

RABİA CEREN KARACA

DIMENSIONS OF STUDENT EVALUATION OF TEACHING FORMS AT
UNIVERSITY LEVEL

RABİA CEREN KARACA

A THESIS SUBMITTED
FOR
THE DEGREE OF MASTER OF ARTS
IN
CURRICULUM AND INSTRUCTION

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

SEPTEMBER 2022

2022



Dedicated to my mom and dad...

As appreciation for their unconditional love, sacrifices, and encouragement...

Dimensions of Student Evaluation of Teaching Forms at University Level

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Rabia Ceren Karaca

In Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in

Curriculum and Instruction

Ankara

September 2022

İHSAN DOĞRAMACI BILKENT UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

Dimensions of Student Evaluation of Teaching Forms at University Level

Rabia Ceren Karaca

September 2022

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Asst. Prof. Dr. İlker Kalender (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Asst. Prof. Dr. Armağan Ateşkan (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum and Instruction.

Prof. Dr. Halil Giray Berberoğlu, Başkent University (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Orhan Arıkan (Director)

ABSTRACT**DIMENSIONS OF STUDENT EVALUATION OF TEACHING FORMS AT
UNIVERSITY LEVEL**

Rabia Ceren Karaca

M.A. in Curriculum and Instruction

Thesis Advisor: Asst. Prof. Dr. İlker Kalender

September 2022

When instructional effectiveness is evaluated, it is essential to gather data from different sources to obtain valid judgments. Among several ways to assess instructional effectiveness, Student Evaluation of Teaching (SET), also known as Student Ratings of Instruction (SRI), is a process applied towards the end of the term in many academic settings to collect data generally through a form on the students' views about an instructor's in-class teaching performance. However, certain factors might affect students' responses to the items in the form during the SET process. That is, the dimensions of the form might be different from the initially developed one and different dimensions with different items might arise. The aim of the current study is to analyze the dimensions of an SET form in the preparatory school of a non-profit private university to see how different groups perceive the items in the SET form. The participants are all English language learners from different English language levels. The data were collected in the Fall term, in the 12th week of teaching, in the first contact hour. The data were analyzed through a software program running principal component analysis (PCA). The findings revealed that the structure of the form can be considered unidimensional and there is one dominant latent factor, named as *instructional effectiveness* by the researcher. In other words, the responses of the different groups are mostly related to instructional effectiveness and are not affected by other potential factors.

Keywords: Student evaluation of teaching, instructional effectiveness, factorial structure

ÖZET

ÜNİVERSİTE DÜZEYİNDEKİ ÖĞRETİME YÖNELİK ÖĞRENCİ DEĞERLENDİRME ANKETLERİNİN BOYUTLARI

Rabia Ceren Karaca

Eğitim Programları ve Öğretim Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi İlker Kalender

Eylül 2022

Öğretim etkililiği değerlendirilirken, geçerli bulgular elde etmek için farklı kaynaklardan veri toplamak esastır. Öğretim etkililiğini değerlendirmenin çeşitli yollarından biri olan Öğretim Elemanı Değerlendirmesi genellikle bir form aracılığıyla bir öğretim elemanının sınıf içi öğretim performansı hakkında öğrenci görüşlerine ilişkin veri toplamak amacıyla birçok akademik ortamda dönemin sonuna doğru uygulanan bir yöntemdir. Ancak bazı faktörler, öğrencilerin değerlendirme sürecinde formdaki maddelere verdiği yanıtları etkileyebilir. Yani, formun boyutları başlangıçta geliştirilenden farklı olabilir ve farklı sorular üzerinde farklı boyutlar ortaya çıkabilir. Bu çalışmanın amacı, bir vakıf üniversitesinin hazırlık okulunda kullanılan öğretim elemanı değerlendirme formunun boyutlarını inceleyerek farklı grupların değerlendirme formundaki maddeleri nasıl algıladıklarını görmektir. Katılımcıların tümü çeşitli seviyelerde İngilizce öğrenmekte olan kişilerdir. Veriler güz döneminde, öğretimin 12. haftasında, ilk ders saatinde toplanmıştır. Veriler, temel bileşenler analizi yapan bir yazılım programı kullanılarak incelenmiştir. Bulgular, formun yapısının tek boyutlu olarak kabul edilebileceğini ve araştırmacı tarafından *öğretim etkililiği* olarak adlandırılan bir baskın gizil faktörün olduğunu ortaya koymuştur. Başka bir deyişle, farklı grupların cevapları çoğunlukla öğretim etkililiği ile ilgili olup diğer potansiyel faktörlerden etkilenmemektedir.

Anahtar kelimeler: Öğretim elemanı değerlendirme, öğretim etkililiği, faktör yapısı

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Asst. Prof. İlker Kalender, for his continuous support, constant encouragement, patience and visionary guidance.

I would also like to extend my gratitude to my thesis committee members, Asst. Prof. Dr. Armağan Ateşkan and Prof. Dr. Halil Giray Berberoğlu for their constructive feedback, valuable contributions and guidance.

It is hard to say “thanks” to some people who deserve much more than mere words. Friends are the family we choose for ourselves. Funda Kelahmetoğlu Tunçer and Merve Sivri Aras, you are my family. Thank you for always being there for me.

No words can express my sincere gratitude to my mom and dad. I am forever indebted to them for their unconditional love, continuous encouragement and incredible sacrifices. I am so lucky to have you both!

