, Conclusion, and Recommendations

The significance of VSMs in science education cannot be ignored, as suggested in the literature (e.g., Mnguni, 2014). However, the challenge that has stood for decades has been the lack of clear and universal theory and definition of what visual literacy is, and how it could be developed and assessed (Avgerinou & Pettersson, 2011). The present research has attempted to demonstrate that visual literacy does not exist independent of content knowledge. Therefore, the present research was informed by Schönborn and Anderson (2009), who argue that content knowledge affects students' ability to reason with VSMs. This is supported in the present research in that content knowledge correlated significantly with the different skills utilized during visualization. As expected, however, the correlation between content knowledge and

skills related to IVM was low. This is because IVM in the present research was defined as those pre-attentive skills that did not require content knowledge related to the present research. Similarly, the present research has demonstrated that visualization skills tested in CVM and EVM are also crucial for the development of content knowledge. While causality was not tested in the present research, the strong association between content knowledge and IVM, CVM and EVM suggest that a lack of visualization skills may impact on students' understanding of related content knowledge.

Related to the difficulty of developing a universal theory and definition for visual literacy, the present researcher argues that visual literacy can be better understood within specific contexts. In line with this reasoning, the researcher is of the view that the instrument presented in this research could be used to design context-specific tests for visual literacy. For example, the present instrument can only be reliable in the context of amino acid structures. However, it cannot be applicable in other contexts. For this reason, the researcher suggests that the instrument be adapted accordingly, for example, by modifying the content knowledge and semiotics. This is because visual literacy is affected by subject-specific knowledge and semiotics (Offerdahl et al., 2017; Schönborn & Anderson, 2009).

Conclusion

The present research has shown that VSR-b can be assessed using the VSR-b Test. In particular, this research has shown that VSR-b can be understood within the context of the theoretical cognitive process of visualization (Mnguni, 2014), the taxonomy for teaching, learning and assessing (Anderson et al., 2001), the model of factors determining students' ability to interpret external representations in Biochemistry (Schönborn & Anderson, 2009), and the taxonomy for characterizing abstraction in instructional representations (Offerdahl et al., 2017) as a theoretical framework. The research has also shown that VSR-b includes students' conceptual knowledge, content-free VSR-b skills for IVM, content-specific VSR-b skills for CVM as well as content-specific VSR-b skills for EVM. The present researcher, therefore, argues that researcher, teachers, curriculum, and instructional designers should consider all these factors as significant in the development and assessment of visualization skills amongst students. The researcher also proposes further researcher to improve the understanding of teaching and learning that utilizes visuo-semiotic models. This could include testing and further developing the present instrument in different contexts.

References

Aboraya, A., France, C., Young, J., Curci, K., & LePage, J. (2005). The Validity of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Validity of Psychiatric Diagnosis. *Psychiatry (Edgmont)*, *2*(9), 48–55.

Airey, J., & Linder, C. (2009). A disciplinary discourse perspective on university science learning: Achieving fluency in a critical constellation of modes. *Journal of Research in Science Teaching*, *46*(1), 27–49.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition*. White Plains, NY: Longman.

Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in bloom: Using bloom's taxonomy to support visual learning skills. *CBE Life Sciences Education*, *17*(1), 1–8. doi: 10.1187/cbe.17-08-0178.

Arsad, N. M., Osman, K., & Soh, T. M. T. (2011). Instrument development for 21st century skills in Biology. *Procedia-Social and Behavioral Sciences*, *15*, 1470-1474.

Avgerinou, M. D., & Pettersson, R. (2011). Toward a Cohesive Theory of Visual Literacy. *Journal of Visual Literacy, 30*(2), 1-19. doi: 10.1080/23796529.2011.11674687.

Avgerinou, M. D., & Quinn Knight, E. (2005). *Assessing the Visual Literacy Skills And Perceptions Of Pre-Service Math Teachers*. International Visual Literacy Conference, Orlando, Florida. October 17-2. 2005. Retrieved from https://www.researchgate.net/publication/313469975_Assessing_The_Visual_Literacy_Skills_And_Perceptions_Of_Pre-Service_Math_Teachers.

Avgerinou, M., & Ericson, J. (1997). A review of the concept of visual literacy. *British Journal of Educational Technology*, *28*(4), 280-291.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1956). *Taxonomy of educational objectives: The classification of educational goals.* New York: McKay.

Hyrkäs, K., Appelqvist-Schmidlechner, K., & Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, *40*(6), 619-625.

Kawahara, J. I., & Yokosawa, K. (2001). Pre-attentive perception of multiple illusory line-motion: A formal model of parallel independent-detection in visual search. *The Journal of General Psychology*, *128*(4), 357-383.

Khodor, J., Halme, D. G., & Walker, G. C. (2004). A hierarchical biology concept framework: a tool for course design. *Cell Biology Education*, *3*(2), 111-121.

Krauss, S. E. (2005). Research paradigms and meaning-making: A primer. *The Qualitative Report, 10*(4), 758–770.

Linenberger, K. J., & Holme, T. (2015). Biochemistry Instructors' Views toward Developing and Assessing Visual Literacy in Their Courses. *Journal of Chemical Education, 92*(1), 23-31. doi: 10.1021/ed500420r.

McKee, T., & McKee, J. R. (2017). *Biochemistry: the molecular basis of life*. International 6th edition. Oxford: Oxford University Press.

Mnguni, L. (2018). A description of visual literacy among third year biochemistry students. *Journal of Baltic Science Education*, *17*(3), 486-495

Mnguni, L. E. (2014). The theoretical cognitive process of visualization for science education. *SpringerPlus*, *3*(1). doi: 10.1186/2193-1801-3-184.

Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge? *Learning and Instruction*, *31*, 13-22.

Offerdahl, E. G., Arneson, J. B., & Byrne, N. (2017). Lighten the load: Scaffolding visual literacy in biochemistry and molecular biology. *CBE Life Sciences Education, 16(1), 1-11*. doi: 10.1187/cbe.16-06-0193.

Schönborn, K. J., & Anderson, T. R. (2009). A Model of Factors Determining Students' Ability to Interpret External Representations in Biochemistry. *International Journal of Science Education*, *31*(2), 193–232.

Van Leeuwen, T. (2005). *Introducing social semiotics*. New York: Routledge

Walsh, G. (2014). *Proteins: Biochemistry and Biotechnology, 2nd Ed.* West Sussex: Wiley-Blackwell.

Weliweriya, N., Sayre, E. C., & Zollman, D. (2018). *Case Study: Coordinating Among Multiple Semiotic Resources to Solve Complex Physics Problems*. *European Journal of Physics*, *40*(2), 1-26.

Eurasian Journal of Educational Research 82 (2019) 137-168

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

Factors Influencing Students' Research Self-Efficacy: A Case Study of University Students in Malaysia

Racheal POH[1], Abdul Ghani Bin KANESAN ABDULLAH[2]

ARTICLE INFO

Article History:
Received: 22 Apr. 2019
Received in revised form: 18 Jun. 2019
Accepted: 15 Jul. 2019
DOI: 10.14689/ejer.2019.82.8

Keywords
Higher education; mentoring; interest in research; research training environment; research knowledge.

ABSTRACT

Purpose: The main aim of the study was to investigate the relationship between research self-efficacy and the perception of the research training environment, interest in research, research mentoring experience, and research knowledge within a sample of Ph.D. students (N=120) at a local university in Malaysia.

Method: Correlation and regression analysis were employed to investigate the impact of research mentoring, research training environment, interest in research and research knowledge on research self-efficacy. Self-reported questionnaires and a research knowledge test were distributed to the Ph.D. students in the university to collect the data.

Findings: An apparent correlation was found between research self-efficacy, research training environment, interest in research and research mentoring while research self-efficacy was not correlated with research knowledge. Research training environment and interest in research contributed to 26 percent of the variance of change in research self-efficacy.

Implications for research and practice: Conducive environment in the institution fosters higher levels of research self-efficacy and encourage Ph.D. students' research skills and development. Furthermore, the incorporation of research methodology as a compulsory course in the Ph.D. study serves to improve students' research interest and knowledge. The self-efficacy theory helps to assist decision making in identifying future researchers in the institution.

[1] INTI International College, Penang, MALAYSIA
[2] University Sains Malaysia, MALAYSIA
For any correspondence, please email the author at racheal.poh@newinti.edu.my ORCID: https://orcid.org/0000-0002-8220-7739

Introduction

Research self-efficacy was described as the confidence in carrying out research activities from organizing a research plan to carrying out the research process from library research and reading to writing and publication (Forester, Kahn & Hesson-McInnis, 2004). The accurate assessment of research self-efficacy may assist the faculty to identify a given students' self-identified strengths and weaknesses with respect to research, thereby facilitating graduate research. High research self-efficacy was an important factor related to students' successfully conducting research and pursuing research beyond graduate study (Forester et al., 2004).

Research is an important area for sustainability in Malaysian higher education because of its volatilities in global higher education (Omar, Mohd Ali, Khalid, Zakaria & Anuar, 2013). The integration of higher education policy with good research was a determining factor in global competitiveness with neighbor countries (Koo, Fazal Mohamed & Kemboja, 2012). Furthermore, research produces new knowledge, and it is essential to the country's growth and development (Yazit & Zainab, 2007). In recent years, the number of Ph.D. enrollments in Malaysia has dropped significantly. For example, in 2009, the number of admission into Ph.D. programs in Malaysia was 4,942 and number of graduate for that year was 750 for both Public and Private University (Ministry of Higher Education, 2015). The shortage of Ph.D. graduates in Malaysia implies that it is crucial that the Ph.D. students are of high quality, well trained and possess a high level of research self-efficacy in order to become competent researchers in the future (Aziz, 2016).

Although there was a positive outlook in the trends of research, output in social science and humanities, further efforts were required to ensure the sustainability and continuous growth towards a knowledge-based country (Ahmad, 2012). The success of the National Higher Education Action Plan Phase 2 (2011 - 2025) or known as Pelan Strategik Pengajian Tinggi Negara (PSPTN) has led to the new Malaysian Education Blueprint 2015-2025 for higher education (Pelan Pembangunan Pendidikan Malaysia (Pendidikan Tinggi) (PPPM) 2015 – 2025 where the success of a giant leap of increase in the output of research was recorded. The "Innovation Ecosystem" agenda in the Malaysian Education Blueprint (2015-2025) encourages more academics to partner with the community to develop and commercialize ideas. Ministry of Higher Education facilitates the development of an innovation ecosystem in selected strategic areas that are critical to the nation's economic growth. Therefore, more Ph.D. graduates are required as the universities now need to improve research skills (Ministry of Higher Education, 2015).

Three (3) important research objectives were constructed in the study.

Research objective 1:

The main research objective was to find out the level of research self-efficacy of Ph.D. students.

The research objective 1 was further divided to:

a) Find out the level of research mentoring experience of Ph.D. students.

b) Find out the level of the research training environment.

c) Find out the level of interest in research among Ph.D. students.

d) Find out the level of research knowledge that Ph.D. students possess.

Research objective 2:

To find out the relationships between research mentoring, research training environment, interest in research, research knowledge and research self-efficacy.

Research objective 3:

To find out the impact of research mentoring, research training environment, interest in research, research knowledge on research self-efficacy.

Problem Statement

The 11th Malaysian Plan, Economic Transformation Programme (ETP) and New Economic Model (NEM) are series of programs introduced by the Malaysian government in order to propel Malaysia towards becoming a high-income nation by 2020 (Chin, 2015). To achieve vision 2020, a highly-skilled, creative and innovative workforce becomes essential to realize such aspiration. Organisation for Economic Cooperation and Development (OECD) reported on the structural policy in Malaysia has found that the talent base of the workforce has been lagged behind the standards of high-income nations (OECD, 2013).

Malaysia suffers from a shortage of skilled workforce and is overly reliant on its low wages and unskilled workers. An example of over-reliance on unskilled workers is the recent Memorandum of Understanding (MoU) signed between Malaysian and Bangladesh where Malaysia agrees to hire 1.5 million of the Bangladeshi workers over the next five years (Carvalho & Rahim, 2016). The Labor Force Report by the Department of Statistics in Malaysia has reported that only 55.5 percent of the employed workforce attained secondary education in 2014, and it has declined by 0.5 percentage points as compared to 2013 (Department of Statistics Malaysia, 2014). Furthermore, the percentage of employed persons with primary education and no formal education declined by 1.1 percentage points and 0.3 percentage points, respectively.

The Malaysian government realized that the education system in Malaysia requires a change to transform itself into a high-income and knowledge-based country. Therefore, the Malaysian Education Blueprint (2015-2025) (or known as Pelan

Pembangunan Pendidikan Malaysia (Pendidikan Tinggi) (PPPM) 2015 – 2025 by the Ministry of Higher Education was launched to transform the education system to be ahead of global trends (Ministry of Higher Education, 2015). The Malaysian Higher Education Blueprint 2015-2025 has reported that the number of research publications was not as widespread and intensive as desired (Ministry of Higher Education, 2015). The MyBrain15 program was implemented in 2008 to increase the number of Ph.D. holders in the workforce (The Star Online, 2008). Ministry of Higher Education realized there was a shortage of critical mass for researchers to drive innovation and economic growth in Malaysia. Malaysian universities faced a shortage of lecturers with Ph.D. qualifications as they push to achieve the target workforce which consists of 75 percent of lecturers with Ph.D. (University World News, 2011). Medical lecturers left the public sector for a better remuneration package and career development (Chin, 2014). The direct effect of the shortage of qualified lecturers in Malaysia has caused the workforce to be lagged behind high-income nations (OECD, 2013).

MyBrain15 was an accelerator program initiated to produce 60,000 Ph.D. holders in Malaysia by 2023 as the country faces a shortage of Ph.D. holders in the workforce (MyBrain15, 2008). Over the years, the Ministry of Higher Education has maintained its budget allocation in providing assistance to those who wish to pursue their Ph.D. through the MyBrain15 program. For example, a total of 20,000 scholarships, where 15,000 were offered for Master's and 5,000 were offered for Ph.D. program in 2016 (Ahmad Kamil, 2016). The objective of the MyBrain15 Program is to produce 60,000 Ph.D. graduates by the year 2023. As of December 2015, the target of 60,000 Ph.D. graduates has not been achieved and 8,205 students have received the offers under MyPh.D. and MyPh.D. Industri program (Ahmad Kamil, 2016).

The number of Ph.D. enrollments has also dropped significantly. For example, in 2009, the number of admissions into Ph.D. programs in Malaysia was 4,942 and number of graduates for that year was 750 for both Public and Private University (Ministry of Higher Education, 2015). The shortage of Ph.D. graduates in Malaysia implies that it was crucial that the Ph.D. students are of high quality, well trained and possess a high level of research self-efficacy in order to become competent researchers in the future (Aziz, 2016). Evans (2012) pointed out that the attitudinal component of researcher development has been largely overlooked and it may be the reason why research behavior and productivity are not enhanced as expected. In fact, Evans (2011) argues that the attitudinal component of researcher development is crucial towards shaping a researcher's intellectual and behavioral development. Attitudinal development includes the researcher's perception of the relevance and usefulness of research and his or her self-efficacy (Evans, 2011).

Higher education institutions have been tasked to develop graduate students to become psychologically competent and effective researchers with sufficient levels of research competency (Unrau & Beck, 2004). Universities are compelled to provide students with an appropriate research environment and experience as well as adequate psychological confidence for challenging research tasks (Bernardin, 1996; Presely & Engelbride, 1998). Ph.D. graduate drop-outs range from 30 to 50 percent, depending on the discipline and country (Cassuto, 2013; Fullick, 2013).

One of the factors that caused the students to drop out their studies was due to the incompletion of the thesis. The studies of Bowen and Rudenstine (2014); Kerlin (1995); and Tinto (1993) further discovered that many Ph.D. students were "caught" in a situation named as "All but Dissertation." The ABD (All but dissertation) refers to a situation whereby the doctorate student has completed all of the requirements of the degree except for the dissertation/thesis. A student usually proceeds as a doctoral candidate once he or she has completed all the coursework required for the degree and has passed the comprehensive exam. The doctoral degree will be conferred to the student who has completed their thesis (Kuther, 2015). Unfortunately, many students remain in the ABD status because of poor research self-efficacy, difficulties faced in research, loss of interest and motivational deficits (Kuther, 2015).

Many studies have demonstrated that the number of Ph.D. graduates in the country has not reached its target number, and the level of research self-efficacy has diminished among the students (Ahmad Kamil, 2016; Aziz, 2016; Evans, 2011). Some scholars have suggested that the lack of research self-efficacy was due to mentors (Hollingsworth & Fassinger, 2002), research training environment (Schlosser & Gelso, 2001), interest in research (Lambie & Vaccaro, 2011) and research knowledge (Lambie et al., 2014). However, to date, no comprehensive study has been carried out to determine the factors influencing students' research self-efficacy at a public university in Malaysia. The study was conducted to solve the problems identified among Ph.D. graduates in the country.

In addition, three (3) important research questions were identified based on the research objectives.

Research question 1:

What is the level of research self-efficacy of Ph.D. students?

a) What is the level of research mentoring experience that Ph.D. students have throughout their study?

b) What is the level of research training environment that Ph.D. students have?

c) What is the level of interest in research that Ph.D. students possess?

d) What is the level of research knowledge that Ph.D. students possess?

Research question 2:

What are the relationships between research self-efficacy and research mentoring, research training environment, interest in research, and research knowledge?

Research question 3:

What is the impact of research mentoring, research training environment, interest in research, and research knowledge on research self-efficacy?

Research hypothesis

Five (5) research hypothesis was identified in the study and illustrated as follows:

H1: There is a positive relationship between research self-efficacy and research training environment.

H2: There is a positive relationship between research self-efficacy and interest in research.

H3: There is a positive relationship between research self-efficacy and research mentoring experience.

H4: There is a positive relationship between research self-efficacy and research knowledge.

H5: There is a positive impact of research mentoring, research training environment, interest in research, and research knowledge on research self-efficacy.

Literature Review

Research self-efficacy

The broad concept of self-efficacy was a relevant indicator many areas such as academic (Collins, 1982; Lee & Ciftci, 2014; Multon, Brown & Lent, 1991; Schunk & Rice, 1993) and research (Bako-Okolo, 1993; Bieschke, Bishop & Garcia, 1996; Faghihi, Rakow & Ethington, 1999; Holden, Barker, Meenaghan & Rosenberg, 1999; Lambie, Hayes, Griffith, Limberg & Mullen, 2014; Pasupathy, 2010; Petko, 2012; Phillips, 1992; Rezaei & Miandashti, 2013; Vaccaro, 2009; Welzer-Ward, Baltes, Hoffman-Kipp & Lynn, 2010). Table 1 displays a summary of self-efficacy research in other areas.

Table 1

Summary of Self-Efficacy Research in Other Areas of Study.