Lastly, I should express my appreciation to myself to be able to overcome all the obstacles along this long and overwhelming journey, thanks to which I have learned a lot. However, this is not the end of the journey since there is a whole other world out there to explore...

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZET.....	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER 1: INTRODUCTION	1
Introduction	1
Background	3
Problem	8
Purpose	10
Research Questions	10
Significance	10
CHAPTER 2: REVIEW OF RELATED LITERATURE	12
Introduction	12
Teaching Effectiveness.....	12
Evaluation Procedures	13
Self-Assessment.....	14
Peer Evaluation	14
Mentor Coaching	14
Teaching Portfolio	15

Video-Based Observation	15
Student Evaluation of Teaching (SET)	16
Pro-Arguments for the Use of SET	17
Counter Arguments Against the Use of SET	19
Potential Factors Affecting SET Scores	20
Characteristics Related to The Administration of Evaluations.....	20
Course Characteristics	21
Instructor Characteristics	22
Student Characteristics	24
CHAPTER 3: METHOD	26
Introduction	26
Research Design	26
Context	27
Participants	29
Instrumentation.....	30
Method of Data Collection	32
Method of Data Analysis.....	33
CHAPTER 4: RESULTS	37
Introduction	37
Descriptive Statistics Results	37
Principal Component Analysis (PCA) Results.....	42
The Whole Sample.....	42

A Level	47
AF Level	52
B Level.....	57
C Level.....	62
CR Level	67
MC Domain	72
RW Domain	77
LS Domain.....	82
CHAPTER 5: DISCUSSION.....	89
Introduction	89
Overview of the Study.....	89
Discussion of Major Findings	90
Findings Regarding the Whole Sample	90
Findings Regarding the English Language Levels	93
Findings Regarding Three Domains	97
Implications for Practice	100
Implications for Further Research	101
Limitations.....	102
REFERENCES.....	103
APPENDIX	119

LIST OF TABLES

Table		Page
1	Descriptive Statistics Regarding Participants Based on English Language Levels and Domains.....	30
2	Items of the SET Form.....	31
3	The Number of Students across English Language Levels and Domains.....	33
4	Descriptive Statistics Based on English Language Levels.....	38
5	Descriptive Statistics Based on Domains.....	38
6	Descriptive Statistics Based on Instructors.....	39
7	Means of Items across Domains.....	39
8	Means of Items across English Language Levels.....	41
9	The Whole Sample - Total Variance Explained Based on the Initial Eigenvalues.....	43
10	The Whole Sample - Component Labeling.....	45
11	The Whole Sample - Factor Loadings.....	46
12	A Level - Total Variance Explained Based on the Initial Eigenvalues.....	48
13	A Level - Component Labeling.....	50
14	A Level - Factor Loadings.....	51
15	AF Level - Total Variance Explained Based on the Initial Eigenvalues.....	53
16	AF Level - Component Labeling.....	55
17	AF Level - Factor Loadings.....	57
18	B Level - Total Variance Explained Based on the Initial	

	Eigenvalues.....	58
19	B Level - Component Labeling.....	60
20	B Level - Factor Loadings.....	62
21	C Level - Total Variance Explained Based on the Initial Eigenvalues.....	63
22	C Level - Component Labeling.....	65
23	C Level - Factor Loadings.....	67
24	CR Level - Total Variance Explained Based on the Initial Eigenvalues.....	68
25	CR Level - Component Labeling.....	70
26	CR Level - Factor Loadings.....	72
27	MC Domain - Total Variance Explained Based on the Initial Eigenvalues.....	73
28	MC Domain - Component Labeling.....	75
29	MC Domain - Factor Loadings.....	76
30	RW Domain - Total Variance Explained Based on the Initial Eigenvalues.....	77
31	RW Domain - Component Labeling.....	80
32	RW Domain - Factor Loadings.....	82
33	LS Domain - Total Variance Explained Based on the Initial Eigenvalues.....	83
34	LS Domain - Component Labeling.....	85
35	LS Domain - Factor Loadings.....	86

LIST OF FIGURES

Figure		Page
1	Means of Items across Domains.....	40
2	Means of Items across English Language Levels.....	42
3	The Whole Sample - Scree Plot.....	44
4	A Level - Scree Plot.....	49
5	AF Level - Scree Plot.....	54
6	B Level - Scree Plot.....	59
7	C Level - Scree Plot.....	64
8	CR Level - Scree Plot.....	69
9	MC Domain - Scree Plot.....	74
10	RW Domain - Scree Plot.....	79
11	LS Domain - Scree Plot.....	84

CHAPTER 1: INTRODUCTION

Introduction

Education is one of the most significant concepts for the development of a country's socio-economic situation. Citizens have some responsibilities to their country and education is the tool which they need to fulfill these responsibilities. This makes education a necessity in the global world today (Zuehlke, 2008). As a consequence of this, the need for higher education institutions has increased over the past number of years (Usher, 2009) and the increase in the number of these institutions has led to a great deal of competition among them (Naidoo, 2016). The competition has been in a number of different areas and teaching quality is one of these areas.