Area of Study	*Authors*
Counseling	Al-Darmaki (2012); Chaney, Hammond, Betz & Multon (2007); Lam, Tracz & Lucey (2013)
Healthcare	Byrne, Barry & Petry (2012); Callaghan (2005); Chiu & Tsai (2014); Fillman (2015); Grembowski et al. (1993); Longmore, Manning, Giordano & Rudolph (2003); Mackenzie & Peragine (2003); Robb (2012); Shao, Chuang & Chen (2015); Torres, Torres, Rodríguez & Lee (2003); Zubaran et al. (2010)
Parenting	Coleman & Karraker (2000)
Social work	Ellet, (2007); Pennanen, Haukkala, Vries & Vartiainen (2011); Reeb, Folger, Langsner, Ryan & Crouse (2010); Rodebaugh (2006)
Teacher	Aydemir, Duran, Kapıdere, Kaleci & Aksoy (2014); Frazier & Osaghae (2011); Jaafar et al. (2012); Morris & Usher (2011); Raudenbush, Rowan & Cheong (1992)
Internet	Chuang, Lin & Tsai (2015); Kaya & Durmuş (2010); Kuo, Walker, Schroder & Belland (2014)
Computer	Hodges (2008); Jegede (2014); Kher, Downey & Monk (2013); Pellas (2014)
Mathematics	Jameson & Fusco (2014); Pampaka, Kleanthous, Hutcheson & Wake (2011)
Sports	Lane et al. (2004)
Online technology	Bergey, Jass, Senfeng & Uma (2015); Lee (2015); Wang, Harris, & Patterson (2013)
Mobile learning	Mahat, Ayub & Luan (2012)
Reading and writing	Leader-Janssen & Rankin-Erickson (2013); Pajares (2003); Prat-Sala & Redford (2012)
Dissertation	Varney (2010)

Hartadiyati (2015) made an astonishing finding in teacher self-efficacy where the concept of self-efficacy was utilized as a bridge that connects the knowledge of instructional strategies of science teaching to students' understanding of science knowledge. The researcher proposed that teacher and students' self-efficacy beliefs was improved in pedagogy content knowledge (PCK) through modeling, sharing the story of self-efficacies in science, constructive feedback and effective use of lesson study. The recommendation of increasing teacher self-efficacy using collaborative skills and classroom instruction was also discussed in Malinen, Savolainen, and Xu (2012). Self-efficacy also influences teachers' academic performance (GPA) in the Iranian English teachers (Ghonsooly, Hassan & Mohaghegh, 2014).

The meta-analysis of Chesnut and Burley (2015) indicates that pre-service and in-service teachers' self-efficacy was positively related to their commitment in the teaching profession. Research self-efficacy was a derivation from the Social Cognitive Theory by psychologist Albert Bandura (Bandura, 1977a, 1982, 1986, 1995, 1997). Social Cognitive Theory also acts as a basis for understanding the influence of teacher self-efficacy beliefs in sustaining their satisfaction and the students' achievement (Caprara, Barbaranelli, Steca & Malone, 2006). While a variety of definitions for research self-efficacy were available in current research, the study used the definition suggested by Forester, Kahn and Hesson-McInnis (2004) who described research self-efficacy as "the confidence in carrying out research activities from the point of organizing a research plan to carrying out the research process from library research and reading to writing and publication."

In fact, social work educators have identified the main reason that caused students to dislike research was the feeling of uncertainty in their own abilities to conduct research (Epstein, 1987; Reissman, 1993; Royse & Rompf, 1992; Wainstock, 1994). Furthermore, Bieschke, Bishop & Herbert (1995) suggested that research self-efficacy determines the perseverance of a student's research behaviour. The accurate assessment of research self-efficacy may assist the education faculty to identify a given student's self-identified strengths and weaknesses with respect to research, thereby facilitating graduate research. For instance, the acceptance of Bandura's theory of self-efficacy forms a basis for measures of self-efficacy; which leads to useful predictions of future productivity and mastery of research skills (Brown, Lent, Ryan & McPartland, 1996; Forester et al., 2004; Lynch, Zhang & Korr, 2009; Multon, Karen, Brown & Lent, 1991; Schunk, 1995; Zimmerman, 1995). Additionally, the completion of the thesis was an essential part of a Ph.D. program; and there were no alternatives to graduation except the completion of the thesis (Isaac, Quinlan & Walker, 1992). Zhao, McCormick and Hoekman (2008) and Vasil (1992) found research self-efficacy and research productivity to be statistically significant and differ between genders. Males had higher research self-efficacies and research productivities than females.

A considerable amount of research has been published on research self-efficacy. These studies have led to significant outcomes on students' tendency towards research. For instance, lower research self-efficacy can interfere with students' research training and students' willingness to conduct research and result in a lower contribution to their field of study (Love et al., 2007). Lynch, Zhang and Korr (2009) conclude that social work practitioners prefer to participate in research activities when they have a stronger level of research self-efficacy. The findings were consistent with those in Unrau and Grinnell (2005) where the authors revealed that the research on confidence levels and gains were important dimensions of students' learning experience in social work research courses.

Forester et al. (2004) found that high research self-efficacy was a crucial factor related to students successfully conducting research and pursuing research beyond graduate study. Lambie, Hayes, Griffith, Limberg and Mullen (2014) and Lambie and Vaccaro (2011) came to a conclusion that research self-efficacy scores predicted higher levels of interest in research and research knowledge. Additionally, the students who

participated in research activities, including publishing manuscripts, scored higher in research self-efficacy than those not participated in the publication process. Faghihi et al. (1999) concluded that higher self-efficacy in research was positively related to students' dissertation progress.

Their study proved that research self-efficacy was the most important factor influencing students' dissertation progress and it confirmed a few of the previous studies that demonstrated self-efficacy as being an important predictor of students' persistence and academic achievement (Landino & Owen, 1988; Lent, Brown & Larkin, 1984; Multon, Karen, Brown & Lent, 1991). On the other hand, Lambie and Vaccaro (2011) also identified that doctoral students in their 3rd year of preparation had higher research self-efficacy scores than did 1st and 2nd year students. The results for the relationship between research self-efficacy and year of study consists of diverse results; for instance, the results in Lambie and Vaccaro (2011) were consistent with those in Kahn and Scott (1997) and Lambie et al. (2014). However, Kahn (2001) found that there was no significant relationship between research self-efficacy and year of study. The difference of the results may be related to the difference in samples or difference in the course of study. Besides, prior studies used different scales in measuring research self-efficacy (Greeley et al., 1989).

Vaccaro (2009) identified that research self-efficacy was positively correlated with interest in research but the results contradicted with the work of Petko (2012) where it was reported that research self-efficacy was positively correlated with interest in research but negatively correlated with research mentoring. Furthermore, Petko (2012) found that interest in research and research mentoring did not have a statistically significant relationship. The research of Love et al., (2007) concluded that research self-efficacy was influenced by a positive mentoring environment while Bishop and Bieschke (1998) came to a conclusion that the factors of research self-efficacy and the mentoring environment had an influence on research interest. The findings from Bishop and Bieschke (1998) was supported by numerous studies concluding that interest in research had an influence on research self-efficacy (Bieschke, Bishop & Herbert, 1995;.Bieschke, Bishop & Garcia, 1996; Bieschke, 2006).

Despite the emergence of growing literature supporting the positive effects of interest in research on research self-efficacy, Kahn (2001) concluded that interest in research only has an indirect relationship to research self-efficacy. Kahn (2001) reported that levels of research self-efficacy increased in tandem with participation in research related activities. Research self-efficacy was found to be a good predictor of student interest in conducting research and related activities (Bishop & Bieschke, 1998; Kahn & Scott, 1997; Lent, Lopez & Bieschke, 1993; Lopez, Lent, Brown & Gore, 1997). Szymanski, Ozegovic, Phillips and Briggs-Phillips (2007) concluded that research self-efficacy and interest in research scores predict levels of scholarly productivity.

The findings from Boswell (2014) were rather intriguing; as it was found that research self-efficacy did not equate to competence, and it was possible that participants might overestimate their abilities in conducting research. Individuals have the tendency to over-rate their competence and base their performance on

incorrect pre-conceived notions. Thus, the situation leads individuals to make incorrect judgments about their own abilities (Dunning, Johnson, Ehrlinger & Kruger, 2003; Dunning, Heath & Suls, 2004). Furthermore, Pan, Sun and Chow (2011) found that the impact of mentoring was weaker on individuals with lower self-efficacy. Huge amount of literature has shown that students scoring at higher levels of interest in research had higher levels of research self-efficacies than those who score lower in interest in research.

Lastly, Kahn (2001) and Phillips and Russell (1994) reported that research training environment was proven as a strong predictor of research self-efficacy. Phillips and Russell (1994) concluded that counseling psychology doctoral students' levels of research self-efficacy and perceptions of the training environment were positively correlated. Although positive correlations exist between the two constructs, Saral and Didem (2015) made an interesting finding in their study where they found that educational students' research self-efficacy varied according to the different departments they came from such as psychological counseling and guidance, elementary school mathematics teaching, science teaching, and computer and teaching technologies.

Method

Research Design

Extensive data filtering was employed to ensure the reliability of the data prior to data analysis. Data mean were used to determine the level of research self-efficacy, research training environment, interest in research, and research mentoring while the frequency of scores was employed to ascertain the level of research knowledge of the respondents. Correlational analysis was conducted to find out the relationships between research self-efficacy, research training environment, and interest in research, research mentoring and research knowledge. Regression analysis was employed to find out the impact of research mentoring, research training environment, interest in research, and research knowledge on research self-efficacy. Moderate levels of research self-efficacy and high levels in the research training environment, interest in research and research mentoring was found among the respondents. Unfortunately, the majority of them have low levels of research knowledge. An apparent correlation was predicated between research self-efficacy, research training environment, interest in research and research mentoring while research self-efficacy was not correlated with research knowledge. Results from regression analysis showed that research training environment, interest in research contributed to 26 percent of the variance of changes in research self-efficacy.

A correlational design was implemented which included the following statistical analyses: Pearson's correlation coefficients (two-tailed) and independent t-test. Data were collected from four faculties in the university. A total of five (5) instruments and one (1) demographic survey was used to measure the constructs involved in the study.

Research Sample

The sample size formula was estimated according to the prescriptions in Krejcie and Morgan (1970). The researcher utilizes a power analysis formula for experiments in calculating the sampling errors (Cohen, 1988; Lipsey, 1990; Murphy & Myors, 1998). A stratified sampling was used by the researcher to obtain all relevant information on research self-efficacy. Stratified sampling is a sampling method whereby the populations are divided into subpopulations (stratas) and random samples are drawn from each stratum (Laerd Statistics, 2012). A first strata was drawn from a population of university students pursuing their Ph.D. study.

The population consists of students from four different faculties (Management, Graduate School of Business, Education and Housing, Building and Planning). Out of a total of 809 students, second strata were drawn and further divided into two groups; mainly students who enrolled for less than 3 years; and another group of students who enrolled for more than 3 years.

Research Instrument and Procedures

The level of research self-efficacy was measured using the self-efficacy in research (SERM) brief form. This is a 12-item self-report measure of a doctoral students' self-efficacy with respect to doing research (Kahn & Scott, 1997). It contains items assessing self-efficacy with respect to research design skills, practical research skills, quantitative and computer skills, and writing skills. Items were derived from Phillips and Russell's (1994) 33-item Self-Efficacy in Research Measure. Secondly, the role of research mentoring was measured by the research mentoring experience scale (RMES). An 18-item short form (Kahn & Miller, 2000) of the Research Training Environment Scale (Gelso et al., 1996) used to measure global perceptions of the RTE of the interpersonal and instructional dimensions.

The third instrument used in the study involved research training environment revised (RTES-r) scale which measures students' perceptions on the research training environment. Interest in research was measured by the interest in research scale. The Interest in Research Questionnaire (IRQ) was developed by Bishop & Bieschke (1994). The IRQ is a 16-item 5-item Likert response scale that contains responses that range from 1 (very disinterested) to 5 (very interested). (Hollingsworth & Fassinger, 2002). The questionnaire intends to measure interest in research-oriented activities.

The respondents evaluate their degree of interest in a particular research task such as taking a research design course and analyzing data on a 5-point Likert scale ranging from 1 (very disinterested) to 5 (very interested).

An example of the items in the IRQ involves reading research from a journal article, conducting a literature review, analyzing data, and designing a study.

Lastly, a research knowledge assessment by Lambie (2012) was used to find out the level of research knowledge among the Ph.D. students. It covers eight subscales:

(a) literature review, (b) ethics in educational research, (c) research designs, (d) sampling, (e) data collection methodologies, (f) data analysis procedures, (g) data reporting and (h) scholarly writing practices. Each participant was given an answer sheet to circle the answers.

Additionally, a short demographic questionnaire was also distributed to the participants to find out general information on the participants such as age, gender, year of study and faculty.

Data Analysis

The information received from the survey packet was scored using the Statistical Package for Social Science (SPSS) software version 23. The objective of the study was to find out the level of research self-efficacy and its relationships with research mentoring, research training environment, interest in research, and research knowledge. Firstly, descriptive statistics, frequencies, mean and standard deviation were used to analyze the demographic factors of the respondents.

Secondly, Pearson's correlation coefficient was used to examine the relationship between research self-efficacy and research mentoring, research training environment, interest in research and research knowledge. Research self-efficacy was treated as the dependent variable while research mentoring, research training environment, interest in research and research knowledge were the independent variables. Thirdly, a multiple regression analysis was conducted to find out the impact of research mentoring, research training environment, interest in research and research knowledge on research self-efficacy.

Results

Reliability Analysis

Reliability analysis was employed to determine whether the items measuring research self-efficacy, research mentoring, research training environment and interest in research was a reliable measure of these variables. In order to measure the reliability of the items, Cronbach alpha's (α) statistics were calculated using Statistical Package for the Social Sciences (SPSS) version 23 (Cronbach, 1951). The observed value of .70 concludes the items included in the survey measured the intended variable (Nunnally, 1978).

The self-efficacy subscale consists of 12 items (α = .86). The Cronbach's alpha for the self-efficacy items was .86 and was found to be highly reliable (12 items; α =.86). The research training environment subscale consists of 18 items (α =.78).

The Cronbach's alpha for the research training environment items was .78. Items deleted were RTE1, RTE4, RTE5, RTE8, RTE9, RTE10, RTE13 and RTE17.

The interest in research subscale consists of 16 items (α =.86). The Cronbach's alpha for the self-efficacy items was .86 and was found to be highly reliable (16 items; α =.86). The research mentoring subscale consists of 28 items (α =.95). The Cronbach's alpha for the research mentoring items was .95 and was found to be highly reliable (28 items; α =.95).

Additionally, the research knowledge assessment was a 50 multiple choice assessment whereby the respondents were required to circle the correct answer on an answer sheet. Then, each respondent was given a score out of a total mark of 50. All items in the assessment was a dichotomous response with a choice of a, b, c and d. Therefore, the Kuder-Richardson Reliability Coefficients (KR20) was used to check the reliability of the 50 items in the RKA. Lord and Novick (1968) strongly proposed the KR20 as a better reliability estimate that KR21. A "Correct" answer was coded as 1.00 and an "Incorrect" answer was coded as .00 in Statistical Package for the Social Sciences (SPSS).

Table 2

Reliability Statistics for Research Self-Efficacy, Research Training Environment, Interest in Research, Research Mentoring Experience and Research Knowledge Assessment.

	Number of items	***Cronbach's Alpha***
RSE	12	.86
RTE	9	.78
IR	16	.86
RME	28	.95
RKA	120	.89

Note: RSE: Research self-efficacy, RTE: Research training environment, IR: Interest in research, RME: Research mentoring experience, RKA: Research knowledge

Table 2 illustrates that the research knowledge assessment subscale consists of 50 items (*a* =.89). The Cronbach's alpha for the research knowledge assessment was .89 and was found to be highly reliable (50 items, *a* =.89).

Correlation Analysis

The findings revealed that the difference in gender resulted in different levels of research self-efficacy while the students' diverse faculty had a difference in their perceptions of the research training environment. An apparent correlation was predicated between research self-efficacy, research training environment, interest in research, and research mentoring.

Table 3 presents the results of correlation analysis between research self-efficacy, research training environment, interest in research, and research mentoring.

Table 3

Correlation Analysis between Research Self-Efficacy, Research Training Environment, Interest in Research, Research Mentoring Experience, and Research Knowledge

	RSE	*RTE*	*IR*	*RME*	*RKA*
RSE	1	.36**	.46**	.31**	-.02
RTE		1	.37**	.49**	.21*
IR			1	.44**	.10
RME				1	.11
RKA					1

***Sig. at p <.01; * sig. at p <.05*

A moderate correlation existed between research self-efficacy and research training environment, $r = .36$, $p < .05$. A moderate correlation existed between research self-efficacy and interest in research, $r = .46$, $p < .05$. A moderate correlation existed between research self-efficacy and research mentoring experience, $r = .31$, $p < .05$. Lastly, a negative correlation existed between research self-efficacy and research knowledge $r = -.02$, $p > .05$.

A moderate correlation existed between research training environment and interest in research $r = .37$, $p < .05$. There was a moderate correlation between the research training environment and research mentoring experience where $r = .49$, $p < .05$ and a weak correlation between the research training environment and research knowledge where $r = .21$, $p < .05$.

A moderate correlation was found between interest in research and research mentoring experience where $r = .44$, $p < .05$. However, there was no correlation between interest in research and research knowledge where $r = .10$, $p > .05$. Finally, there was no correlation between research mentoring and research knowledge where $r = .11$, $p > .05$.

Table 4 presents the results of the mean square and F value multiple regression analysis.

Table 4

Impacts of Research Training Environment, Interest in Research, Research Mentoring Experience, and Research Knowledge on Research Self-Efficacy

Predictor	***Research self-efficacy***
RTE	.23*
IR	.36*
RME	.05
RKA	-.11
R Value	.51
R^2 Value	.26
Adjusted R^2	.24
F Value	10.32

*Note: *p<.05;*
RSE: Self-efficacy, RTE: Research training environment, IR: Interest in research, RME: Research mentoring experience, RKA: Research knowledge

A significant regression equation was found (F (4,115 = 10.323, p < .00) with an R^2 of .26. Interest in research contributed to 36 percent of the variance of changes on research self-efficacy. The result showed that research training environment (β = .23; p = .019) and interest in research (β = .36; p = .00) had significant impacts on research self-efficacy. Unfortunately, research mentoring experience (β = .05; p = .62) and research knowledge (β = -.11; p = .20) had no impact on research self-efficacy. The results of research training environment, interest in research, research mentoring experience, and research knowledge on research self-efficacy are presented in Table 5.

Table 5

Impacts of Research Training Environment, Interest in Research, Research Mentoring Experience, and Research Knowledge on Research Self-Efficacy

Predictor	*Research self-efficacy*
RTE	.23*
IR	.36*
RME	.05
RKA	-.11
R Value	.51
R^2 Value	.26
Adjusted R^2	.24
F Value	10.32

*Note: *p<.05; RSE: Self-efficacy, RTE: Research training environment, IR: Interest in research, RME: Research mentoring experience, RKA: Research knowledge*

Research training environment, interest in research contributed to 26 percent of the variance of changes in research self-efficacy. All predictors except research mentoring experience and research knowledge had a significant impact on research self-efficacy with a range of $.23 \leq \beta \geq .36$ at $p < .05$. Therefore, the null hypothesis was rejected as there was a positive impact of the research training environment and interest in research on research self-efficacy and a negative impact of research mentoring experience, and research knowledge on research self-efficacy.

Table 6 presents the summary of findings for all research hypothesis set in the study.

Table 6

Summary of Findings for Research Hypothesis

Hypothesis	*Statement*	*Findings*
Ho1	There is a positive relationship between research self-efficacy and research mentoring experience.	**Accepted**
Ho2	There is a positive relationship between research self-efficacy and research training environment.	**Accepted**
Ho3	There is a positive relationship between research self-efficacy and interest in research.	**Accepted**
Ho4	There is a positive relationship between research self-efficacy and research knowledge.	**Rejected**
Ho5	There is a positive impact of research mentoring, research training environment, interest in research, and research knowledge on research self-efficacy.	**Rejected**

Note: Ho: Hypothesis

Discussion, Conclusion and Recommendations

In conclusion, this study was the first to examine the specific constructs of research self-efficacy such as the perceptions of the research training environment, research mentoring, interest in research, and research knowledge within a sample of Ph.D. students in a local university. The current study was initiated to provide an original contribution to the literature in the education field as research self-efficacy is the most popular and predictive construct in behavior research.

Self-efficacy ratings are highly predictive of behavior as the ratings reflect a broad range of predictive motives (Williams & Rhodes, 2014). One of the purported strengths of the self-efficacy is the one that provides the reasoning as to why individuals have the tendency to engage in specific behaviors rather than merely just predicting the behavior. The self-efficacy theory by Bandura (1977) clearly explains that self-efficacy determines the behavior of an individual and the setting of relevant goals, and challenges are essential towards sculpting the right behavior.

A few significant findings were drawn from the study. Generally, the results suggested that Ph.D. students possess moderate levels of research self-efficacy and low levels of research knowledge. On the contrary, the students have shown high levels in the research training environment, research mentoring and interest in research. Remedial actions such as online delivery and including the research as a module in the Ph.D. study was essential to provide additional assistance to the students to improve their research knowledge. Other viable suggestion included peer modeling which serves as a solution towards improving the research knowledge of the Ph.D. students. Scaffolding through mentoring and the opportunity for the students to present in academic conferences and publish their research skills could help to increase the student's research self-efficacy (Wyatt & Dikilitas, 2015).

Secondly, positive correlations were found between the research training environment, research mentoring, interest in research, and research self-efficacy. Other than that, interest in research and research training environment was accountable for twenty-six percent of the variation in research self-efficacy. The level of interest in research enabled the Ph.D. student to actively conduct research. The self-fulfilling prophecy method was recommended to the students, in which good students convey to their peers to the belief that they are capable in conducting research and likely to excel; and the peers, in turn, become more confident and believe they will excel in research and continuously achieve higher levels of performance. Self-efficacy beliefs do not form in isolation but are influenced by the forces that shape the understanding and intensity of the belief. Mutual cooperation between the student and the faculty involved was important to ensure success in actively fostering and improve the research self-efficacy, research training environment, interest in research, research mentoring, and research knowledge of the students.

Implications for self-efficacy theory

The self-efficacy theory explains that reciprocal interactions between personal, environmental and behavioral factors play an important role in shaping the desired behavior of an individual. Perceived self-efficacy influences an individual's choice of activities and behavioral settings, how much effort they spend, and how long they will persist in the face of obstacles. Understanding research self-efficacy is important to provide guidance to scholars on the deficiency of status quo, improve and evaluate university education.

Overall results have shown that the students possess a moderate level of research self-efficacy. On the positive side, students have shown high levels of the research training environment, interest in research and research mentoring. Correlation findings have shown a positive relationship between research self-efficacy, research training environment, interest in research, and research mentoring. Negative relationships were found between research self-efficacy and research knowledge. The implication of self-efficacy was useful for selection decisions where students with better confidence in research are selected to represent the university in academic conferences and presentation.

Besides, the theory also enabled the faculty to carefully select students with high research self-efficacy to be groomed as future researchers. The accurate selection of student results in higher productivity in research for the university. The theory was also useful in goal setting where specific goals set towards the completion of research will ensure that the students graduate on time and not extend their candidature period.

Implications for practices in research

The findings served as an essential input to the faculty as it provided robust understandings on research self-efficacy and research knowledge that the Ph.D. student possessed, level of interest and mentoring, and their perceptions of research training environment throughout their study. Furthermore, the study was the first to incorporate the few constructs relevant to the area of research self-efficacy.

The findings on the perception of the research training environment provided a better understanding of the current research training environment and all the forces that were within the faculty. It served as an indicator to inform whether the facilities, peers and support staff should be modified to improve the research experience of the Ph.D. student. The moderate level of research self-efficacy and low levels of research knowledge has shown that the faculty must take initiatives to improve the confidence level and provide essential support to the Ph.D. students. Remedial classes, online delivery of research courses, and incorporation of a research module in the program could serve as a viable solution to improve the research self-efficacy and research knowledge of the students.

High interest in research was evident among the respondents in the study and this helped the faculty to identify the potential student for succession planning in research in order to sustain for the future developments in the faculty. Additionally, the students also had rated their Ph.D. supervisors highly. This was a very encouraging and positive finding for the faculty as the commitment of the supervisors were appreciated by the students. Nevertheless, the faculty must continue to strengthen the research clusters within the academic staff and provide the highest level of mentoring to the students.

The findings from the study play an important role to create an innovative ecosystem that facilitates research and development by understanding the basics of research confidence amongst the students. By detecting the flaws in the level of research confidence which affects the research behavior of the students, it enables the faculty to improvise the current situations to facilitate a better environment towards research.

Lastly, the findings from the study provide insights for Ph.D. preparation programs to continue to develop and prepare students for success. It provides beneficial information to assist universities to develop and design policies to encourage and foster research amongst their Ph.D. students. It helps to further enhance the success rate and quality of the Ph.D. programs offered by the university.

Implications for the Ph.D. student

The findings largely contribute to the gaps in the studies in research self-efficacy and support Ph.D. students' development in research. The findings from the study have several important implications for Ph.D. student. Results showed that students possessed moderate levels of research self-efficacy and low levels of research knowledge but high levels of interest in research, research mentoring, and research training environment. Research training environment plays an important role in shaping the research behavior and attitudes of the student as it encompasses a wide range of forces such the faculty, students, and support staff. The findings serve as an important indicator for the student as he/she may improve on the deficiencies in the identified area. Besides, findings were very useful for ambivalent students who have just enrolled in the Ph.D. programs as a positive experience in the training environment contributes to greater self-efficacy. The incremental development of research self-efficacies is crucial in the early part of their career.

Limitations

A few limitations were identified in the study. The current study was non-experimental in design; the investigation employed an ex-post-facto or correlation research design. For studies that are not by nature experimental, a cause and effect (or causality) relationship cannot be inferred as a result since the research is related to the associations between the variables (Sproull, 1995; Tuckman, 1972).

The study did not involve random sampling but rather employed stratified sampling because the sample of the Ph.D. students was drawn from students in the applied arts faculties. Therefore, the results may not be generalized to other fields of study. In addition, the study was ex-post-facto in design. Ex-post-facto suggests "what was done afterward." Kerlinger (1964) defined ex post facto research as a "research in which the independent variable or variables have already occurred and in which the researcher starts with the observation of a dependent variable or variables.

Recommendations for future research

The following are several suggestions for future areas of investigation related to the present study. It is prescribed that the current study be replicated over a period of time which permits a pre and post-study for the Ph.D. students and extend into the participants' post-doctoral work. Longitudinal research design would allow students' progress to be tracked from the beginning of their programs through completion, and into the professional arena. The inclusion of the constructs investigated in the study with selected demographic variables such as gender provides a possible new finding in research self-efficacy. Findings would provide valuable information on gender as a potential factor which influences research self-efficacy and contributes to what is, at the present, a paucity of literature. Further research on research self-efficacy and gender present new ideas for research as women consistently underrated their self-confidence when compared to men (Edward & Hopkins, 2005; Roest & Kleiner, 2010).

References

Ahmad Kamil, Y. (2016). 20,000 MyBrain15 scholarships this year. Retrieved February 25, 2016, from http://www.thestar.com.my/news/nation/2016/02/25/20000-mybrain15-scholarships-this-year/.

Ahmad, S. S. (2012). Performance Indicators for the Advancement of Malaysian Research with Focus on Social Science and Humanities. *Procedia - Social and Behavioral Sciences*, *68*, 16–28. doi:10.1016/j.sbspro.2012.12.203.

Aziz, H. (2016). Making waves with our Ph.D. programmes in Malaysia. Retrieved March 3, 2016, from http://www.nst.com.my/news/2016/03/130569/making-waves-our-Ph.D.-programmes.

Al-Darmaki, F. R. (2012). Relationships Between Career Counseling Attitudes and Self-Esteem and Self-Efficacy Among Emirati University Students. Journal of Career Development, 39(2), 143–161. doi:10.1177/0894845310380046.

Aydemir, H., Duran, M., Kapıdere, M., Kaleci, D., & Aksoy, N. D. (2014). Self-efficacy of Teacher Candidates Intended Teaching Profession. Procedia - Social and Behavioral Sciences, 152, 161–166. doi:10.1016/j.sbspro.2014.09.174

Bishop, R. M., & Bieschke, K. J. (1994). Interest in Research Questionnaire (Unpublished Scale). United States.

Bernardin, H. J. (1996). Academic research under siege: Toward better operational definitions of scholarship to increase effectiveness, efficiencies and productivity. Human Resource Management Review, 6(3), 207–229.

Bowen, W. G., & Rudenstine, N. L. (2014). In pursuit of the Ph.D. Princeton, NJ: Princeton University Press.

Bandura, A., & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. Cognitive Therapy and Research, 1(4), 287–310. doi:10.1007/BF01663995.

Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37(2), 122–147.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs: NJ: Prentice Hall.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In Self-efficacy in changing society. New York NY: Cambridge University Press.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York NY: Freeman.

Bandura, A. (1988c). Self-regulation of motivation and action through goal systems. Cognitive Perspectives on Emotion and Motivation. Dordrecht: Kluwer Academic Publishers.

Bieschke, K. J., Bishop, R. M., & Herbert, J. T. (1995). Research interest among rehabilitation. Rehabilitation Education, 9, 51–66.

Bieschke, K. J. J., Bishop, R. M., & Garcia, V. L. (1996). The Utility of the Research Self-Efficacy Scale. Journal of Career Assessment, 4, 59–75.

Boswell, S. S. (2014). Change In Undergraduates' Efficacy: A Pilot Study Self-Efficacy. The Online Journal of New Horizons in Education, 4(4), 90–96.

Brown, S. D., Lent, Lent, R. W., Ryan, N. E., & McPartland, E. B. (1996). Self-efficacy as an intervening mechanism between research training environments and scholarly productivity: A theoretical and methodological extension. The Counseling Psychologist, 24(3), 535–544. doi:0803973233.

Bako-Okolo, S. (1993). The effects of selected variables on the research self-efficacy of doctoral candidates in the college of education at the University of Akron. Retrieved June 3, 2014, from http://Ph.D.tree.org/pdf/24825111-the-effects-of-selected-variables-on-the-research-self-efficacy-of-doctoral-candidates-in-the-college-of-education-at-the-university-of-akron/.

Byrne, S., Barry, D., & Petry, N. M. (2012). Predictors of weight loss success. Exercise vs. dietary self-efficacy and treatment attendance. Appetite, 58(2), 695–698. doi:10.1016/j.appet.2012.01.005.

Bergey, B. W., Jass, D., Senfeng, K., & Uma, L. (2015). Scientific Inquiry Self-Efficacy and Computer Game Self-Efficacy as Predictors and Outcomes of Middle School Boys' and Girls' Performance in a Science Assessment in a Virtual Environment. Journal of Science Educational Technology. doi:10.1007/s10956-015-9558-4.

Cassuto, L. (2013). Ph.D. attrition: How much is too much? Chronicle of Higher Education. Retrieved May 1, 2015, from http://chronicle.com/article/Ph.D.-Attrition-How-Much-Is/140045/.

Collins, J. L. (1982). Self-efficacy and ability in achievement behavior. In American Educational Research Association. New York NY.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Chin, J. (2015). The 11th Malaysia Plan and Economic Highlights. Retrieved February 25, 2015, from http://www.thestar.com.my/business/business-news/2015/05/21/11th-malaysia-plan-economic-highlights/.

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. Educational Research Review, 15, 1–16. doi:10.1016/j.edurev.2015.02.001.

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(6), 473–490. doi:10.1016/j.jsp.2006.09.001.

Carvalho, M., & Rahim, R. (2016). Malaysia and Bangladesh ink deal to recruit 1.5 million workers. Retrieved February 25, 2015, from http://www.thestar.com.my/news/nation/2016/02/18/bangladeshi-workers-malaysia-mou/.

Chaney, D., Hammond, M. S., Betz, N. E., & Multon, K. D. (2007). The Reliability and Factor Structure of the Career Decision Self-Efficacy Scale (SES) With African Americans. Journal of Career Assessment, 15(2), 194–205. doi:10.1177/1069072706298020.

Callaghan, D. (2005). Healthy behaviors, self-efficacy (SE), self-care, and basic conditioning factors in older adults. Retrieved May 1, 2014, from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2005380060.

Chiu, Y. L., & Tsai, C. C. (2014). The roles of social factor and internet self-efficacy in nurses' web-based continuing learning. Nurse Education Today, 34(3), 446–450. doi:10.1016/j.nedt.2013.04.013.

Chuang, S., Lin, F., & Tsai, C. (2015). Computers in Human Behavior An exploration of the relationship between Internet self-efficacy and sources of Internet self-efficacy among Taiwanese university students. Computers in Human Behavior, 48, 147–155. doi:10.1016/j.chb.2015.01.044.

Coleman, P. K., & Karraker, K. H. (2000). Parenting Self-Efficacy Among Mothers of School-Age Children: Conceptualization, Measurement, and Correlates. Family Relations, 49(1), 13–24.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. doi:10.1007/BF02310555.

Department of Statistics (Malaysia). (2014). The Annual Labour Force Survey Report and Census in Malaysia: A concise report. Retrieved from https://www.statistics.gov.my/dosm/index.php?r=column/ctheme&menu_id=U3VPMldoYUxzVzFaYmNkWXZteGduZz09&bul_id=NHUxTlk1czVzMGYwS29mOEc5NUtOQT09.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment Implications for Health,Education and the Workplace. Psychological Science in the Public Interest, 5(3), 69–106.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognise their own incompetence. Current Directions in Psychological Science: A Journal of the American Psychological Society, 12(3), 83.

Else, H. (2014). Malaysia changes targets for postgraduates. Retrieved March 3, 2016, from https://www.timeshighereducation.com/news/malaysia-changes-targets-for-postgraduates/2014159.article.

Evans, L. (2011). The scholarship of researcher development: mapping the terrain and pushing back boundaries. International Journal for Researcher Development, 2(2), 75–98. doi:10.1108/17597511111212691.

Evans, L. (2012). Leadership for researcher development: What research leaders need to know and understand. Educational Management, Administration & Leadership, 40(4), 423–435.

Epstein, I. (1987). Pedagogy of the pertubed: Teaching research to the reluctants. Journal of Teaching in Social Work, 1, 71–89.

Edward, W., & Hopkins, S. A. (2005). Aligning Organizational Subcultures for Competitive Advantage. New York NY: Basic Books.

Forester, M., Kahn, J. H., & Hesson-McInnis, M. S. (2004). Factor Structures of Three Measures of Research Self-Efficacy. *Journal of Career Assessment*, *12*(1), 3–16. doi:10.1177/1069072703257719.

Faghihi, F., Rakow, E. A., & Ethington, C. (1999). A study of factors related to dissertation progress among doctoral candidate: Focus on students' self research efficacy as a result of their research training and experience. In

American Educational Research Association. Montreal. doi:ERIC Document Reproduction Service No. ED430491.

Fullick, M. (2013). War of attrition – Asking why Ph.D. students leave. Retrieved April 11, 2015, from http://www.universityaffairs.ca/speculative-diction/war-of-attrition-asking-why-Ph.D.-students-leave/.

Fillman, V. M. (2015). Career interest, self-efficacy and perception in undecided and nursing undergraduate students : A quantitative study. Nurse Education Today, 35(1), 251–255. doi:10.1016/j.nedt.2014.08.011.

Frazier, P., & Osaghae, O. J. (2011). From Maybe I Can to Yes I Can: Developing Preservice and Inservice Teachers' Self-Efficacy to Teach African American Students. Journal of Negro Education, 80(3), 209–222.

Greeley, A. T., Johnson, E., Seem, S., Braver, M., Dias, L., & Evans, K. (1989). Research Self-Efficacy Scale (unpublished scale).

Grembowski, D., Patrick, D., Diehr, P., Durham, M., Beresford, S., Kay, E., & Hecht, J. (1993). Self-efficacy and health behavior among older adults. Journal of Health and Social Behavior, 34(2), 89–104. doi:10.2307/2137237.

Ghonsooly, B., Hassan, G., & Mohaghegh, F. (2014). Self-efficacy and Metacognition as Predictors of Iranian Teacher Trainees' Academic Performance: A Path Analysis Approach. Procedia-Social and Behavioral Sciences, 98, 590–598. doi:10.1016/j.sbspro.2014.03.455.

Gelso, C. J., Mallinckrodt, B., & Judge, A. B. (1996). Research Training Environment, Attitudes toward Research, and Research Self-Efficacy: The Revised Research Training Environment Scale. The Counseling Psychologist, 24(2), 304–322. doi:10.1177/0011000096242010.

Hollingsworth, Merris A.; Fassinger, R. E. (2002). The role of faculty mentors in the research training of counseling psychology doctoral students. Journal of Counseling Psychology, 49(3), 324–330. doi:10.1037/0022-0167.49.3.324.

Holden, G., Barker, K., Meenaghan, T., & Rosenberg, G. (1999). Research self-efficacy: A new possibility for educational outcomes assessment. Journal of Social Work Education, 35(3), 463–477.

Hodges, C. B. (2008). Implications of Domain-General "Psychological Support Skills" for Transfer of Skill and Acquisition of Expertise. Performance Improvement Quarterly, 20(3-4), 7–25. doi:10.1002/piq.