Today, teaching quality is regarded as one of the most important criteria and a reshaping power in academic settings. Teaching quality can be achieved with effective teaching. Even though there seems to be no consensus on what effective teaching is (Hornstein, 2017), there are several researchers who examined it (Aregbeyen, 2010; Berk, 2005; Ramsden, 2003; Young & Shaw, 1999) and it has been defined in various ways. Allan and Clarke (2007) defined it as an environment where students are independent learners, feedback is valuable, meta-cognitive skills are developed, a supportive learning atmosphere is created, student autonomy is given importance and skill-oriented outcomes are obtained. Additionally, Ramsden (2003) stated that effective teaching means efficient teaching which encourages deep learning, raises interest in the subject and causes changes in students' behaviors.

In order to prove their teaching quality, most universities today are expected to apply appropriate evaluation practices and processes as a requirement (Kember et

al., 2002). These practices are self-evaluation which aims to raise instructional awareness generally with a checklist, peer evaluation which enables instructors to evaluate each other by an evaluation form and provide feedback on their teaching practices, classroom research techniques like portfolios, student reviews, minute papers, and lastly faculty evaluation which aims to obtain feedback from the students by an evaluation form (Dechesne, 1995). Also, in other studies, these practices are listed as peer reviews, observations, self-ratings, peer-ratings, videos, student interviews, employer ratings, administrative ratings, learning outcome measures, mid-semester reviews, criterion-referenced evaluation, teaching portfolios and evaluation by students (Beck, 1990; Berk, 2005; Ray et al., 2018). When all are considered, evaluation by students on instructional effectiveness has been the subject of a great deal of research all over the world (Marsh, 1984).

Instructor evaluations by students are supposed to provide valuable feedback to academic staff on their teaching practices, information for students to select their courses and data to administrators in order to make personnel decisions. So, the system designed for instructor evaluations should have certain characteristics. The evaluation process in an academic setting should be reliable, multidimensional, rigorous, defensible, grounded on research in the field and relatively unaffected by the potential biases that may distort the interpretation of the results (Marsh, 1984).

Even though commonly employed, instructors' being evaluated by students has been a controversial issue for a great many years. A lot of research on the contaminating factors which have the potential to create validity issues has been conducted so far and many researchers have focused on the factors related to instructor characteristics, student characteristics or implementation procedures during the process of their studies. However, the findings of these studies are in

conflict. Kulik (2001) stated that though the findings of the studies are in conflict, a large number of them indicated that the results of the evaluations by students are affected by some factors and these factors should be taken into consideration for possible validity issues.

Background

While evaluating instructional effectiveness, obtaining data from different sources is critical to reach valid judgments. Among several ways to assess instructional effectiveness, student evaluation of teaching (SET), also known as Student Rating of Instruction (SRI), is a process applied in many academic settings towards the end of the term to collect data from the students on their ideas about instructors' teaching performance (Mace, 1997). However, among other alternatives, data obtained through SET are considered more reliable since these data are gathered through the observation of a great many students facing multiple teaching situations (Benton & Cashin, 2004). This may be the reason why McKeachie (1979) regarded it as the most reliable and valid data source while evaluating instructional effectiveness and why it has been applied in academic settings for many years.

SET dates back to the early 20th century. The first use of SET to evaluate instructors began in American colleges in the early 1920s and the oldest ongoing SET program belongs to the University of Washington (Canelos, 1985). In the 1960s, SET was used to evaluate the faculty and to select a course effectively due to the requests for public transparency. Public accountability was critical for higher education institutions and SET was used primarily for this purpose. In the 1970s, academic departments started to use the data collected through SET for various aims and this process started to go beyond evaluating academic staff. Generally, evaluation data were gathered to help instructors develop themselves professionally.

In the 1980s and 1990s, it started to be used for administrative decisions in addition to faculty and student needs (Ory, 2000). Since then, SET has been increasingly common while hiring new academic staff even though they were first introduced to provide student input in order to enhance teaching (Galbraith et al., 2011). Today, SET has been used in a large number of higher education institutions for different purposes.

Administered either electronically or in paper form, using an evaluation form is one of the common ways used to evaluate teaching performance. Dechesne (1995) stated that these forms might be in different formats like open-ended questionnaires and closed-question instruments. However, the most common one is an evaluation form including both open-ended and close-ended items with additional comments part at the end. Tennant and Khamis (2017) indicated that these forms typically use Likert scales having points like strongly agree, agree, disagree, strongly disagree.

Marsh (1984) noted that SET has been applied in academic institutions to provide feedback to academic staff, to help students while selecting a course and its instructor, to obtain research data for the improvement of teaching and learning quality and to monitor the process of effective teaching and learning. Barth (2008) also added that data obtained at the end of the SET process have been used in order to make administrative decisions during tenure and promotion processes. These purposes are categorized under two headings: summative and formative by researchers (Hornstein, 2017; Penny, 2003; Seldin, 1993; Zabaleta, 2007). If the data collected in the rating process are used for tenure and promotion decisions, it has a summative purpose. When the results guide instructors to enhance their teaching skills and enable the institution to detect the needs in order to improve effective

teaching and learning, it has a formative purpose (Hobson & Talbot, 2001; Macfeydan et al., 2016).

SET process in academic settings is one of the most controversial issues that has been open to discussion for years. However, it exists as the most frequently used method (Marsh, 1984). One problem with SET is the lack of participation in both traditional and online methods due to some reasons like lack of time, the process's being monotonous, being implemented quite often and students' being unaware that instructors will make use of the data. Another problem is related to the quality of student feedback in SET. It is questioned whether students are mature enough to make objective judgments while evaluating instructional effectiveness. Students might not take the evaluation process seriously, respond truthfully on purpose or answer carefully. Explaining the rationale behind SET, giving detailed instructions and asking students to rate each item carefully and separately are a few suggestions for this problem (Darling-Hammond et al., 2017). Although these two issues are commonly encountered problems, the debate in literature has centred around the possibility of biases that might influence the interpretation of the results (Liu, 2012). Obtaining reliable and valid evaluation results at the end of the process is critical as the results have been used by academic staff for feedback, by the institution for personnel decisions and by the students for course selections.