Hartadiyati, E. W. H. (2015). Self-efficacy as affective domain of pedagogy content knowledge (PCK): The Implications for Pre-Service and In-Service teachers in Science teaching. International Journal of Education and Research, 3(1), 391–404.

Isaac, P. D., Quinlan, V., & Walker, M. M. (1992). Faculty Perceptions of the Doctoral Dissertation. The Journal of Higher Education, 63(3), 241–268.

Jaafar, N., Tamuri, A. H., Muhamad, N. A. F., Ghazali, N. M., Amat, R. A. M., Raus, N. M., & Hassan, S. N. S. (2012). The Importance of Self-Efficacy: A Need for Islamic Teachers as Murabbi. Procedia - Social and Behavioral Sciences, 69(International Conference on Education and Educational Psychology (ICEEPSY 2012)), 359–366. doi:10.1016/j.sbspro.2012.11.421.

Jegede, P. O. (2014). Journal of Psychology in Africa Factors in Computer Self-Efficacy Among Nigerian College of Education Teachers Factors in Computer Self-Efficacy Among Nigerian College of. Journal of Psychology in Africa, 17(1), 39–44. doi:10.1080/14330237.2007.10820142.

Jameson, M. M., & Fusco, B. R. (2014). Math Anxiety, Math Self-Concept, and Math Self-Efficacy in Adult Learners Compared to Traditional Undergraduate Students. Adult Education Quarterly, 64(4), 306–322. doi:10.1177/0741713614541461.

Krejcie, R. V, & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, *38*, 607–610.

Kaya, S., & Durmuş, A. (2010). Pre-service teachers' perceived internet self-efficacy and levels of internet use for research. Procedia - Social and Behavioral Sciences, 2(2), 4370–4376. doi:10.1016/j.sbspro.2010.03.695

Koo, Y. L., Fazal Mohamed, M. S., & Kemboja, I. (2012). *Memperkukuh Dasar Pengajian Tinggi Melalui Penyelidikan* (IX-X). Ministry of Higher Education.

Kahn, J. H., & Scott, N. A. (1997). Predictors of Research Productivity and Science-Related Career Goals among Counseling Psychology Doctoral Students. The Counseling Psychologist, 25(1), 38–67. doi:10.1177/0011000097251005.

Kahn, J. H. (2001). Predicting the scholarly activity of counseling psychology students: A refinement and extension. Journal of Counseling Psychology, 48, 344–354.

Kahn, J. H., & Miller, S. A. A. (2000). Measuring global perceptions of the research training environment using a short form of the RTES-R. Retrieved May 5, 2015, from http://psycnet.apa.org/psycinfo/2000-05649-003.

Kerlin, S. P. (1995). Pursuit of the Ph.D.: Survival of the fittest, or is it time for a new approach. Education Policy Analysis Archives, 3(16), 41–68.

Kuther, T. (2015). What is a doctoral candidate? Retrieved May 1, 2015, from http://gradschool.about.com/od/survivinggraduateschool/g/doctoralcand.htm.

Kerlinger, F. N. (1964). Foundations of behavioral research. New York NY: Rinehart & Winstron.

Kuo, Y. C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. Internet and Higher Education, 20, 35–50. doi:10.1016/j.iheduc.2013.10.001.

Kher, H. V., Downey, J. P., & Monk, E. (2013). A longitudinal examination of computer self-efficacy change trajectories during training. Computers in Human Behavior, 29(4), 1816–1824. doi:10.1016/j.chb.2013.02.022.

Leader-Janssen, E. M., & Rankin-Erickson, J. L. (2013). Preservice Teachers' Content Knowledge and Self-Efficacy for Teaching Reading. Literacy Research and Instruction, 52(3), 204–229. doi:10.1080/19388071.2013.781253.

Lambie, G. W. (2012). Research Knowledge Assessment (RKA) Unpublished Scale. Orlando, Florida, USA.

Lam, S., Tracz, S., & Lucey, C. (2013). Age, Gender and Ethnicity of Counsellor Trainees and Corresponding Counselling Self-Efficacy: Research Findings and Implications for Counsellor Educators. International Journal of Advance Counselling, 35, 172–187. doi:10.1007/s10447-012-9175-3.

Laerd Statistics. (2012). Research strategy. Retrieved April 29, 2015, from http://dissertation.laerd.com/purposive-sampling.php total.

Lambie, G. W., & Vaccaro, N. (2011). Doctoral Counselor Education Students ; Levels of Research Self-Efficacy, Perceptions of the Research Training Environment and Interest in Research, 50(June), 243–258.

Lent, R. W. W., Brown, S. D. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. Journal of Counseling Psychology, 33(3), 356–362.

Landino, R. A., & Owen, S. V. (1988). Self-efficacy in university faculty. Journal of Vocational Behavior, 33, 1–14.

Longmore, M. A., Manning, W. D., Giordano, P. C., & Rudolph, J. L. (2003). Contraceptive Self-Efficacy : Does It Influence Adolescents' Contraceptive Use? Journal of Health and Social Behavior, 44(1), 45–60.

Lane, A. M., Devonport, T. J., & Horrell, A. (2004). Self-Efficacy and Research Methods. The Journal of Hospitality Leisure Sport and Tourism, 3(2), 25–37. doi:10.3794/johlste.32.59.

Lambie, G. W., Hayes, B. G., Griffith, C., Limberg, D., & Mullen, P. R. (2014). An Exploratory Investigation of the Research Self-Efficacy, Interest in Research, and Research Knowledge of Ph.D. in Education Students. Innovative Higher Education, 39(2), 139–153. doi:10.1007/s10755-013-9264-1.

Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1993). Predicting mathematics-related academic choice and success: Test of an expanded social cognitive model. Journal of Vocational Behavior, 42, 223–236.

Lee, J. Y., & Ciftci, A. (2014). Asian international students' socio-cultural adaptation: Influence of multicultural personality, assertiveness, academic self-efficacy, and social support. International Journal of Intercultural Relations, 38(1), 97–105. doi:10.1016/j.ijintrel.2013.08.009.

Lee, C.Y. (2015). Changes in self-efficacy and task value in online learning. Distance Education, 36(1), 59–79. doi:10.1080/01587919.2015.1019967.

Lipsey, M. W. (1990). Design sensitivity: Statistical power for applied experimental research. Newbury Park CA: Sage Publications.

Lynch, M. T., Zhang, L., & Korr, W. S. (2009). Research Training, Institutional Support, and Self-Efficacy: Their Impact on Research Activity of Social Workers. Advances in Social Work, 10(2), 193–210.

Love, K. M., Bahner, A. D., Jones, L. N., & Nelson, J. E. (2007). An investigation of early research experience and research self efficacy. Professional Psychology: Research and Practice, 38(3), 314–320.

Lopez, F. G., Lent, R. W., Brown, S. D., & Gore, P. A. (1997). Role of social cognitive expectations in high school students' mathematics-related interest and performance. Journal of Counseling Psychology, 44, 44–52.

Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

Ministry of Higher Education. (2015). Malaysia Education Blueprint 2015-2025 (Higher Education). Retrieved May 23, 2014, from http://hes.moe.gov.my/event/docs/4. Executive Summary PPPM 2015-2025.pdf.

Mahat, J., Ayub, A. F. M., & Luan, S. (2012). An Assessment of Students' Mobile Self-Efficacy, Readiness and Personal Innovativeness towards Mobile Learning in Higher Education in Malaysia. Procedia - Social and Behavioral Sciences, 64, 284–290. doi:10.1016/j.sbspro.2012.11.033.

Malinen, O. P., Savolainen, H., & Xu, J. (2012). Beijing in-service teachers' self-efficacy and attitudes towards inclusive education. Teaching and Teacher Education, 28(4), 526–534. doi:10.1016/j.tate.2011.12.004

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah,NJ: Erlbaum.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38, 30–38.

Multon, K. D., Karen, D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38(1), 30–38.

Mackenzie, C. S., & Peragine, G. (2003). Measuring and enhancing self-efficacy among professional caregivers of individuals with dementia. American Journal of Alzheimer's Disease and Other Dementias, 18(5), 291–299.

MyBrain15. (2008). MyBrain15 programme in Malaysia. Retrieved September 5, 2014, from https://biasiswa.moe.gov.my/MyBrain15/.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38, 30–38.

Morris, D. B., & Usher, E. L. (2011). Developing teaching self-efficacy in research institutions: A study of award-winning professors. Contemporary Educational Psychology, 36(3), 232–245. doi:10.1016/j.cedpsych.2010.10.005.

Nunnally, C. J. (1978). Psychometric theory. Psychometric Theory (2nd ed.). New York NY: McGraw-Hill.

Organisation for Economic Corperation and Development (OECD). (2013). Structural policy and country notes: Malaysia. Retrieved from http://www.oecd.org/dev/asia-pacific/Malaysia.pdf.

Omar, M. W., Mohd Ali, M. N., Khalid, K., Zakaria, A., & Anuar, M. (2013). Managing the time to completion of the PH.D, (October), 262–266.

Prat-Sala, M., & Redford, P. (2012). Writing essays: does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing. Educational Psychology, 32(1), 9–20. doi:10.1080/01443410.2011.621411.

Pennanen, M., Haukkala, A., Vries, H. D., & Vartiainen, E. (2011). Academic achievement and smoking: Is self-efficacy an important factor in understanding social inequalities in Finnish adolescents? Scandinavian Journal of Public Health, 39(7), 714–722.

Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. Research in Mathematics Education, 13(2), 169–190. doi:10.1080/14794802.2011.585828.

Pan, W., Sun, L. Y., & Chow, I. H. S. (2011). The impact of supervisory mentoring on personal learning and career outcomes: The dual moderating effect of self-efficacy. Journal of Vocational Behavior, 78(2), 264–273. doi:10.1016/j.jvb.2010.05.001.

Phillips, J. C., & Russell, R. K. (1994). Research Self-Efficacy, the Research Training Environment, and Research Productivity among Graduate Students in Counseling Psychology. The Counseling Psychologist, 22(4), 628–641. doi:10.1177/0011000094224008.

Pasupathy, R. (2010). A mixed methods study examining that influence faculty members' research self-efficacy and research productivity. Texas Tech University.

Petko, J. (2012). Counselor education doctoral students' level of research self efficacy, interest in research and research mentoring: A cross-sectional investigation (Ph.D. thesis). Central Florida.

Phillips, J. C. (1992). Research self-efficacy and research training environment in counseling psychology. Dissertation Abstracts International, 53(9), 4934–B.

Presely, J. A., & Engelbride, E. (1998). Accounting for faculty productivity in the research university. Review of Higher Education, 22(1), 17–37.

Pellas, N. (2014). The influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: Evidence from the virtual world of Second Life. Computers in Human Behavior, 35, 157–170. doi:10.1016/j.chb.2014.02.048.

Pajares, F. (2006). Self-efficacy during childhood and adolescence: Implications for Teachers and Parents. Information Age Publishing.

Reeb, R. N., Folger, S. F., Langsner, S., Ryan, C., & Crouse, J. (2010). Self-Efficacy in Service-Learning Community Action Research: Theory, Research, and Practice. American Journal of Community Psychology, 46(3), 459–471. doi:10.1007/s10464-010-9342-9.

Rodebaugh, T. L. (2006). Self-efficacy and social behavior. Behaviour Research and Therapy, 44(12), 1831–1838. doi:10.1016/j.brat.2005.11.014.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual Effects on the Self-perceived Efficacy of High School Teachers. Sociology of Education, 65(2), 150–167.

Roest, D. V., & Kleiner, K. (2010). Self efficacy (SE) theories: The biology of confidence. Retrieved November 15, 2015, from http://www.calstate.edu/pa/News/2010/advisory/BOT_11.1.10.shtml.

Rezaei, M., & Miandashti, N. Z. (2013). The Relationship Between Research Self-Efficacy, Research Anxiety and Attitude Toward Research : A Study Of Agricultural Graduate Students. Journal of Educational and Instrutional Studies in the World, (November), 69–78.

Reissman, C. (1993). Teaching research: Beyond the storybook image of positivist science. Journal of Teaching in Social Work, 8, 281–303.

Royse, D., & Rompf, E. (1992). Math anxiety: A comparison of social work and non-social work students. Journal of Education for Social Work, 28, 270–277.

Robb, M. (2012). Self-Efficacy With Application to Nursing Education: A Concept Analysis. Nursing Forum, 47(3), 166–172. doi:10.1111/j.1744-6198.2012.00267.x.

Sproull, N. L. (1995). Handbook of research methods: A guide for practitioners and students in the social sciences. Metuchen, N.J: Scarecrow Press.

Saral, D. G., & Didem, R. (2015). An analysis of educational faculty students' research self-efficacy in terms of a number of variables. Procedia - Social and Behavioral Sciences, 174, 1138–1145. doi:10.1016/j.sbspro.2015.01.729.

Schlosser, L. Z., Knox, S., Moskovitz, A. R., & Hill, C. E. (2003). A qualitative study of the graduate advising relationship: The advisee perspective. Journal of Counseling Psychology, 50, 178–188.

Schunk, D. H., & Rice, J. M. (1993). Strategy fading and progress feedback: Effects of self-efficacy and comprehension among students receiving remedial reading services. Journal of Special Education, 27, 257–276.

Szymanski, D. M., Ozegovic, J. J., Phillips, J. C., & Briggs-Phillips, M. (2007). Fostering scholarly productivity through academic and internship research training environments. Training and Education in Professional Psychology, 1, 135–146.

Shao, J., Chuang, Y., & Chen, S. (2015). A multi-perspective focus-group approach to revise items in a dietary self-efficacy. Australian College of Nursing.

Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed.). Chicago, IL: University of Chicago Press.

Tuckman, B. W. (1972). Conducting educational research. New York NY: Harcourt.

Torres, M. M., Torres, R. R., Rodríguez, A. M. P., & Lee, C. D. (2003). Translation and Validation of the Breastfeeding Self-Efficacy Scale Into Spanish: Data From a Puerto Rican Population. Journal of Human Lactation, 19(1), 35–42.

Unrau, Y. A., & Beck, A. R. (2004). Increasing research self-efficacy among students in professional academic programs. Innovative Higher Education, 28(3), 187–205. doi:10.1023/B:IHIE.0000015107.51904.95.

Vasil, L. (1992). Self-efficacy expectations and causal attributions for achievement among male and female university faculty. Journal of Vocational Behavior, 41(3), 259–269.

Vaccaro, N. (2009). The relationship between research self-efficacy, perceptions of the research training environment and interest in research in counselor education doctoral students: An ex-post-facto, cross-sectional correlational investigation (Ph.D. thesis). Central Florida.

Varney, J. J. (2010). The role of dissertation self-efficacy in increasing dissertation completion: sources, effects and viability of a new self-efficacy construct. College Students Journal, 44(4).

Wang, C., Harris, J., & Patterson, P. (2013). The Roles of Habit, Self-Efficacy and Satisfaction in Driving Continued Use of Self-Service Technologies : A Longitudinal Study. Journal of Service Research, 16(3), 400–414. doi:10.1177/1094670512473200.

Welzer-Ward, L., Baltes, B., Hoffman-Kipp, P., & Lynn, L. (2010). Students' Research Self-Efficacy (RSE) During Online Doctoral Research Courses. Retrieved May 6, 2014, from http://jamiesmithportfolio.com/EDTE800/wp-content/Self-Efficacy/Baltes.pdf.

Wainstock, S. (1994). Swimming against the current: Teaching research methodology to reluctant social work student. Journal in Teaching Social Work, 9, 3–16.

Zimmerman, W. S. (1995). Self-efficacy and educational development. In A. Bandura

(Ed.), Self-efficacy in changing society (pp. 202–231). New York NY: Cambridge University Press.

Zhao, J., McCormick, J., & Hoekman, K. (2008). Idiocentrism-allocentrism and academics' self-efficacy for research in Beijing universities. International Journal of Educational Management, 22(2), 168–183. doi:10.1108/09513540810853567.

Zubaran, C., Foresti, K., Schumacher, M., Thorell, M. R., Amoretti, A., Müller, L., & Dennis, C. L. (2010). The Portuguese version of the Breastfeeding Self-Efficacy Scale-Short Form. Journal of Human Lactation:Official Journal of International Lactation Consultant Association, 26(3), 297–303. doi:10.1177/0890334409359916.

Yazit, N., & Zainab, A. N. (2007). Publication Productivity of Malaysian Authors and Institutions in LIS. *Malaysian Journal of Library & Information Science*, *12*(2), 35–55.

Eurasian Journal of Educational Research 82 (2019) 169-190

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

Direct and Indirect Effects of Learning Strategies and Reading Enjoyment on PISA 2009 Reading Performance

Ezel TAVSANCIL[1] , Ozen YILDIRIM [2] , Safiye BILICAN DEMIR[3]

ARTICLE INFO

Article History:

Received: 4 Mar. 2019

Received in revised form: 17 Jun. 2019

Accepted: 8 Jul. 2019

DOI: 10.14689/ejer.2019.82.9

Keywords

Reading performance, learning strategies, reading enjoyment, direct-indirect effect, Hierarchical linear modeling (HLM)

ABSTRACT

Purpose: The aim of the study was to investigate the extent students' learning strategies and reading enjoyment variables predicted PISA 2009 reading achievement, and to examine whether reading enjoyment had a mediator effect in the relationship between reading achievement and learning strategies.

Methods: In this correlational study, Turkey PISA 2009 reading comprehension cognitive test and student questionnaires were used for data collection. Learning strategies and reading pleasure which was related to students' reading comprehension achievement were identified as predictive variables and index values of these variables were used. The data were analysed using hierarchical linear models (HLM).

Findings: The results of the HLM analysis showed that the students who used control strategies more frequently had higher reading scores. Memorization was negatively associated with reading literacy and the elaboration strategies had no effect on achievement. Reading enjoyment was identified as a variable that improved students' reading performance. It was found out that control and elaboration strategies directly affected student's reading enjoyment whereas memorisation did not have such a direct effect. The results related to the indirect effect between variables showed that reading enjoyment had a partial mediator effect for the control strategy and had a full mediating effect for elaboration.

Implications for Research and Practice: Teachers are encouraged to use deep learning strategies instead of surface learning strategies. Teachers can inform their students about what deep learning strategies are, why they are more effective than others, and how to use them. The think aloud technique can be used to show how this strategy can be used in daily life.

[1] Ankara University, Faculty of Educational Science, Ankara, Turkey, etavsancil@gmail.com, ORCID ID: https://orcid.org/0000-0002-8318-2043

[2] Pamukkale University, Faculty of Education, Denizli, Turkey, ozenyildirim@pau.edu.tr, ORDIC ID: https://orcid.org/0000-0003-2098-285X

[3] Corresponding Author, Kocaeli University, Faculty of Education, Kocaeli, Turkey, safiye.demir@kocaeli.edu.tr, ORCID ID: https://orcid.org/0000-0001-9564-9029

Introduction

Reading is one of the best strategies for acquiring knowledge and learning. The importance of reading for the individual's personal development, daily and work life cannot be ignored (Holden, 2004). Reading allows us to learn about the events that happen around us or we are curious about them, and to learn about the history, social studies, language arts, science, mathematics and subjects needed for school learnings (Lyon, 1997). Reading is a dynamic and complex skill that is influenced by the reader's personal characteristics, learning processes, and prior knowledge (Koda, 2005; Wilkinson & Son, 2011).