Although there are leading researchers adopting a positive perspective on the usefulness of SET (Marsh, 1987; Wilson et al., 1997), many other researchers have a skeptical view towards the use of SET in academic settings. They have questioned the usefulness of SET and focused on the validity issues because biased results which are not related to effective teaching are likely to be obtained (Koh & Tan, 1997; Neath, 1996; Sinclair & Kunda, 2000; Zhao & Gallant, 2011). In addition to

the issues arising due to the instrument used for measurement in the SET process, issues related to students like misinterpreting the rationale behind the process, not taking it seriously and filling out the forms in a deceptive way might cause biased results (Clayson & Haley, 2011).

Bias is defined as the fact that the results of research or an experiment are not accurate because a particular factor has not been considered when collecting the information (Oxford University Press, n.d.). It is crucial to understand the impact of bias on each stage of the research as misinterpretation of the findings may be possible. If irrelevant variables during research are not detected, validity concerns may occur. These irrelevant variables are considered as a threat to the credibility of the SET results (Kalender, 2011) because they may cause an unjust evaluation process.

There are a variety of factors reported on bias related to SET in literature and these factors lead to measuring something unintended (Andersen et al., 1997). The halo effect, a type of cognitive bias where individuals' overall impression of a person affects other traits, has been an issue of controversy in certain studies (Clayson & Haley, 2011; Marks, 2000; Orsini, 1988). Also, gender issue has caused biased results in the analysis of the SET results. The gender of both students and instructors might be a contaminating effect on refined evaluation results as indicated by many researchers (Lueck et al., 1993; Macnell et al., 2015; Nargundkar & Shrikhande, 2014; Weinberg et al., 2007). Other factors related to the course characteristics like the course's being elective or required, the course's being offered in the Fall, Spring or Summer term, the year when the course is taught, course workload and the level of the course are also regarded as factors affecting SET results (Beran & Violato, 2005; Beran & Violato, 2009; Feldman, 1978). Class size, class shape, lighting, the

level of noise and technological equipment a class has might be other possible effects on SET (Lei, 2010). Grade expectation factor is another characteristic related to SET and its relationship with SET scores was discussed in certain studies as well (Clayson, 2009; Clayson et al., 2006; Marsh & Roche, 2000; Stroebe, 2020). Identifying these factors might help to assess the SET results and the fact that these factors may affect the scores should be taken into consideration and controlled by certain statistical techniques during the evaluation process (Kalender, 2011).

In educational measurement, it is essential to assess the validity of the SET process due to the possible factors stated above. Typically, the assessment is made through evaluation forms aiming to measure different dimensions based on instructional effectiveness. The responses to the items in these forms are statistically analyzed to see whether they are in the same line with certain criteria of teaching effectiveness. McKeachie (1979) stated that effective teaching includes a great many dimensions and through the analysis of 21 studies of teaching effectiveness, Wotruba and Wright (1975) highlighted 10 criteria. They are having up-to-date knowledge on the subject, encouraging students to think for themselves, being enthusiastic about the subject, open communication, explaining abstract concepts clearly, clarifying concepts with examples, facts and comparisons, fair evaluation, being helpful, creating a motivating atmosphere and enjoying while teaching. According to Baliyan and Moorad (2018), certain dimensions of effective teaching are using effective teaching methods, having classroom management skills, fair assessment and giving feedback. Professionalism, leadership, analytical thinking skills, being confident and reliable are also elements related to effective teaching (Mcber, 2000, as cited in Antony et al., 2019). And in a study by Wimberly et al. (1978), five dimensions were listed. These dimensions were noted under five headings: student development (new

perspectives, learning encouragement, interest in the subject, etc.), task responsiveness (fair assessment, indicating expectations clearly, etc.), respect for students (not insulting, respect for student comments, etc.), teacher capability (enjoying while teaching, being devoted to teaching, etc.) and encouragement (encouraging participation, motivating students, etc.).

Although SET forms are developed considering effective teaching dimensions, it is students' responses that define the final characteristics of the statistical properties such as the factorial structure. These forms may be designed according to certain factorial structures; however; different factorial dimensions may form due to students' perceptions. Items initially developed aiming to measure several concepts may be located under different dimensions at the end of the evaluation process.

Problem

SET is a valuable tool to enhance teaching performance and improve learning outcomes for all stakeholders in an educational institution, among other benefits stated above. So, all around the world, the use of SET has been common to evaluate teaching performance and competence. However, there has been an ongoing issue on the validity, reliability, and application of the data (Tsou, 2020) because how valid and reliable analysis results are obtained at the end of a data collection process is of significance.