The main purpose of reading is developing comprehension skills. Individuals need to use more complex thinking processes such as establishing relationships, making comparisons, presenting evidence, making generalizations and inferences or predicting events. The fact that most of the learning tools used in schools are reading-based resources shows the importance of reading comprehension skills for academic achievement. Research has shown that students who are good readers are successful in school life (e.g. Duncan et al., 2007; Guthrie, Schafer & Huang 2001; Juel, 1988).

Many countries acknowledge the importance of raising individuals who understand what they are reading, and they conduct research to identify to what extent their students are successful in reading by using different evaluation methods and the factors affecting this success. In recent years, several international studies have been conducted on the education policies and comparison of them with other countries. PISA is one of the international projects that evaluate students' knowledge and skills in the fields of mathematics, science, and reading. This research evaluates to what extent students have the basic knowledge and skills required in modern society (OECD, 2010).

One of the most intensively studied aspects of reading in the last few decades concerns learning strategies and methods (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Learning strategies can be defined as techniques that support student's learning and related behaviors, and thinking process of students (Oxford, 2003). Students need to develop a series of strategies to understand the gist of any given text, to make a judgment and make personal connections with text, and to make sense of unknown words in the text in order to be a good reader (Antonacci, 2000). Today, even young learners are expected to comprehend complex texts and answer complicated questions that require independent interpretation and integration of numerous knowledge sources (Ortlieb, 2013). Therefore, improving reading skills effectively is an important research question than ever. Lack of adequate reading skills impedes educational progress (OECD, 2010). Learning strategies can play a crucial role in closing these "reading gaps". Research shows that lower-achieving students also have lower awareness about how to read and learn efficiently (Alderson, 2000; Baker & Brown, 1984; Pitts, 1983).

Classification of strategies is also somewhat complex and unclear (Kang, 1997; Oxford, 1990), one of the most frequent taxonomies categorizes them as surface, deep and meta strategies (Chiu, Chow & McBride-Chang, 2007). Surface strategies are often

associated with rote learning and memorising of material by using crude techniques, such as rehearsal. Deep strategies are related to more refined processing of knowledge, e.g., transferring information previously learned to new knowledge areas, whereas meta-strategies are described as processes that regulate actions (Hacker, 1998). Another classification separates strategies as cognitive and metacognitive (Phakiti, 2006). This study focuses on metacognitive strategies. PISA 2009 datasets reflect the use of the tree metacognitive strategies: memorisation, elaboration and control.

Memorization is defined as recalling texts through repetitive reading. This is an ineffective strategy, leading to lower re-acquisition and transient learning impacts (Mayer, 2008; Pressley & Harris, 2017). Elaboration is another strategy that helps to link old and new knowledge or enables to make a link between knowledge to everyday experience (OECD, 2010). It leads to a deeper understanding of concepts, more efficient application of knowledge in real life, and makes it easier to retrieve the knowledge from memory (Chiu et al., 2007). Lastly, the control strategy is seen as the most characteristic examples of metacognitive strategies (Kraayenoord, 2010) and it monitors reading purpose, planning and literate activities (Iwai, 2011; Phakiti, 2006).

Using of appropriate strategies are important as they are associated with greater reading enjoyment and better performance (Carretti, Caldarola, Tencati & Cornoldi, 2014; McDaniel, Howard & Einstein, 2009; OECD, 2010; Yip, 2007). Oxford (2003) states that students use learning strategies to make learning more fun and to be more successful by self-direction. When the student uses the appropriate strategy, the learning process becomes more enjoyable. Therefore, the mediating role of reading enjoyment in the relationship between learning strategies and reading achievement is also taken into consideration in this research.

Successful reading comprehension is also related to reading enjoyment (Zasacka & Bulkowski, 2017). Reading enjoyment refers to the satisfaction we receive from reading in our daily lives. It can also start with the motivation of someone else and then the individual becomes interested in reading. Many studies and political activities aiming at increasing student's reading success have focused on the cognitive aspects of reading such as word recognition and comprehension. However, the enjoyment of reading is more important for student achievement than the socio-economic level of their families (OECD, 2002). Reading enjoyment could, therefore, be an important way to reduce social exclusion and raise educational standards. Research-based on a reciprocal causality model shows that school success is at risk for individuals reading without having fun (Mol & Bus, 2011; Zasacka & Bulkowski, 2017). Secondary school students who prefer reading in their spare time are more successful than those who do not (Mol & Jolles, 2014; OECD, 2010). A recent meta-analysis by Mol and Bus (2011) showed that leisure readers in College and University had higher GPAs than their non-leisure reading peers.

It is important to raise individuals who understand what they read in today's rapid development and competitive world. According to PISA 2009 results, 56.7% of the age group of 15 students in Turkey have either Level 2 (basic proficiency) or below the level of proficiency. It shows that there are no students with level 6 who have high-

level reading skills (EARGED, 2010). In PISA 2009, Level 2 has been accepted as the basic level of competence in which students have started to show their skills in reading that will enable them to be a success in social life. These results make it necessary to identify the causes of difficulties experienced by Turkish students in terms of reading skills and the variables that play a role in this situation. Thus, it will inform teachers and families on how to develop these skills. In addition, research findings will provide scientific data to relevant institutions in the development of appropriate educational settings and the creation of educational programs to ensure the development of these skills.

The aim of the study is to understand the extent students' learning strategies and reading enjoyment variables predict PISA 2009 reading achievement, and to examine whether reading enjoyment has a mediator effect in the relationship between reading achievement and learning strategies.

Method

Research Design

This research was a correlational study that aimed to reveal the relationships between student's reading achievement and learning strategies variables.

Research Sample

Data in PISA applications were collected by random stratified sampling to represent the 15-age group of each country. The sample was selected from different units such as students, teachers, and schools. Turkey PISA 2009 sample consisted of 4996 students from 170 schools in the age group of 15. After missing value analysis, HLM analyses were conducted on 4648 students in 169 schools. 51.64% of the participants were female and 48.36% were male. PISA data set has predicted weighting value for student and school samples. Sample weights should be included in the analysis to make accurate predictions about the PISA population using the sample (OECD, 2012). Weightings for student and school sampling were used in the study.

Research Instruments and Procedures

In this study, Turkey PISA 2009 reading comprehension cognitive test and the data collected from student questionnaires were used. Data were downloaded from the official web page of the OECD (http://pisa2009.acer.edu.au/downloads.php). Learning strategies and reading pleasure which is related to students' reading comprehension achievement were identified as predictive variables, and index values of these variables were used. Scale scores for these variables were estimates of latent traits using by Item Response Theory (IRT) scaling of depended on item types. Brief descriptions of these indices from PISA technical report are given below (OECD, 2012):

Reading enjoyment (ENJREAD) *(Mediator):* Eleven items (e.g."*I read only if I have to, reading is one of my favourite hobbies, I find it hard to finish books*") were used to measure this variable. The scale includes four response categories from "strongly disagree",

"disagree", "agree" to "strongly agree". The positive scores from this scale indicate higher levels of enjoyment of reading. The alpha reliability of scale for this research was reported as 0.84 in PISA Technical Report.

Learning strategy (Initial variables): This scale includes three subscales: memorisation (MEMSTR), elaboration (ELBSTR) and control strategies (CTRSTR). Positive scores from the scale indicate the use of reading strategies more often. In the scale, thirteen items measured learning strategies; four items measured memorisation (e.g. "*When I study, I try to memorize as many details as possible; when I study, I read the text so many times that I can recite it. etc.*"), four items measured elaboration strategies (e.g *"When I study, I try to relate new information to prior knowledge acquired in other subjects, When I study, I try to understand the material better by relating it to my own experience etc"*), and five items measuredcontrol strategies (*"When I study, I check if I understand what I have read; When I study, I start by figuring out what exactly I need to learn, etc").* There are four response categories varying from "almost never", "sometimes", "often" to "almost always". Positive scores indicate greater use of that learning strategy. The alpha reliability of memorisation is 0.67, elaboration was 0.68 and the control was 0.74.

Reading performance (PV1- PV5) (outcome): Students participating in PISA applications have not responded to all questions of the reading literacy test. PISA uses the imputation methodology usually referred to as plausible values (PVs). Using item parameters from the international calibration, the plausible values are randomly estimated from the marginal posterior of the latent distribution for each student (OECD, 2012). In this research, five possible values for reading literacy were separately analysed. Table 1 presents the descriptive statistics of the variables.

Table 1

Descriptive Statistics

Variable name	N	MEAN	SD	MIN	MAX
CTRSTR	4648	0.24	0.82	-2.05	2.50
ELBSTR	4648	0.46	0.78	-1.52	2.76
MEMSTR	4648	-0.04	0.77	-1.97	2.52
ENJREAD	4648	0.62	0.76	-1.82	2.66
PV1 (Reading)	4648	468.29	77.71	229.01	706.04
PV2	4648	467.04	78.19	234.06	726.06
PV3	4648	467.61	78.44	225.00	698.26
PV4	4648	467.23	77.96	229.49	694.46
PV5	4648	467.36	77.95	234.70	698.26

There are large differences in standard deviation between students' reading comprehension scores and differences in the assessment of the index values of the variables. When Table 1 is examined, it can be seen that, for example, PV1 had SD = 78.12 while CTRLSTR had SD = 0.82. Kline (2011) emphasizes that the variance of the variables should first be analysed and then the variance values should be scaled by

various conversion methods, including iterative processes of analysis. Standardized scores of dependent variables (mean 0.00, standard deviation 1.00) were calculated to equalize the scale levels of dependent and independent variables.

Data Analysis

In PISA, sampling was selected from different units such as students, teachers, and schools. In studies involving such sampling structures, it is assumed that the previous level is not independent of its subsequent levels (Raudenbush & Bryk, 2002). Multi-level models take into account dependence between observations in nested data, and so the result was estimated more accure (Krull & MacKinnon, 2001). This results in fewer errors. In this research, direct effects of memorisation, elaboration and control strategies and reading enjoyment variables on reading comprehension were tested. In conditions when reading enjoyment was mediator, the direct effects of memorisation, elaboration and control strategies on reading comprehension were analysed using hierarchical linear models (Figure1).

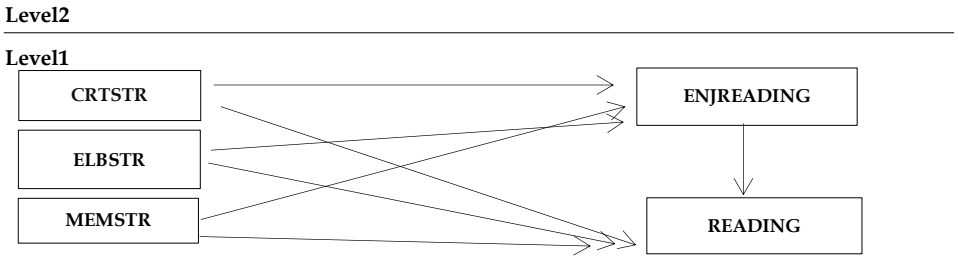


Figure 1. *The Tested Model*

In this research, a mediational model was formed at the lower level by considering the hierarchical structure of the PISA sample. All variables were measured at Level-1, the model was labelled as 1-1-1, the Level-1 units (students) were nested in Level-2 units (schools). The direct and indirect outcomes of Initial variables (CRTSRT, ELBSTR, and MEMSTR) on reading achievement were analysed. The effect of reading enjoyment (mediator variable, M) on reading literacy (outcome variable, Y) was also been tested. Considering the HLM model tested in the study, the following regression equations were formed. (Krull & Mckinnon, 2001; Yildirim, 2012; Zhang, Zyphur & Preacher, 2009):

Model 1-1-1

(1) L-1: $PVReading_{(ij)}\ (Y) = \beta_{0j} + \beta_{1aj}(CRTSTR)_{ij} + \beta_{1bj}(ELBSTR)_{ij} + \beta_{1cj}(MEMSTR)_{ij} + r_{ij}$

L-2: $\beta_{0j}=\gamma_{00}+u_{0j}$

$\beta_{1aj}=\gamma_{10}$

$\beta_{1bj}=\gamma_{20}$

$\beta_{1cj}=\gamma_{30}$

(2) L-1: $ENYREADING_{(ij)}(M)= \beta_{0j}+ \beta_{1dj}(CRTSTR)_{ij}+ \beta_{1ej}(ELBSTR)_{ij}+ \beta_{1fj}(MEMSTR)_{ij}+ r_{ij}$

L-2: $\beta_{0j}=\gamma 00 +u0j$

$\beta_{1dj}=\gamma 10$

$\beta_{1ej}=\gamma 20$

$\beta_{1fj}=\gamma 30$

(3) L-1: $PVReading_{(ij)} = \beta_{0j} + \beta_{1gj}(CRTSTR)_{ij} + \beta_{1hj}(ELBSTR)_{ij} + \beta_{1kj}(MEMSTR)_{ij} + \beta_{1mj}(ENJREADING)_{ij} + r_{ij}$

L-2: $\beta_{0j}=\gamma_{00}+u_{0j}$

$\beta_{1gj}=\gamma_{10}$

$\beta_{1hj}=\gamma_{20}$

$\beta_{1kj}=\gamma_{30}$

$\beta_{1mj}=\gamma_{40}$

Model refers to the model where all variables measured at level 1. Equations presume that predictive variables are fixed; that is, they do not allow to vary across higher-level units. The effect of the mediator was determined by multiplying the alpha levels of the variables in the second module and the alpha level of reading enjoyment variable in the fourth model ($\gamma_{10}(2) * \gamma_{40}(3)$). Sobel (1982) test was used to calculate the significance of the effect. The variances explained in the reading scores of the variables included and the effect sizes based on these variances were calculated. The following formula is used in the calculation of the explained variances

(Snijders & Bosker; 2012):

$R^2=1- (\sigma^2_F+\tau_F) / (\sigma^2_E+\tau_E)$

Where σ^2_F represents the variance of eij for Model with coefficients; τ_F represents the variance of u0j for the same model. σ^2_E represents the level-one random error variance for the empty model, and τ_E represents the level-two random error variance for the null model.

The effect size f^2 was computed as (Cohen, 1992):

$f^2 = R^2 / (1-R^2)$

The explanation of f^2 indicates that "*0.02 is a small effect, 0.15 is a medium effect, and 0.35 is a large effect*" (Cohen, 1992).

Before the analysis of the data, the assumptions with regards to the regression analysis and hierarchical models were tested. Multilevel modeling was conducted using Hierarchical Linear Modelling (HLM6) (Raudenbush, Bryk, Cheong & Congdon, 2004).

Results

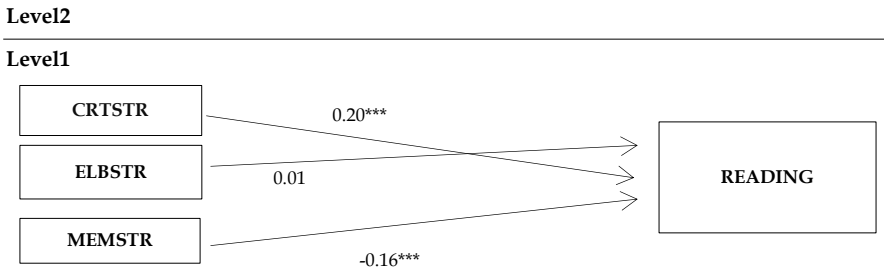
Null model (One-way ANOVA) and ICC (intra-class correlation) were calculated before testing the direct effects of learning strategies on reading performance. This shows that 68% of the variance in reading performance was due to the difference between the schools and 32% of it was related to the differences among students. The overall effects of the learning strategies on reading performance were tested by Model 1 before including reading enjoyment to the model as a mediator. Table 2 presents the findings of the model and the tested model.

Table 2

The Overall Effect on Reading Attainment

Model1	β Coefficients	Se	t	p
CNTRSTR (Xaj) → Reading(Y)	0.202	0.029	7.011	0.000
EBRSTR (Xbj) → Reading(Y)	0.016	0.024	0.665	0.506
MEMSTR (Xcj) → Reading(Y)	-0.160	0.023	-6.763	0.000
Random Effects	Variance Component	df	χ2	p
U0j	*0.89*	*168*	10100.05	0.000
rij	*0.434*			

(Y)=Outcome, (X)= Initials, (M)= Mediator



***p<0.001

Figure 2. *The relationship between learning strategies and reading performance*

Table 2 shows that the control and memorisation strategy were significant predictors of reading performance ($p < 0.01$), and the elaboration did not significantly predict the students' reading performance ($p > 0.01$). When other variables were controlled, the control strategy resulted in a standard deviation increase of 0.20 in the frequency use of the control strategy. Memorisation is a variable that decreases the student's reading scores. A one-unit increase in the standard deviation of memorisation strategy causes 0.16 decrease in the standard deviations of reading scores. The strategies explained 9% (R^2) of variance on reading performance. The variables in the model could explain 9% of the change in reading attainment. Accordingly, the effect size of reading variables on reading performance was 0.10 in (f^2). The effect of variables on reading attainment was minimum. Model 2 tests whether learning strategies predict reading enjoyment, and the findings are presented in Table 3.

Table 3

The Effects on Reading Enjoyment

Model 2	β Coefficients	Se	t	p
CNTRSTR (Xdj) → ENJREADING(M)	0.271	0.024	10.930	0.000
EBRSTR(Xej) → ENJREADING(M)	0.117	0.018	6.424	0.000
MEMSTR (Xfj) → ENJREADING(M)	-0.016	0.025	-0.640	0.522
Random Effects	**Variance Component**	**df**	**χ2**	**p**
u0j	0.014	168	416.518	0.000
rij	0.482			

(Y)=Outcome, (X)= Initials, (M)= Mediator

Table 3 shows that that control and elaboration strategies significantly predicted reading enjoyment (p <0.01); however, the memorisation strategy was not a significant predictor (p> 0.01). Therefore, memorisation had no indirect effect on reading attainment. These variables explained 13% of the variance in the variable of reading enjoyment. As the frequency of using control and elaboration strategy increased, the level of reading enjoyment also increased.