Since the pioneering research on SET, a great many studies have been carried out on this issue and a great majority of them have focused on the relationship between students' responses and the possible biasing factors (Abrami, 2001; Aleamoni, 1999; Kalender, 2011; Marsh & Roche, 1997; Spooren et al., 2013; Wachtel, 1998). Even though most of the research conducted on the SET has enabled

the stakeholders in higher institutions to gain useful information, the validity issue of the SET process continues to remain a question as indicated in many studies (Clayson, 2009; Hornstein, 2017; Kalender, 2015)

The validity issues occur during the analysis of the collected data to evaluate instructors. It is much more reliable to gather data to evaluate instructors with various evaluation procedures; however, the most often employed one is asking students to fill out an anonymous paper-based or online SET form having a Likert-type rating scale. The form enables students to express how much they agree or disagree with the items written in it (Constantinou & Wijnen-Meijer, 2022). The responses are analyzed and the scores obtained are interpreted. At this point, it becomes important for all stakeholders in an academic setting to have a clear understanding of the determinants in an SET form as this enables a more refined evaluation process in a more reliable atmosphere. That is, the collected data are expected to be in the same line with the expected structure at the end of the measurement process. The instruments are supposed to measure what is intended at the beginning (Matsunaga, 2010). However, there are some areas still underexplored. A large body of research focuses on undergraduate students in their departments. Yet, preparatory school students, who are at the beginning of the university life, are also a group reporting about instructional effectiveness via SET forms. Therefore, the examination of this group is especially important because younger students may not know how to evaluate an instructor. It might cause that the items in the evaluation form can be perceived in a different way by the students. The factorial structure of the form might change and differ from the initially developed one. Different dimensions with different items might arise across groups. Such

differentiations might create unfair results for instructors due to factors beyond their control. Thus, an analysis of how different groups perceive SET forms is needed.

Purpose

The purpose of this research study is to analyze the dimensions of an SET form used to evaluate preparatory school instructors in a non-profit private university. To this end, the factorial pattern of the SET form was analyzed for the whole participants, across English language levels and across three different courses called main course, reading and writing course and listening and speaking course.

Research Questions

This study addresses the following question:

1. What are the dimensions of the SET form designed in a non-profit private university to measure instructional performance?
 - a. What are the dimensions of the SET form for all the participants in the preparatory school?
 - b. What are the dimensions of the SET form in the preparatory school across English language proficiency levels of students?
 - c. What are the dimensions of the SET form in the preparatory school across three domains (MC, RW and LS)?

Significance

SET is a significant part of higher education for a great many reasons. Thanks to SET in academic settings, instructors get valuable feedback that will be useful to see how effective their teaching practices are (Bangert, 2006), administrators obtain data for promotion, tenure and merit decisions (Clayson & Haley, 2011) and students have information for the selection of courses and instructors (Marsh & Roche, 1993). Also, SET plays an important role in accreditation processes (Pepe & Wang, 2012).

When all these are considered, it is clear that SET is of importance for all stakeholders in academic settings. So, the results of the current study might have both local and general importance.

At the local level, the current study may help the institution in which the study was conducted to have a clear understanding of the evaluation process. Also, the study is useful for the institution to reanalyze the evaluation process which they have been implementing for years thanks to the theoretical knowledge presented in the study and the findings obtained. If necessary, changes are made, the administration might make fair decisions on career advancement opportunities and have an opportunity to reward teaching competence and performance, which is critical to retaining effective instructors. The study is also valuable for the instructors as they have the opportunity to get precise feedback on their teaching practices. With the help of this feedback, they may design a new plan for their professional development. They may have the awareness of their strong sides, identify the areas they need more strength in and focus on how to better shape in terms of teaching skills. In other words, they may learn about, reflect on and enhance their practices.

In general, the current study might contribute to literature as it provides valuable information on the refined instructor evaluation process to other faculties and higher education institutions. They might be more aware of the need for a properly functioning instructor evaluation process. This might eliminate the possible mistrust in the current system they apply for academic staff. Even though this study focused on the instrument used at a non-profit private university, the methodological ways that I follow apply to other institutions as well.

CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

This chapter aims to introduce the theoretical framework of the current study by presenting the existing literature on the issue. Firstly, the concept of effective teaching was clarified with definitions from literature and the evaluation procedures used to measure effective teaching were provided. After that, student evaluation of teaching (SET) was explained by underlining the pro and counter arguments for and against the use of it with certain studies from the literature. Lastly, potential factors that might create biased results in the SET process were discussed in the studies carried out on the issue by different researchers.

Teaching Effectiveness

Teaching quality has become an issue of debate in higher education institutions recently (Chen & Hoshower, 2003). Consequently, administrators of the universities have focused on the improvement of teaching quality by making some reforms in curricula, hiring qualified academic staff, giving priority to professional development, etc. (Baliyan & Moorad, 2018). Therefore, a need to measure teaching quality has occurred, which means that teaching effectiveness, the enhancement of which is critical for teaching quality, has to be evaluated.

Teaching effectiveness has been a debatable term in literature for years as it is quite difficult to define it in a clear way. Despite not having an exact definition, the term has been explained by many authors in various ways. Ding and Sherman (2006) underlined the difference between effective teaching and efficient teaching

while explaining the term. For them, effective teaching is about considering the needs of students instead of applying the right teaching strategies and techniques in classes.

According to Seidel and Shavelson (2007), teaching effectiveness means how satisfied students are at the end of their learning practice, which means that teaching effectiveness is about learning experiences. Galbrath and Merrill (2012) noted that teaching effectiveness is related to learning outcomes. McKeachie (1997) stated that whether students achieved learning goals or not defines effective teaching; however, there might be other factors affecting teaching effectiveness like student ability. Also, the achievement scores of the students might not be the same in different courses. So, how teaching effectiveness can be measured becomes a questionable topic in literature today.

Since there is not an agreed acceptable model explaining teaching effectiveness, there might be different factors affecting it. However, Ding and Sherman (2006) stated that the instructor factor has the strongest effect on teaching effectiveness. A strong cause and effect pattern between student learning outcomes and instructor characteristics is presented in the literature. Though not much is known about the exact psychological patterns causing a powerful relationship between student learning and instructor characteristics, the general conclusion is that when the characteristics of the instructors create a positive impact on students, students can feel less anxious, constitute structured knowledge easily and form a more developed self-awareness (Murray, 2007).

Evaluation Procedures

Multiple evaluation procedures should be applied to get a full picture of an instructor's teaching performance. If the biggest purpose of instructor evaluations is

professional growth, various kinds of evidence are required for a fair evaluation process.

Self-Assessment

Self-assessment is a kind of reflection on instructors' own teaching practices. It is a part of formative assessment because instructors assess their own teaching and share the findings with their peers, consultants or mentors for professional development (Sayavedra, 2014). According to Airasian and Gullickson (1994), it has the steps of identifying the problem, analysis of practice, reflection and decision-making. Instructors focus on students' paying attention to the course, body language, knowledge of the subject, mimics or the questions asked during the course to evaluate their own instructional success (Jones & Airasian, 1996). However, being a novice instructor, lack of training or some uncontrollable situations might not lead to positive changes at the end of self-assessment procedures (Pourjamal et al., 2018).

Peer Evaluation

Peer evaluation means two instructors work together by observing each other's classroom environment, instructional practices, lesson delivery and discussion what might be different for professional growth (Sayavedra, 2014). According to Huston and Weaver (2007), one of the instructors is expected to be an experienced one. For an effective peer evaluation process, first they identify a focus theme, they have an in-class observation and they arrange a session to share the observation notes with feedback. Training is very significant for this evaluation procedure as well because a judging atmosphere is likely to occur (Sayavedra, 2014).

Mentor Coaching

Mentor coaching includes getting formative feedback from a supervisor, consultant, an experienced instructor or a mentor. How it is applied in educational

settings is quite similar to peer evaluation. According to a study by Dewi (2021), mentor coaching is the cause of successful increase in professional competence. A bridge between mentors and mentees are built and they are able to identify and interpret the issues in the classroom, work on them and find some solutions (Sayavedra, 2014).

Teaching Portfolio

A teaching portfolio consists of the works created by instructors and the portfolio serves as a reflection tool. It enables an instructor to evaluate his/her own teaching practices and evaluate the effectiveness of the lessons. A teaching portfolio includes an instructor background section, class description, examinations, a teaching philosophy page, videos of classroom practices, lesson plans, notes, documents related to seminars or trainings, written reflections on classroom experiences and observation notes (Doolittle, 1994). The use of portfolios provides an alternative for instructors to other evaluation methods and it is beneficial for them in terms of professional development. However, as Lengeling (1996) stated, collaboration with a mentor is needed while creating portfolios. They are supposed to be tested a couple of times by an experienced instructor. Guidelines related to portfolio formation must be defined clearly in order to avoid confusion. The method is criticized as it takes a considerable amount of time and requires a lot of energy, but it is an effective method for professional growth of the instructors and the growth of the institution.

Video-Based Observation

For an ongoing instructor evaluation, self-observation with a video recording enables instructors to reflect on in-class teaching experiences (Mercado & Baecher, 2014). Video-based observation requires two important steps, one is collaborating with a mentor and getting professional guidance on the process. The other one is

using video recording to be able to evaluate in-class teaching practices (Linden et al., 2022). In a study conducted by Eröz-Tuğa (2013) in Turkey, it was aimed to help trainees to discover their own strengths and the sides that need development by means of video-recorded courses. It was highlighted in the study that regular feedback by mentors is very important for video-based observation. Even though instructors may feel discouraged and disappointed when seeing themselves on the screen and watching videos may take a lot of time for mentors, it is a beneficial component of instructor evaluations.

As stated above, teaching effectiveness can be measured in different ways. In addition to the procedures above, the evaluation of instructors by students is an important way of measuring teaching effectiveness (Marsh & Hattie, 2002). Also, Miguel et al. (2019) stated that a great many sources were regarded so far to set metrics for teaching effectiveness, but student responses to evaluation forms were focused the most. In spite of disagreements on the use of evaluation by students, it has been used by governments for the measurement of teaching effectiveness for years. Since direct feedback from students who observe the in-class performance of an instructor best is provided through SET (Alok, 2011), instructor evaluations by students continue to keep its primacy.

Student Evaluation of Teaching (SET)

SET is defined as an important quality assurance way used to provide data with the aim of improving courses and guiding instructors (Tennant & Khamis, 2017). Another definition was given by Mace (1997) who stated that SET is a process through which the views of the students reach to the course's instructor and the instructor is able to make changes to the course for the following terms.

In academic settings, SET plays a major role in many different ways. Firstly, valuable feedback is provided to instructors of the course and they can enhance their way of teaching accordingly (Mace, 1997). Additionally, SET is regarded as a process description outcome for research and teaching quality. Also, it plays a vital role for students' course selection, for the selection of the award winners and for institutional review of the curriculum. And lastly, it affects administrators' decisions on promotion, tenure and salary judgments (Beck, 1990; Benton & Cashin, 2014; Kember et al., 2002; Marsh, 1984; Marsch & Roche, 1993; Spooren et al., 2013). When the importance given to SET is regarded, the arguments for and against the use of SET need to be addressed.