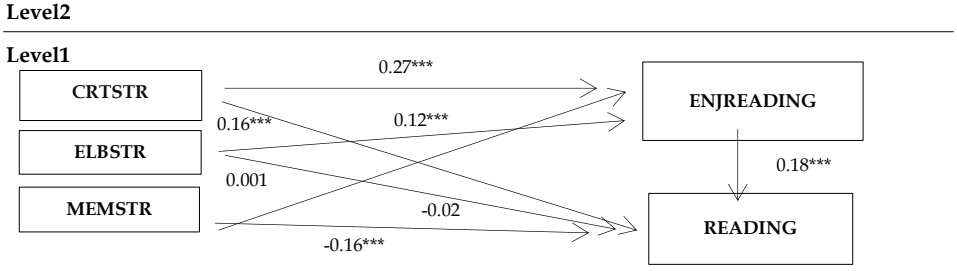
The mediator effects of reading enjoyment were tested via Model 3. Table 4 presents the findings of the model, and the tested model is shown in Figure 3.

Table 4

The Direct Effects on Reading attainment

Model3	β Coefficients	Se	t	p
CNTRSTR (Xaj) → Reading(Y)	0.155	0.030	5.069	0.000
EBRSTR (Xbj) → Reading(Y)	-0.005	0.025	-0.020	0.839
MEMSTR (Xcj) → Reading(Y)	-0.158	0.023	-6.823	0.000
ENJREADING(M) → Reading(Y)	0.176	0.023	7.466	0.000
Random Effects	Variance component	df	χ^2	p
u0j	0.886	168	10347.168	0.000
rij	0.420			

(Y)=Outcome, (X)= Initials, (M)= Mediator



***p<0.01

Figure 3. *The relationship between learning strategies, reading enjoyment and reading attainment*

Table 4 shows that control and memorisation strategy predicted reading scores (p<0.01), and elaboration strategy was not a significant predictor in this model (p>0.01). According to the model, reading scores of the students who used control strategies more frequently were higher, while reading scores of the students who used

the memorisation strategies were low. The more pleasure student gets from reading, the higher their reading scores are. When other variables were controlled, a standard deviation increase in the reading enjoyment increased the standard deviation of the students' reading points by 0.18 units. When reading enjoyment mediator was included in the model, all variables explained 10% (R^2) of the reading scores, and the effect size was calculated as 0.11 (f^2). The variance explained by the variables was small.

The effects of control and elaboration strategies on reading scores were examined with the effect of reading enjoyment mediator. It was found out that the use of control and elaboration strategies' frequency significantly predicted students' reading performance ($p < 0.01$). This situation showed that reading enjoyment had a mediating effect on predicting reading scores of control and elaboration strategies. As the frequency of using a control strategy for reading increased, level of reading enjoyment increased. Student's reading score increased as the level of enjoyment increased (Mediation effect (10 (2) * (40 (3) = 0.27 * 0.18 = 0.05, $z_{sobel} = 6.33$ $p < 0.01$). A one-unit increase in index values related to the use of control strategy led to a 0.27 increase in reading enjoyment scores, and 0.18 of this increase was transferred to reading scores. Thus, the indirect effect of the frequency of the use of control strategy on reading scores was 0.05.

It is observed that the students who use the elaboration strategy frequently increase their level of reading enjoyment and reading scores. Since there was no significant relationship between the elaboration strategy and reading performance, the reading enjoyment was a full mediator for the elaboration strategy. (Mediation effect $\gamma_{20}(2)$* $\gamma_{40}(3)=0.12*0.18=0.02$, $z_{sobel}=5.03$ $p<0.01$). A one-unit increase in the elaboration strategy led to an increase of 0.12 in reading enjoyment, and 0.18 of this increase was transferred to reading scores. Thus, the indirect effect of the elaboration strategy on reading scores was 0.02.

Discussion, Conclusion, and Recommendations

This research analysed the extent the effects of the learning strategies and reading enjoyment variables predict students' PISA 2009 reading performance, and tested the mediator effect of the reading enjoyment variable on reading the relationship between reading performance and learning strategies. The results of the study showed that students who used control strategies more frequently had higher reading scores. PISA reading skills are measured in the form of simple analysis of a text or deriving extensive, realistic or figurative meanings from the text, and understanding the theme of a long text written for discussion or storytelling (OECD, 2012). Therefore, students' use of higher-level learning strategies to understand relevant texts is a factor that increases their reading scores. Many studies conducted in different countries using PISA 2000 data found positive correlations between the frequency of using control strategies of students and reading performance (Artelt, Baumert, Julius-McElvany &Peschar, 2003; Bortoli & Cresswell, 2004; Muszyński & Jakubowski, 2015).

Memorization was negatively associated with reading performance. Yet, this is not surprising; researchers have argued that a surface strategy such as memorization is insufficient for deep learning. (e.g. Chiu et al.,2007; Czuchry & Dansereau, 1998; Mayer, 2008). However, the relationship between these two variables for different countries where PISA data were used showed the difference. Using of memorisation shows a less consistent relationship with performance than the other two strategies (control and elaboration). For example, in Hong Kong, Hungary, and the Russian Federation, it is found that students who frequently use the memorization strategy are more successful than those who used it less. In Italy, Austria, and Poland, the success of students who use this strategy frequently is low (Artelt et al., 2003; Bortoli & Cresswell, 2004; Li & Chu, 2012; Muszyński & Jakubowski, 2015). These findings do not, however, support that memorisation strategies never contribute to effective learning. Researchers point out that poor readers memorise more to make up for other learning limitations (Artelt et al., 2003; Weinstein & Mayer, 1986). Several other factors may support the conflict between results. First, socio-cultural factors can play a role in the emergence of the results. Cultural context is a factor that affects students' learning strategies and approaches to that strategy (Oxford, 1996). For example, memorization is common in Asian countries, and students from the Confucian heritage culture generally prefer memorization strategies (Kember, Biggs & Leung, 2004; McInerney, 2011; Politzer & McGroarty, 1985). Second, the relationship between using strategies and performance may be moderated by other student or family characteristics. Third, it should be noted that high frequency use, and more types of learning strategies do not necessarily lead to better academic success (Wen & Wang, 2004). Finally, although the same PISA data is used in these studies, the variables discussed for each study are different. Accordingly, the effects of the learning strategies of students on reading literacy differ depending on the direct and indirect effects of these variables and their interaction with each other.

The elaboration strategies had no effect on achievement, which could be considered as an unexpected but not a surprising finding. The research of Muszyński and Jakubowski (2015) and Chiu et al., (2007) also reached a similar finding. There may be some reasons for this result. First, questionnaire questions may be insufficient to adequately measure student's transfer skills through detailing. For example, students may not have an objective view of the extent to which and how to link their previous knowledge to their new knowledge. How to transfer knowledge across situations or context remains one of the greatest challenges for educators (Halpern, 1998). Second, it is possible that the use of the elaboration strategy should not be measured according to self-reports as students report that they struggle to use this strategy. It is difficult to successfully implement; thus, they would mostly then report failed attempts (Muszyński & Jakubowski, 2015). Despite these explanations, there are also studies showing that there is a positive significant relationship between elaboration strategies and reading literacy. For example, Artelt et al. (2003) conducted a study using PISA 2000 data from 26 countries. They assert that stronger readers are more likely to use elaboration strategies than poor readers. Similarly, Li and Chun (2012) state that there is a strong positive relationship between elaboration strategies and reading literacy for Hong Kong in the top rankings of PISA. This result can be explained by the theory of

learning pattern dissonance which asserts that high-achieving students tend to benefit from using deep cognitive strategies and self-regulated learning whereas low-achieving students succeed by virtue of externally-regulated learning (e.g., by teacher or parents) and using surface strategies (Beishuizen & Stoutjesdijk 1999; Meyer, 2000).

Reading enjoyment is identified as a variable that improves students' reading performance. Research findings show that a large achievement gap between the secondary school students who read and who do not read books is the most important finding (e.g. OECD, 2010; Mol & Bus, 2011; Mol & Jolles, 2014; Rao & Moely, 2000). Readers who read frequently are more likely to enjoy reading, and it improves their performance in school (Baker, Dreher & Guthrie, 2000; Oatley, 2012). This finding is pointing a causal relationship between the variables.

When the effects of learning strategies in the related model on the reading enjoyment were examined, it was found out that control and elaboration strategies directly affected student's reading enjoyment whereas memorisation did not have such a direct effect. This shows that students who use higher-level strategies for learning enjoy reading more. When reading the text, students using higher-level skills take more pleasure in reading. Previous studies with adults found that feeling highly vividness can increase the enjoyment of reading books (e.g., Green et al., 2008; Weibel, Wissmath & Mast, 2011).

When the results showing the indirect effect between variables were examined, reading enjoyment had a partial mediator effect for the control strategy and had a full mediating effect for elaboration. This shows that students who employ control and elaboration strategies enjoy reading more, and they are more successful. Research on the effects of learning strategies shows that the use of learning strategies has positive effects on students' attitudes (Carroll & Leander, 2001; Huffman & Spiers, 1992; Keller, 1990) and their academic achievement (e.g. Ho, 1998; Vermunt & Vermetten, 2004; Ward & Rosetta 2001). Successful students were found to be more enthusiastic and conscious about the use of learning strategy (Loranger, 1994; Schutz, Drogozs, White & Distefeno 1998).

In conclusion, this research shows that students who use learning strategies with high-level skills enjoy reading, and they are more successful in reading comprehension. Teachers are encouraged to use deep learning strategies instead of surface learning strategies. Teachers can inform their students about what deep learning strategies are, why they are more effective than others, and how to use them. The think-aloud technique based on how this strategy can be used in daily life can be used. Future studies may examine cognitive or affective factors in natural settings or explore the effects of learning strategies on reading literacy experimentally. Large scale studies that distinguish between higher and lower-achieving countries are definitely needed in order to assess the contribution of the learning strategies to reading literacy.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Antonacci, P. A. (2000). Reading in the zone of proaximal development: Mediating literacy development in beginner readers though guided reading. *Reading Horizons*, 41(1), 19-33. Retrieved February 24, 2019, from https://scholarworks.wmich.edu/cgi/viewcontent.cgi?referer=https://scholar.google.com.tr/&httpsredir=1&article=1194&context=reading_horizons

Artelt, C., Baumert, J.,Julius-McElvany, N., & Peschar, J. (2003). Learners for Life. Student Approaches to Learning. Results from PISA 2000. Retrieved April 14, 2019, from https://files.eric.ed.gov/fulltext/ED480899.pdf

Baker, L. & Brown, A. (1984). Meta cognitive skills and reading. In P. D. Pearson (Ed.), *Handbook of reading research.* New York: Longman.

Baker, L., Dreher, M.J. & Guthrie, J.T. (2000). *Engaging young readers: Promoting achievement and motivation*. New York: Guilford Publishing.

Beishuizen, J. J., & E. T. Stoutjesdijk, E. T. (1999). Study strategies in a computer assisted study environment. *Learning and Instruction*, 9(3), 281–301. Doi: 10.1016/S0959-4752(98)00027-9

Bortoli, L. D., & Cresswell, J. (2004). Australia's indigenous students in PISA 2000: Results from an international study. ACER Research Monograph No 59. Retrieved March 10, 2019 from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=indigenou _education

Carretti, B., Caldarola, N., Tencati, Ch., & Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: the effect of two training programmes focusing on metacognition and working memory. *The British Psychological Society*, *84*(Pt. 2), 194–210. Doi: 10.1111/bjep.12022

Carroll, L., & Leander, S. (2001). *Improving student motivation through theuse of active learning strategies*.Masters of arts action research Project, Saint Xavier University and Skylights Field-Based Masters Program, Eric Identifier:ED455961.

Chiu, M. M., Chow, B. W-Y., & McBride-Chang, C. (2007). Universals and specifics in learning strategies: Explaining adolescent mathematics, science, and reading achievement across 34 countries. *Learning and Individual Differences*, *17*(4), 344-365. Doi:10.1016/j.lindif.2007.03.007

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. Doi: 10.1037/0033 2909.112.1.155

Czuchry, M., & Dansereau, D. F. (1998). The generation and recall of personally relevant information. *Journal of Experimental Education*, *66*, 293–315. Doi: 10.1080/00220979809601403

Duncan G.J., Dowsett C.J., Claessens, A, Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science*, *14*(1), 4–58. Doi:10.1177/1529100612453266

EARGED (2010). Uluslararası Öğrenci Değerlendirme Programı PISA 2009 Ulusal Ön Raporu [PISA 2009 National Preliminary Report]. Retrieved May 20, 2019 from http://pisa.meb.gov.tr/wp-content/uploads/2013/07/PISA-2009-Ulusal-On-Rapor.pdf

Guthrie, J. T, Schafer, W D., & Huang, C. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research*, 94, 145-162. Doi: 10.1080/00220670109599912

Green,M.C.,Kass,S.,Carrey,J.,Herzig,B.,Feeney,R., & Sabini,S.(2008). Transportation across media: Repeated exposure to print and film. *MediaPsychol*. 11, 512- 539. Doi:10.1080/15213260802492000

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky and A. C. Graesser (eds.), *Metacognition in educational theory and practice* (pp. 1-20). Mahwah: Erlbaum.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449 455. Doi:10.1037/0003-066X.53.4.449

Ho, I. (1998). Relationships between motivation\attitude, effort, English proficiency, and social cultural educational factors and Taiwan and technological university \institute students' English learning strategy use, Unpublished doctoral dissertation, Auburn University, Alabama.

Holden, J. (2004). *Creative Reading*. London: Demos.

Huffman, L. E., & Spires, H. A. (1992). Effects of explicit instruction in notetaking on sixth graders' lecture comprehension and attitudes toward note taking, Paper presented at the Annual Meeting of the North Carolina Association for Research in Education, ChapelHill, NC.

Iwai, Y. (2011). The effects of metacognitive reading strategies: pedagogical implications for EFL/ESL teachers. *The Reading Matrix*, *11*(2), 150–159. Retrieved January 15, 2019 from https://pdfs.semanticscholar.org/85f6/c512258e86ee797f2d0973d8b123d4a1d9b4.df

Juel, C. (1988).Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437-447. Doi: 10.1037/0022 0663.80.4.437.

Kang, D. H. (1997). *Assessing Korean middle school students' language learning strategies in input-poor environments*. (ERIC Document Reproduction Service No. ED 413 778). Retrieved January 10, 2019, from http:// files.eric.ed.gov/fulltext/ED413778.pdf

Keller, J. J. (1990). Strategy Games: Developing positive attitudes and perseverance toward problem solving with fourth graders (ERIC Document Reproduction Service No. ED323013).

Kember, D., Biggs, J., & Leung, D. Y. P. (2004). Examining the multidimensionality of approaches to learning through the development of a revised version of the Learning Process Questionnaire. *British Journal of Educational Psychology*, *74*, 261–280. Doi: 10.1348/000709904773839879

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. 3rd Edition, New York City, The Guilford Press.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* Cambridge: Cambridge University Press.

Kraayenoord, C. E. van (2010). The role of metacognition in reading comprehension. In H. P. Trolldenier, W. Lenhard and P. Marx (eds.), *Brennpunkte der Gedächntisforschung* (pp. 277- 304). Gottingen: Hogrefe.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249-277. Doi: 10.1207/S15327906MBR3602_06

Li, J., & Chun, C. K. W. (2012). Effects of learning strategies on student reading literacy performance. *The Reading Matrix*, *12*(1), 30–37. Retrieved March 10, 2019, from https://pdfs.semanticscholar.org/d39f/9864620a80030cf1d19f4dd912b5318 bb404. df

Loranger, A. L. (1994). The study strategies of successful and unsuccessful high school students. *Journal of Reading Behavior,* 26 (4), 347-360. Doi: https://doi.org/10.1080/10862969409547858

Lyon, R. (1997). Statement before the Committee on Education and Workforce. U.S. House of Representatives. Retrieved January 9, 2019, from http://mirror.apa.org/ppo-OLD/lyon.html.

Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River: Pearson Merrill Prentice Hall.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy. *Psychological Science*, *20*(1), 516–522. Doi: 10.1111/j.1467-9280.2009.02325.x

McInerney, D. M. (2011). Culture and self-regulation in educational contexts. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 442- 464). New York, NY: Routledge.

Meyer, J. H. F. (2000). The modeling of "dissonant"study orchestration in higher education. *European Journal of Psychology of Education*, *15*(1), 5–18. Doi:10.1007/BF03173163

Mol,S.E., & Bus, A. G. (2011).To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychol.Bull.* 137, 267–296. Doi: 10.1037/a0021890

Mol, S. E., & Jolles, J. (2014). Reading enjoyment amongst non-leisure readers can affect achievement in secondary school. *Front.Psychol*, 5, 1214. Doi: 10.3389/fpsyg.2014.01214

Muszynski, M., & Jakubowski, M. (2015). Learning strategies and reading performance: PISA 2009 results for Poland. Edukacja, 3(134), 5-25.

OECD (2002). *Reading for change: Performance and engagement across countries. Results from PISA 2000*. New York: Organisation for Economic Cooperation and Development.

OECD (2010). *PISA 2009 results. Learning to learn: student engagement, strategies and practices.* Paris: OECD Publishing.

OECD (2012). *PISA 2009 technical report.* Paris, France: OECD Publications.

Oatley, K.(2012).The cognitive science of fiction. *Wiley Interdiscip.Rev.Cogn.Sci.* 3, 425-430. Doi:10.1002/wcs.1185

Ortlieb, E. (2013). Using anticipatory reading guides to improve elementary students'comprehension. *International Journal of Instruction*, *6*(2), 145–162. Retrieved January 10, 2019, from https://files.eric.ed.gov/fulltext/EJ1085407.pdf

Oxford, R. L. (1990). Styles, strategies, and aptitude: connections for language learning. In T. S. Parry and C. W. Stansfield (eds.), *Language aptitude reconsidered* (pp. 67-125). Englewood Cliffs: Prentice Hall Regents.

Oxford, R. L. (2003). Language learning styles and strategies: An overview. Learning Styles and Strategies. Oxford, GALA. pp. 1-25.

Phakiti, A. (2006). Modeling cognitive and metacognitive strategies and their relationships to EFL reading test performance. *Melbourne Papers in Language Testing*, 1, 53 95. Retrieved March 12, 2019, from http://

sydney.edu.au/education_social_work/research/publications/resources/phakiti.pdf

Pitts, M. M. (1983). Comprehension monitoring: definition and practice. *Journal of Reading*, *26*(6), 516–523. Retrieved March 16, 2019, from https://www.jstor.org/stable/40031758

Politzer, R. L., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *TESOL Quarterly, 19*,103-123. Doi: 10.2307/3586774

Pressley, M., & Harris, K. R. (2017). Cognitive strategies ınstruction: From basic research to classroom instruction. *Journal of Education,* 189(1-2), 77-94. Doi:10.1177/002205740918900120 6

Rao, N., & Moely, B. E. (2000). Motivational beliefs, study strategies and mathematics attainment in high- and low-achieving Chinese secondary school students. *Contemporary Educational Psychology*, 25, 287-316. Doi:10.1006/ceps.1999.1003

Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical Linear Models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S. W., Bryk, S., Cheong, Y. F., & Congdon, R. (2004). HLM6: Hierarchical linear and nonlinear modeling. Chicago: Scientific Software International.