Pro-Arguments for the Use of SET

A growing body of research has examined the usefulness of SET so far and there are several researchers adopting a generally positive position on the issue. Murdoch (2000) claimed that plenty of time and effort should be spent on SET in an institution instead of issues like developing curricula or revising teaching materials. He also underlined that it is vital for academic institutions to have an SET system. Theall et al. (2001) also indicated that students' opinions are quite necessary for the development of teaching quality and reported that ratings by students provide a reliable and valid assessment for instructors to improve their teaching performance. Zhao and Gallant (2012) questioned the effectiveness of SET process and focused on the validity and internal consistency reliability of the questionnaires after 11 years of implementation at a university in the USA. 73,500 completed forms by undergraduate students were analyzed. The results of the study supported the construct validity and reliability of the questionnaire. In addition, Renaud and Murray (2005) examined the factorial validity of SET at the University of Western

Ontario by comparing in class teaching behaviors obtained from video type with evaluation scores. A strong relation was found in the study. And, Chen and Hoshower (2003) investigated students' perception of SET and their motivation to participate at a mid-west university. The results showed that if students believe their feedback is significant, they give meaningful feedback.

In literature, there are many other researchers who believe that students are valid judges of teaching quality as well. They underlined some misconceptions about the effectiveness of SET and refuted these misconceptions with some research facts (Aleamoni, 1987, 1999; Feldman, 2007; Kulik 2001; Theall & Franklin, 2007). The most commonly held arguments by these researchers in literature are:

- Students are mature enough to make consistent judgments.
- There is no significant relationship between instructors' research productivity and high evaluation results.
- Being a humorous and friendly instructor does not mean high ratings.
- Alumni can rate instructors accurately.
- Rating instruments are valid and reliable.
- Having large classes does not mean lower ratings,
- There are no differences between instructor ratings made by female and male students.
- Time of the course does not affect evaluation results.
- The course's being elective or mandatory does influence evaluation results.
- Instructors can improve their teaching thanks to SET implementation.
- Single general items are not accurate teaching quality measures.
- Disciplinary differences exist in SET.

- Students do not rate more highly when they receive high grades.
- Instructor rank does not have an influence on evaluation results.
- There is a significant relationship between student status and evaluation results.

Counter Arguments Against the Use of SET

On the other hand, some researchers' viewpoints about the use of SET are negative as they claim that it is a process full of shortcomings. Yunker and Yunker (2003) found a negative correlation between student achievement scores and evaluation results at a regional Mid-western school in Accounting Program after obtaining data through a sample of 283 students. A study by Gonyea et al. (2018) questioned if students' pre- and post- perceptions of the course's instructor are positively correlated with student ratings. With a sample of 408 college students, the study was conducted at a regional institution in the North-eastern United States. The findings revealed that while pre-course perceptions of appearance were negatively correlated with SET ratings, end-of course perceptions are affected by faculty members' appearance. Moreover, Chen et al. (2017) explored whether SET influences instructors and they tend to give higher grades to the students. The correlation between evaluation scores and students' final grades were analyzed in the Fall, 2014 and the Spring, 2015 at a university in Taiwan. The results showed that teaching evaluation scores may cause instructors to give better grades in order to get higher evaluation scores. Moreover, Kalender (2011) aimed to identify the variables that contaminate the validity of SET applied in a higher education institution and analyzed the mean scores of the ratings by 3094 students from 2006 to 2009. The results indicated that course-credits, the number of students, how many sections the courses have, grade levels of the course and mean of end-semester grade of the

students are influential on the students' opinions while evaluating the instructor and the course.

Despite the ongoing controversy in literature today, SET is a valuable process for the improvement of student learning outcomes and teaching performance in addition to critical decisions such as retention, tenure and promotion of academic staff. This is the reason why there has been a lot of research in the field and why a great number of questions have been placed around the potential factors impacting on evaluation results and causing validity issues.

Potential Factors Affecting SET Scores

Biases, background factors that invalidate a good evaluation of teaching quality, exist in literature (Marsh & Overall, 1981). It is critical to consider the existing knowledge of these contaminating factors affecting the evaluation results since they are used in various ways from course selection to administrative decisions. These factors can be grouped under four headings: (1) characteristics related to the administration of evaluations, (2) course characteristics, (3) instructor characteristics and (4) student characteristics.

Characteristics Related to The Administration of Evaluations

It is possible to invalidate a good evaluation process with poor administration procedures. First of all, instructor's being in the classroom during the evaluation might threaten the results and it is recommended that a separate evaluator should be in charge of the evaluation process to protect the anonymity of students (Wachtel, 1998). Anonymity of students is also critical since students have tendency to rate highly unless they remain anonymous (Blunt, 1991; Feldman, 1979). Lastly, Aleamoni (1981) stated that timing of evaluation is of significance since it might

distort evaluation results and Seldin (1989) suggested that evaluation should be carried out in the fourth or sixth week after the academic term begins.