Schutz, P. A, Drogoz, L.M.,Vhite,V.E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in a introduction to statistics course. *Learning & Individual Differences*, 10 (4), 291-308. Doi: 10.1016/S1041-6080(99)80124-1

Snijders, T., & Bosker, R. (2012) *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* 2nd Edition, London, Sage Publications.

Sobel, Michael E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models". *Sociological Methodology*. 13, 290–312. Doi:10.2307/270723

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*(4), 359–384. Doi: 10.1007/s10648-004-0005-y

Ward, W., & Rosetta, R. (2001). *The effectiveness of instruction in using reading comprehension strategies with eleventh grade social studies students*. Unpublished master thesis, The University of Mississippi.

Wen, Q. F., & Wang, L. F. (2004). Challenging the effectiveness of L2 learning strategy studies. *Foreign language World, 2*, 2-7.

Weinstein, C. E., & Mayer, R.E. (1986). The teaching of learning strategies .*Handbook of research on teaching.*Ed: M.C.Wittock. New York: MacmillanCompany.

Weibel,D.,Wissmath,B., & Mast,F.W.(2011).Influence of menta limagery on spatial presence and enjoyment assessed in different types of media. *Cyberpsychol. Behav.Soc.Netw.* 14, 607–612. Doi:10.1089/cyber.2010.0287

Wilkinson, I. A. G. & Son, E. H. (2011). A dialogic turn in research on learning and teaching to comprehend. In M.L. Kamil. P.B. Rosenthal, P. D. Pearson, & R. Barr (Eds.*), Handbook of reading research: Volume IV (pp. 359-387)*. New York: Routledge.

Yildirim (2012). *Determination of factors relating reading comprehension achievement with hierarchical linear models (PISA 2009 comparison of the Netherlands, Korea and Turkey)*. (Doctoral dissertation, Ankara University, Ankara, Turkey) Retrieved January 18, 2019, from http://acikarsiv.ankara.edu.tr/browse/24723/

Yip, M. C. W. (2007). Differences between high and low academic-achieving university students in learning and study strategies: a Hong Kong study. *Educational Psychology*, 27(1), 597–606. Doi:10.1080/01443410701309126

Zasacka, Z., & Bulkowski, K. (2017). Reading engagement and school achievement of lower secondary school students. *Edukacja,* 2, 78-99. Doi:141. 10.24131/3724.170205.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12(4), 695-719. Doi: 10.1177/1094428108327450

Okumaktan Zevk Alma ve Öğrenme Stratejilerinin PISA 2009 Okuma Performansı Üzerindeki Doğrudan ve Dolaylı Etkilerinin İncelenmesi

Atıf:

Tavsancil, E., Yildirim, O., & Demir Bilican, S. (2019). Direct and indirect effects of learning strategies and reading enjoyment on PISA 2009 reading performance. *Eurasian Journal of Educational Research,* 82, 169-190, DOI: 10.14689/ejer.2019.82.9

Özet

Problem durumu: Gelişen ve sürekli rekabetin yaşandığı dünyamızda, okuduğunu anlayan bireyler yetiştirmek önemlidir. Bireyin kişisel, sosyal ve ekonomik olarak kendinin geliştirebilmesi için okuma önemli bir role sahiptir. Bireylerin okuduğunu anlama sürecinde, ilişki kurma, karşılaştırma yapma, kanıt sunma, genelleme yapma, çıkarımda bulunma ya da olayları yordama gibi daha karmaşık düşünme süreçlerini kullanmaları gerekmektedir. Bu da akademik başarıyı destekleyen bir faktördür. Bu

bakımdan öğrencilerin okuduğunu anlama performansı bakımından durumlarının değerlendirilmesi, bu beceriyi geliştirmek üzere yapılacak uygulamalara yön vermesi açısından bu beceriyle doğrudan ya da dolaylı olarak ilişkili faktörlerin belirlenmesi giderek daha önemli hale gelmektedir. Son yıllarda okuma becerisinin geliştirilmesi için öğrencinin kullandığı stratejiler ve metodlar üzerinde ağırlıklı olarak durulmaktadır. Bu çalışmada, öğrencilerin okuma performansı ile ilişkisini belirlemek üzere metabilişsel stratejiler (kontrol, ezberleme ve detaylandırma) ele alınmıştır. Ayrıca, bu iki değişken arasındaki ilişkide okumaktan zevk almanın olası aracı etkisi de incelenmiştir.

Araştırmanın amacı: Bu çalışmanın amacı, öğrencilerin kullandıkları öğrenme stratejileri ve okumaktan zevk alma değişkenlerinin öğrencilerin PISA 2009 okuma başarılarını ne derece yordadığı ve okuma başarısı ile öğrenme stratejileri arasındaki ilişkide okumaktan zevk alma değişkeninin aracı etkisinin olup olmadığının incelenmesidir.

Araştırmanın Yöntemi: İlişkisel tarama modelindeki bu araştırmada örneklemi, PISA 2009 uygulamasına katılan 170 okuldan 4996 Türk öğrenci oluşturmaktadır. Verilerin düzenlenmesinden sonra analizler 4648 öğrenci ve 169 okul üzerinden yürütülmüştür. Araştırmada, Türkiye PISA 2009 okuduğunu anlama bilişsel alan testi ve öğrenci anketinden toplanan veriler kullanılmıştır. İlgili veriler OECD'nin resmi web sayfasından elde edilmiştir. Öğrencilerin okuduğunu anlama başarısı ile ilişkili olduğu düşünülen "öğrenme stratejileri ve okumaktan zevk alma" yordayıcı değişken olarak; öğrenme stratejiler olarak tanımlanan kontrol, hatırlama ve detaylandırma stratejileri içsel değişkenler, okumaktan zevk alma ise aracı değişken olarak belirlenmiştir. Kurulan model yoluyla okuma başarısı üzerinde öğrenme stratejilerinin dolaylı etkisi incelenmiştir. Ölçeklerin geçerlik ve güvenirlik çalışmaları PISA 2009 teknik raporunda ayrıntılarıyla raporlanmıştır. Araştırmada PISA örnekleme yaklaşımının hiyerarşik yapısı dikkate alınarak HLM'ye dayalı analizler yapılmıştır.

Araştırmanın Bulguları: Öğrenme stratejilerinin okuma performansı üzerindeki doğrudan etkileri test edilmiş, kontrol stratejisi ve hatırlama stratejisi kullanma sıklığının öğrencilerin okuma performansını yordamada anlamlı bir değişken olduğunu ($p<0.01$), detaylandırma stratejisi kullanımının ise anlamlı bir yordayıcısı olmadığı belirlenmiştir ($p>0.01$). Kontrol ve detaylandırma stratejilerini kullanım sıklığı, öğrencilerin okumaktan zevk almasını anlamlı bir şekilde yordamakta ($p<0.01$); ancak hatırlama stratejisini kullanım sıklığı öğrencilerin okumaktan zevk almasını anlamlı olarak yordamamaktadır ($p>0.01$). Bu bakımdan hatırlama stratejisinin okuma başarısı üzerinde dolaylı etkisi olmadığı belirlenmiştir. Okumaktan zevk alma değişkeninin aracı etkileri test edildiğinde ise kontrol stratejisi, hatırlama stratejisi ve okumaktan zevk alma stratejisinin okuma performansının anlamlı olarak yordamış ($p<0.01$); detaylandırma stratejisi ise bu modelde de anlamlı bir yordayıcı olmamıştır ($p>0.01$). Okumaktan zevk alma değişkeninin aracı etkisiyle, kontrol ve detaylandırma stratejilerinin okuma puanları üzerindeki etkisi incelenmiş; bu iki stratejinin kullanım sıklığının okumaktan zevk alma değişkeni üzerinden öğrencilerin okuma performansını anlamlı olarak yordadığı görülmüştür ($p<0.01$).

Araştırmanın Sonuç ve Önerileri: Okuma sırasında kontrol stratejilerini daha sık kullanan öğrencilerin okuma puanlarının daha yüksek olduğu belirlenmiştir. Öğrencilerin ilgili metinleri anlamak üzere daha üst düzey öğrenme stratejileri kullanmaları onların okuma puanlarını artıran bir faktör olmaktadır. Hatırlama stratejisinin okuduğunu anlama başarısı ile negatif yönlü bir ilişkisi bulunmuştur. Bu durum ezberleme ağırlıklı strateji kullanımının, farklı becerileri bir arada içeren okuma sürecini desteklemediğini işaret etmektedir. Beklenilenin aksine bu araştırmada detaylandırma stratejisinin okuma başarısı üzerinde doğrudan bir etkisi bulunmamıştır. Okumaktan zevk alma, öğrencilerin okuma performansını artıran bir değişken olarak belirlenmiştir. Bu durum, daha iyi okuyucuların, okuduklarından zevk aldığını, gönüllülükle okumaya devam ettiklerini ve ayrıca okul performanslarını arttırma eğiliminde olduğunu göstermektedir. İlgili modelde öğrenme stratejilerinin okumaktan zevk alma değişkeni üzerindeki doğrudan etkileri incelenmiş; kontrol ve detaylandırma stratejilerini kullanma sıklığının öğrencinin okumaktan zevk almasını doğrudan etkilediği, hatırlama stratejisinin ise öğrencinin okumaktan zevk almasını doğrudan etkilemediği görülmüştür. Bu durum, öğrenmeye yönelik üst düzey stratejileri daha sıklıkla kullanan öğrencilerin, okumaktan daha çok zevk aldığını göstermektedir. Değişkenler arasındaki dolaylı etkiyi gösteren sonuçlar incelendiğinde, okumaktan zevk alma kontrol stratejisi değişkeni için kısmi aracı etkiye sahipken, detaylandırma stratejisi için tam aracı bir etkiye sahip olmuştur. Sonuç olarak, bu araştırma üst düzey beceri içeren öğrenme stratejilerini kullanan öğrencilerin okumaktan daha çok zevk aldığını ve bu öğrencilerin okuduğunu anlama konusunda daha başarılı olduğunu göstermektedir. Araştırma sonuçları, üst düzey beceri içeren öğrenme stratejilerini kullanan öğrencilerin okumaktan daha çok zevk aldığını ve bu öğrencilerin okuduğunu anlama konusunda daha başarılı olduğunu göstermektedir. Bu durumda, öğrencilerin okumaktan zevk almalarını sağlayacak farklı öğretim materyalleri ve etkinliklerin öğretmenler tarafından kullanılması, öğrencilerin okuma sırasında üst düzey becerilerini geliştirmelerini ve kullanmalarını sağlayacak öğretim ortamlarının oluşturulması ve bu becerilerin nasıl kullanılacağının öğretilmesine yönelik etkinliklerine öğretim programlarında yer verilmesi önerilmektedir. Okul dışında ise, aileler öğrencilerin ilgisini çekebilecek farklı türde kitaplar edinerek okumayı çocuklar için daha zevkli hale getirebilir. Aileler, çocukların okunan metne ilişkin var olan bilgilerini kullanarak, yorum yapma ve yeniden anlam verme gibi becerilerini kullanmalarına yönelik birlikte etkinlikler yapabilir.

Anahtar Kelimeler: Okuma performansı, öğrenme stratejisi, okumaktan zevk alma, doğrudan-dolaylı etki, aşamalı doğrusal model.

Eurasian Journal of Educational Research 82 (2019) 191-202

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

The Contribution of Test Type and Curriculum Difference on the Effect of the National Test Score at International Mathematic Test Score: The Challenge of IR 4.0 Curriculum

Setyabudi INDARTONO[1], Anwaril HAMIDY[2]

ARTICLE INFO

Article History:
Received: 13 Feb. 2019
Received in revised form: 8 Apr. 2019
Accepted: 28 Jun. 2019
DOI: 10.14689/ejer.2019.82.10

Keywords
Learning evaluation, educational policy, learning result

ABSTRACT

Purpose: The students' competencies of math are required to be assessed by using an international scale of evaluation. Thus, education policies and learning evaluation are developed by educational policymakers to improve the education system and result quality. Hence, the purpose of this research was to investigate The Contribution of International Mathematics Test Type and Curriculum Difference on the Effect of the National Test Score on International Mathematics Test Score.

Research Methods: This study employed 600 students of private and public Junior High Schools in Indonesia. The Mathematics Test Types included were PISA-like and TIMSS-like. They were used to represent the type of test. The model of curriculums applied in Junior High School in Indonesia used in this study was KTSP and K13. The multiple regression was used to analyze the data.

Findings: The National Test Score had a stronger effect on International Mathematics Test Score for KTSP curriculum implementation rather than K13 curriculum, and National Test Score had a stronger effect on International Mathematics Test Score for PISA type of International mathematic test rather than TIMSS one. These results showed that the different curriculum used at the mathematics learning process and test types had a moderate effect of National on International test score in a mathematics course.

Implications for Research and Practice: It implies an increase in the policymakers to pay more attention to the curriculum design and type of learning evaluation. It is suggested that future research should include learning and teaching types to find out a deeper behavioral and attitudinal understanding of learning and teaching of mathematics.

[1] Universitas Negeri Yogyakarta, INDONESIA; Economy Department, Email: Setyabudi_indartono@uny.ac.id ; ORCHID: http://orcid.org/0000-0003-2289-9216

[2] Corresponding Author; Institut Agama Islam Negeri Samarinda, INDONESIA; Economy and Business Department, Email: serahdeh@gmail.com

Introduction

The issues of Industrial Revolution 4.0 (IR 4.0) are believed to affect the curriculum design to adjust student learning (Coşkun, Kayıkcı, & Gençay, 2019; Fung, 2019; Lan & Vu, 2018; Lieu, Duc, Gleason, Hai, & Tam, 2018). The IR 4.0 requires the standards of learning processes and output in global competition. Hence students' mathematics skill is recently required to be assessed by using international scale of evaluation (Chow & Ekholm, 2018; Moyer-packenham et al., 2018). Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) are usually used to measure international mathematical skills. PISA is used to measure the reading literacy, math literacy, and science literacy. The term of literacy in PISA is related to the readiness of students on the future workplace, i.e. the ability to apply knowledge and skills to a subject that includes analysis skills, give a reason and make an effective communication for various problems in different situations (OECD, 2013). TIMSS is used to measure Mathematics and Science achievement (Mullis, Martin, Ruddock, Sullivan, & Preuschoff, 2009). Hence, the results of PISA and TIMSS can compare the capability of different students on mathematics among countries. Those international tests show the rank of the country on the result of their students' scores on PISA and TIMSS.

The PISA scores of Indonesian students showed that the literacy score is under the average of OECD. Their TIMSS score indicated that their mathematics skill has got lower and lower from 1999 to 2011. In the period of 1999/ 2000, the rank of Indonesian students among all the participating countries was 39 in 41 (95.12) for PISA and 34 in 38 (89.47) for TIMSS, whereas in 2003 period, Indonesian students reached the rank of 38 in 40 (95.00) for PISA and 35 in 46 (76.09) for TIMSS. In the period of 2006/ 2007, the rank of Indonesian students among the countries was 50 in 57 (87.72) for PISA and 36 in 49 (73.47) for TIMSS, then reached the rank of 61 in 65 (93.85) for PISA and achieved 64th place in 65 (98.46) for PISA and 38 in 42 (90.48) in 2011/ 2012. Finally, in the period of 2015 PISA achievement for Indonesian student reached at the rank of 66 in 72 (91.67) (Mullis, Martin, Foy, & Arora, 2011; OECD, 2013, 2015). Hence these PISA and TIMSS achievements encouraged the education policy makers to develop the mathematical learning system to reach higher achievements. Scholars were also promoted to investigate various antecedents related to the international test results.

The Rationale for the Study and Design

Since various approaches to mathematics tests are designed by education policymakers, the curriculum is believed to increase the international test scores. The curriculum is designed to increase students' achievement in mastering various abilities on a subject (Dekant, Sungur, & Yerdelen, 2018; Mullis et al., 2009). The results of PISA and TIMSS are used to evaluate and re-formulate the education policies by various countries (Carnoy, Khavenson, Loyalka, Schmidt, & Zakharov, 2016; Jansen, Schroeders, Lüdtke, & Marsh, 2019). Accordingly, it is plausible that this study aimed to investigate the effect of the result of national tests on international ones, where the curriculum and type of mathematics tests are used to moderate the model. Hence, the

contribution of test type and curriculum on the model is clearer. The novelty of this study is to examine and verify the contribution of different type of PISA and TIMMS, and different curriculum of KTSP (stand for *Kurikulum Tingkat Satuan Pendidikan* or Education Unit Level Curriculum) and K13 (stand for *Kurikulum* 2013 or Curriculum 2013).

Scholars believed that theoretically national test scores are related to the international scale of mathematics achievement (Carnoy et al., 2016; Stacey, 2011). Hence the mathematic test in the level of national policy is designed to predict the international scale one. It is found that when students achieve a higher score of the school or nation test, they also get a better TIMSS score (Kaleli-Yılmaz & Hanci, 2015). Accordingly, governments try to include the PISA and TIMSS test type in their national test to increase international mathematics' test score. The total items of PISA and TIMSS type tests were increased in national tests to adapt students, and similarity of those tests on the international test types were also increased (Kaleli-Yılmaz & Hanci, 2015; Retnawati, 2017). Replication of Yilmaz & Hanci is needed to demonstrate the relationship between national and international test scores (Kaleli-Yılmaz & Hanci, 2015). However, few researchers investigated the effect of national test score on the international test score. PISA and TIMSS tests have a different objective to measure (Wu, 2010). The content of PISA and/or TIMSS within the national test is believed to have various effects both in national and the international test scores. Empirically, different achievements of Indonesian students on international mathematics score could be predicted by various antecedents. Many researchers found that characteristics and quality of school and students contribute to student achievement. They argue that learning opportunity and teacher quality have different effects on mathematics score of students. The number of students in a class and socioeconomic status of the school is indicated to be a strong predictor of PISA score (Argina, Mitra, Ijabah, & Setiawan, 2017; Carnoy et al., 2016; Lam & Lau, 2014; Mcconney & Perry, 2010). Hence the study of the effect of national test scores on international test scores in different types of tests is needed.

The curriculum is believed to influence the mathematics skills of students. Scheerens found that educational effectivities are related to student achievement (Scheerens, 2019). The intended, implemented, and achieved curriculum are used to plan, control, and evaluate student achievement (Martin & Mullis, 2016). The KTSP and 2013 curriculum is used in different ways to apply education in Indonesia. The difference of both curriculum types is related to knowledge and skill intention, and learning and evaluation system (Retnawati, Hadi, & Nugraha, 2016). The 2013 Curriculum is close to a student-centered rather than a teacher-centered approach. It uses authentic tests to assess student attitudes, knowledge and skill whereas KTSP tends to use knowledge and skill test. Those different types of curriculum are argued to have various results in student achievements as well as PISA and TIMSS score. The 2013 curriculum type is used as an alternative of KTSP type to increase international mathematics skill. However, both curricula used national tests to measure the mathematics score of student achievement in the scale on the national area. Hence, it is plausible if the 2013 curriculum achieves a better score of international tests rather

than KTSP. However, the studies of evaluation of international mathematics score on both curricula are not conducted. Thus, there is a need for a study investigating the influence of different curriculums on the effect of national test scores on international mathematics score.

This study aimed to investigate the different types of test and curriculum on the effect of national test scores on international mathematics score. It is hypothesized that the score gained from a national test of mathematics has an effect on the score of international tests and different types of the curriculum (K13 and KTSP), and international test types (PISA and TIMSS) have different effects on the national-international test scores. Concordantly, this study is significant in three aspects. Firstly, it strengthens the evidence of the importance of the effect of national mathematics test score on the international score. Secondly, it may explain the significant contribution of International Mathematic Test Type on the effect of National Test Score on International Mathematics Test Score. Finally, it can make significant contributions to curriculum difference on the effect of national test score on an international mathematics test score.

Method

Research Design

The moderation model was used to investigate the different types of tests and curriculum to predict the international mathematics test score based on the national test scores. Multiple regression was used to analyze the model. The moderation effect of curriculum and test types were analyzed by using the multiple regression (Hayes, 2013).

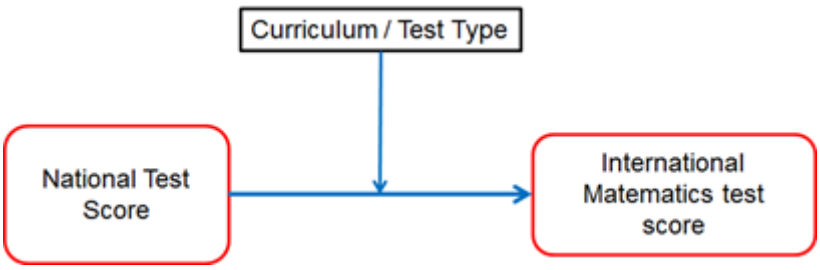


Figure 1: *Research Model*

Sample

This study employed 600 students with different achievement levels of national test scores on a random base. The post-facto data of the result of the national test score in 2016 were used in both KTSP and 2013 curriculum.

Table 1.

Participants

	Category	Participant
National test score	Low	200
	Average	200
	High	200
International test	PISA	300
	TIMSS	300
curriculum	KTSP	304
	2013 curriculum	296
School	Public	263
	Private	337
Gender	Male	279
	female	321

Research Instruments

This study used PISA model of mathematics test (16 items, Cronbach α= 0.835) based on process, content, and context domain (OECD, 2013). The TIMSS model of mathematics test was used (16 items, Cronbach α= 0.837) based on the cognitive and content domain (Mullis et al., 2009).

Data Analysis

The data were gathered from the scores of students' tests of mathematics course in public and private schools. The schools that apply K13 and KTSP curricula were chosen in this study. The Test of Kl-Smirnov showed that the unstandardized residual of the model was .104, the result of ANOVA showed significance value was higher than .05, and tolerance result was higher than 0.1. Hence, the data showed normality, linearity, multicollinearity, and heteroskedasticity of the model. The average score of national mathematics test was 28.71 with a minimum score of 3.23 and a maximum score of 96.77. The average TIMSS score was 8.6 with a minimum score of 1 and the maximum of 30, and the average PISA score was 5.9 with a minimum of 1 and maximum of 18. The descriptive analysis revealed that the average national mathematics score was higher for public schools rather than the private ones (32.37 vs. 25.86), whereas TIMSS score was higher in public schools rather than private ones (9.78 vs. 7.75), and PISA scores for public schools were higher than private ones (6.63 vs. 5.36).

Results

Multiple regression analysis was used to analyze the hypotheses. The result of this study showed that initially, the National Test Score had a significant effect on the International Mathematics Test Score (β=.545***; ΔR^2=10%). A higher score of national tests indicated a better international score of mathematics skills. Whereas the different type of tests had no significant effects on International Mathematics Test Score (β=.043; ΔR^2=18%). The interaction of national test score and test type had a significant effect on the International Mathematics Test Score (β=-.252; ΔR^2=50%). The high contribution of the model (ΔR^2=50%) showed that the type of test has a sufficient ability to explain the different findings of the effect of National Test Scores on International Mathematics Test Score. Surprisingly, Figure 2 explains that PISA type of tests contributes to having a better effect of national test score of mathematics on the international one rather than TIMSS. Hence, the hypotheses were proven.

Table 2.

Contribution of IMTT on the Effect of NTS on IMTS

	step 1	step 2	step 3	step 4
School Status	,173***	,271** *	,270** *	,274** *
National test (UNmc)		,545** *	,545** *	,785** *
Test type (IMT)			0,043	0,043
UNmc*IMT				-0,252*
R^2	0,23	0,33	0,41	0,73
ΔR^2	0,23	0,1	0,18	0,5

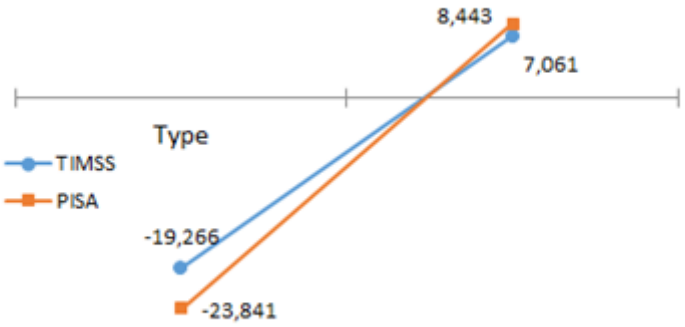


Figure 2. *Contribution of IMTT on the Effect of NTS on IMTS*

Table 3 and Figure 3 shows that interestingly, the National Test Score had a consistent effect on International Mathematics Test Score on different curriculum (β=,486, ΔR^2=10%). A higher score of national tests indicated a better international score of mathematics skills. The different curricula had a significant effect on International Mathematics Test Score (β=,353, ΔR^2=2%). The interaction of national test score and test type had a significant effect on the International Mathematics Test Score (β=,195, ΔR^2=12%). The contribution of the model (ΔR^2=12%) revealed sufficient data to explain the different findings of the effect of the National Test Score on International Mathematics Test Score. Figure 2 explains that KTSP type of curriculum contributed to the better effect of national test score of mathematics on the international one rather than 2013 one.

Table 3.

Contribution of Curriculum on the Effect of NTS on IMTS

	step 1	step 2	step 3	step 4
School Status	,173***	0,101**	0,234***	0,257***
Curriculum Type (CT)		,353***	0,131**	0,125**
National Test (UNmc)			,486***	0,346***
UNmc *CT				0,195***
R^2	0,31	0,33	0,41	0,43
ΔR^2	0,31	0,02	0,1	0,12

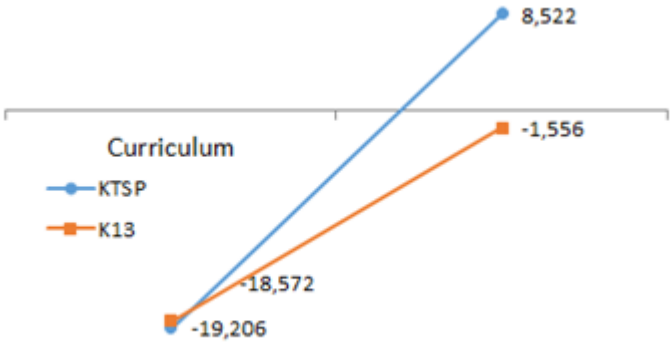


Figure 3. *Contribution of Curriculum on the Effect of NTS on IMTS*

Discussions, Conclusion, and Recommendation

Table 3 and 4 showed consistent findings related to the effect of the National Test Score on International Mathematics Test Score. It is related to the findings of Yilmaz & Hanci (Kaleli-Yılmaz & Hanci, 2015). A higher score of national indicated a better international score of mathematics skills. The reasons are national tests have an equal function with the international one to qualify the mathematical abilities of students, objectives of competencies of national and international test are related (Carnoy et al., 2016; Stacey, 2011), the type of national test of mathematics is widely imitated to mathematics national test (Retnawati, 2017), and the trend of national and international test scores are equal.

The findings indicated that different types of mathematical problems (PISA/TIMSS) are likely to have different impacts of national examinations on international math tests. PISA and TIMSS results showed that the rank on PISA was equal to the ratings on TIMSS. Thus, this confirms the fundamental difference between PISA and TIMSS in measuring mathematical ability (Wu, 2010). The different concepts of PISA and TIMSS provide various benefits to each country. PISA type seems to have a better effect on the national score to the international mathematical score rather than TIMMS. Indonesian students are expected to have better scores on PISA as a contextual test rather than TIMMS. It is related to the decline of the TIMSS scores for Indonesian students since 1999/2000. In the beginning, TIMSS score at 2003/2006 was 411, 2006/2007 was 397, 2011/2012 was 386, and 2015 was 397 (Mullis et al., 2011; OECD, 2013, 2015). However, the scores of PISA are likely to increase. This result seems opposite to the findings of Wu (Wu, 2010). He found that most of the students in Asia and East European countries performed better at TIMMS type than the PISA, yet the western countries concerned for the PISA type.

Different types of the curriculum were also found to have different effects of national to the international score of mathematics test. This finding supported

Scheerens findings (Scheerens, 2019). Hence, it is plausible that the development of the curriculum aimed to increase the result of the learning processes. The KTSP type of curriculum was found to have a better effect of national on the international score of mathematics test. It seems to contradict with the concept of the curriculum development objectives. It is expected that application of KTSP had deeper concerns on knowledge rather than 2013 version of the curriculum which was stressing on attitude and skill (Ilma & Pratama, 2015; Retnawati et al., 2016). It is related to the previous findings indicating that teachers have reached no optimal level of 2013 curriculum implementations (Amat Jaedun, V. Lilik Hariyanto, Nuryadin, 2014; Trisnawati, Gunawan, & Nongkeng, 2016). Hence, the implementation of the 2013 version should be evaluated.

It is concluded that both different curriculum and test types of mathematics have a significant effect on the influence of national to the international mathematics test score. The higher effect of national to international mathematics test scores depended on the context of the curriculum and test type used. Hence school leaders must pay more attention to the implementation of curriculum and choice of the type of test. However, the sample of this study was limited to a specific region in Indonesia. The conclusion of this study shows the effect of national test scores on international test scores is higher when the type of international test is PISA rather than TIMSS, and school which applies KTSP curriculum on their learning process have a better effect of national test scores of mathematics on international test scores rather than 2013 Curriculum.

The result of this study recommends the following studies to investigate broader sampling which represents the model in Indonesia. The implementation of the 2013 curriculum is still in development. Hence, the application of the 2013 curriculum in the ideal stage is needed to represent the concept of the 2013 curriculum. Different objectives of PISA and TIMSS are challenging. Accordingly, future research is suggested to employ a bigger sample implementing the 2013 curriculum and to analyze the details of PISA and TIMSS domain.

References

Amat Jaedun, V. Lilik Hariyanto, Nuryadin, E. R. (2014). An evaluation of the implementation of Curriculum 2013 at the building construction department of vocational high schoos in Yogyakarta. *Journal of Education*, *7*(1), 14–22.

Argina, A. W., Mitra, D., Ijabah, N., & Setiawan, R. (2017). INDONESIAN PISA RESULT : WHAT FACTORS AND WHAT SHOULD BE FIXED ? *Proceedings Education and Language International Conference*, *1*(1), 69–79.

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W. H., & Zakharov, A. (2016). Revisiting the Relationship Between International Assessment Outcomes and Educational Production : Evidence From a Longitudinal. *American Educational*

Research Journal, *53*(4), 1054–1085. https://doi.org/10.3102/0002831216653180

Chow, J. C., & Ekholm, E. (2018). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, *46*, 179–186. https://doi.org/10.1016/j.ecresq.2018.02.011

Coşkun, S., Kayıkcı, Y., & Gençay, E. (2019). Adapting Engineering Education to Industry 4.0 Vision. *Technologies*, *7*(10), 1–13. https://doi.org/10.3390/technologies7010010

Dekant, K., Sungur, S., & Yerdelen, S. (2018). Predicting Science Engagement with Motivation and Teacher Characteristics : a Multilevel Investigation. *International Journal of Science and Mathematics Education*, *17*(1), 67–88.

Fung, D. (2019). Re-thinking the higher education curriculum: Challenges, possibilities, and dreams. *Places of Engagement*, 82–87.

Hayes, A. (2013). *PROCESS model templates*. *New York, NY: Guilford*. https://doi.org/978-1-60918-230-4

Ilma, Z., & Pratama, R. K. (2015). Transformation in Indonesian Language Curriculum: Pros and Cons between KTSP 2006 and Curriculum 2013 in Indonesia, 2013–2016. https://doi.org/10.15242/icehm.ed0315007

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2019). The dimensional structure of students ' self-concept and interest in science depends on course composition. *Learning and Instruction*, *60*(October 2018), 20–28. https://doi.org/10.1016/j.learninstruc.2018.11.001

Kaleli-Yılmaz, G., & Hanci, A. (2015). Examination of the 8th grade students ' TIMSS mathematics success in terms of different variables. *International Journal of Mathematical Education in Technology*, *47*(5), 674–695. https://doi.org/10.1080/0020739X.2015.1102977

Lam, T. Y. P., & Lau, K. C. (2014). International Journal of Science Examining Factors Affecting Science Achievement of Hong Kong in PISA 2006 Using Hierarchical Linear Modeling. *International Journal of Science Education*, *36*(15), 2463–2480. https://doi.org/10.1080/09500693.2013.879223

Lan, T., & Vu, A. (2018). Building CDIO Approach Training Programmes against Challenges of Industrial Revolution 4 . 0 for Engineering and Technology Development, *11*(7), 1129–1148.

Lieu, T. T. B., Duc, N. H., Gleason, N. W., Hai, D. T., & Tam, N. D. (2018). Approaches in Developing Undergraduate IT Engineering Curriculum for the Fourth Industrial Revolution in Malaysia and Vietnam. *Creative Education*, *9*(December), 2752–2772. https://doi.org/10.4236/ce.2018.916207

Martin, M. O., & Mullis, I. V. (2016). TIMSS: Purpose and design. Dalam S. J. Howie & T. Plomp (Eds.), Contexts of learning mathematics and science—Lessons

learned from TIMSS. In *Contexts of learning mathematics and science – Lessons learned from TIMSS*. Routledge.

Mcconney, A., & Perry, L. B. (2010). Science and mathematics achievement in Australia : the role of school socioeconomic composition in educational equity and effectiveness. *International Journal of Science and Mathematics Education*, *8*(3), 429–452. https://doi.org/10.1007/s10763-010-9197-4

Moyer-packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E., Roxburgh, A. L., … Symanzik, J. (2018). How Design Features in Digital Math Games Support Learning and Mathematics Connections. *Comupter in Human Behavior*, *91*, 316–332. https://doi.org/10.1016/j.chb.2018.09.036

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2011). *TIMSS 2011 International Results in Mathematics*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College Chestnut Hill, MA, USA.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., Sullivan, C. Y. O., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD.

OECD. (2015). *PISA 2015 Results (Volume I): Excellence and Equity in Education, PISA* (Vol. I). OECD Publishing, Paris.

Retnawati, H. (2017). Why are the Mathematics National Examination Items Difficult and What Is Teachers ' Strategy to Overcome It ?, *10*(3), 257–276.

Retnawati, H., Hadi, S., & Nugraha, A. C. (2016). Vocational High School Teachers ' Difficulties in Implementing the Assessment in Curriculum 2013 in Yogyakarta Province of Indonesia. *International Journal of Instruction*, *9*(1), 33–48.

Scheerens, J. (2019). School self-evaluation : Origins , definition , approaches , methods and implementation. In *School-based evaluation: An international perspective* (pp. 35–69).

Stacey, K. (2011). The PISA View of Mathematical Literacy in Indonesia. *Journal on Mathematics Education*, *2*(2), 95–126.

Trisnawati, Gunawan, & Nongkeng, H. (2016). Perbandingan implementasi kurikulum tingkat satuan pendidikan (KTSP) dan kurikulum 2013 di SMAN 1 Sinjai Utara. *Jurnal Mirai Management*, *1*(1), 1–9.

Wu, M. (2010). Comparing the Similarities and Differences of PISA 2003 and TIMSS. *OECD Education Working Papers*, (32).

Submission Checklist

Makale Sunumu Kontrol Çizelgesi

ÖNEMLİ: EJER Dergisi makale yazarlarından yayın ücreti talep etmez. Derginin yayın dilinin İngilizce olması sebebiyle dergi kurullarınca yayın kararı verilen makaleler için alan ve dil uzmanlarına proofreading (son okuma) yaptırılır. Dolayısıyla sadece proofreading (son okuma) ücreti alınır.

IMPORTANT: EJER does not charge Article Publication Fee and does not ask authors to make payment. Since the publication language of the EJER is in English, any paper is accepted by the editorial board after the review process is completed sent for the proof reading. Therefore, only the proofreading fee will be charged.

** Aday makalenin değerlendirilmeye hazır olduğunu aşağıdakilerin her biri ile karşılaştırarak kontrol ediniz.*

* Indicate that this submission is ready to be considered by this journal by checking off the following.

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	☒	*The author agrees to pay the amount to be notified to the proofreading of the article and 10 loose copies if the article is accepted.* Yazar, aday makalenin kabul edilmesi halinde makalenin İngilizce son okuma (proofreading) ile 10 serbest kopya gönderimi için kendisine bildirilecek tutarı ödemeyi kabul eder.
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 kaçınınız**.
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; http://www.tandf.co.uk/journals/authors /rereabstracts.asp

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 bitten 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:** 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:…*

♦ Reliability and the validity of the research instrument used or adapted in the work must be provided, and explained in detail.

		Araştırmanın Amacı:… *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 .....: 200 TL*
***Institution** / Kurum Aboneliği .....: 600 TL*
***Abroad** / Yurtdışı Aboneliği:.....: 300 $*
***Student** / Öğrenci Aboneliği:.....: 150 TL*

For single issue / *Tek sayılar için*

***Individual** / Kişiler .....: 40 TL*
***Institution** / Kurumlar.....: 140 TL*

Payment / *Ödeme*

Dollar ($) Account / *Dolar Hesabı*
Account Information for Payment
Account Name: Anı Yay. Rek. Dan. Eğit. Ltd. Şti.
Bank Name: Is Bank
IBAN Number: TR82 0006 4000 0024 2141 3300 29
SWIFT CODE: ISBKTRISXXX
TURKEY

Turkish Lira (TL) Account / *Türk Lirası Hesabı*
Banka: İşbank
Hesap Adı: Anı Yay. Rek. Dan. Eğit. Ltd. Şti.
Şube Kodu: 4214
Hesap No: 962561
IBAN: TR14 0006 4000 0014 2140 9625 61