Course Characteristics

Class size, the first course characteristic considered, is defined as the number of students in a course section. It has been regarded as a possible factor that has an influence on the results of evaluation process. Although this belief was not supported by the results of certain studies (Hoyt & Lee, 2002; Lin, 1992; Solomon, 1966), the findings of many other studies in literature indicated that teaching in small classes might be the indicator of high evaluation scores for instructors. In other words, the ratings of the instructors teaching in large classes are lower than the ones teaching in small ones (Centra, 2009; Feldman, 1978, 1984; Kinney & Smith, 1992; Marsh, 2007; Narayanan et al., 2014; Scott, 1977). The reason why students tend to rate more highly in small classes might be that there is more interaction in small classes (Cranton & Smith, 1986). Workload is another course characteristic, which might be correlated with SET scores. There are different findings about the relationship between workload and the SET scores. Centra (2003) reported that if courses are considered either too difficult or too easy, they were rated in a lower way. Marsh & Roche (2000) noted that there is a positive relationship between them as workload might mean more effort and bigger challenge for students. If students spend more time and exert more effort on a course, they might think that they learn more. Also, Marsh (2001) pointed out that if students perceive workload as valuable and useful, they give higher ratings to instructors. However, Beran and Violato (2009) stated that workload might cause negative views about instructors indirectly, so they might get lower ratings. Lastly, a course's being elective or required might have an impact on the scores. Some research has shown that instructors of elective courses receive

higher scores than those of required ones. For example, Peterson et al. (2008) stated that SET scores are higher in elective courses compared to required business courses. Furthermore, in a study by McPherson and Jewell (2007) there is a correlation between low SET scores and the proportion of students taking required courses. McKeachie (1979) also focused on the course requirement factor and indicated that student ratings are lower for required courses. Similarly, Theall and Franklin (2001) stated that if courses are provided in a required way, they tend to receive lower ratings. Similarly, students may give higher ratings to courses outside their major area than to courses within their program if those courses are taken for interest. Wachtel (1998) claimed that the reason of this might be due to higher prior subject interest to elective courses.

Instructor Characteristics

Instructor characteristics consist of instructor rank or teaching experience, research productivity and gender of the instructor. All these have the potential to produce meaningful statistical differences when the evaluation results of instructors are examined in an academic setting. Firstly, teaching experience has been the subject of debate in literature as a validity issue. Not surprisingly, experienced instructors receive higher scores than inexperienced ones. The findings of a study by Kinney and Smith (1992) showed that there is a significant relationship between instructors' rating scores and teaching experience. Also, Braskamp and Ory (1994) pointed out that students give higher rankings to experienced instructors than teaching assistants. So, rank gets important at this point. Secondly, to what extent a faculty member's level of research productivity is related to SET is also another possible validity issue. Research productivity of an instructor is given importance in academic settings since it is thought that it affects teaching quality in a positive way.

However, it was reported that research productivity has little correlation with evaluation scores (Centra, 1993; Feldman, 1987). Marsh and Hattie (2002) also found that the relation between evaluation scores and the number of publications is close to zero. Additionally, a great deal of research has been conducted to see the effect of instructor gender on evaluation scores. Even though some research findings indicated that there is minimal relationship between SET scores and instructor's gender (Feldman, 1993; Zabaleta, 2007), a number of studies revealed that female and male instructors are rated differently even though they have the same teaching performance. For instance, Miles and House (2015) worked on big evaluation data in a Business College and the findings revealed that female instructors get lower SET scores. Also, Martin (2016) carried out a study and it demonstrated that the scores of female instructors were lower in political science courses. Another study was conducted by Miller and Chamberlin (2000) who reported that male instructors are perceived superior when compared to female ones and rated more highly. It is quite difficult for female instructors to gain acceptance in the eyes of students while male instructors are perceived as a figure who deserves more credibility (Merritt, 2008). This gender factor needs to be taken into consideration due to promotion opportunities for women in academia if they get lower ratings from their students just because they are female (Macnell et al., 2015). And lastly, grading leniency factor has the potential to create biased results. That is, students might have tendency to rate higher the instructors who are lenient graders whereas instructor with harsh grading policies might be penalized with low ratings (Griffin 2004). This means that instructors might increase their ratings by giving out inflated grades and they might not have fair grading practices in order to buy higher ratings (Krautmann & Sander, 1999). Bias resulting from grading practices might be regarded under the headings of

both instructor characteristics and student characteristics because the higher grades students expect or get, the higher ratings they assign to their instructors and the higher ratings instructors get, the more inflated grades they tend to give out. Both sides have the potential to affect each other.

Student Characteristics

Gender of the students has been regarded a potential biasing factor in literature. Basow (2000) found that gender factors affect students' perceptions to instructors. Female students tend to give more value to female instructors, so ratings are higher. And, male students tend to give higher ratings to male instructors. Also, there are many other studies whose findings revealed that female students tend to rate female instructors more highly whereas male instructors receive higher ratings from male students (Bachen et al., 1999; Centra & Gaubatz, 2000; Feldman, 1993). In addition to gender factor, to what extent students' grade expectations have an impact on SET scores is another subject of debate in literature. Clayson et al. (2006) pointed out that changes in expected grades are correlated with changes in SET scores in a study conducted with 499 students in business courses. There are also other studies where positive correlations have been found out (Blackhart et al., 2006; Cohen, 1981; Harrison et al., 2004; Isely & Singh, 2005) and these correlations are explained with three hypotheses (Marsh, 1987; Marsh & Roche, 1997). The validity hypothesis asserts that if students learn more, they assign higher scores to their instructors. Grading leniency hypothesis states that instructors are rewarded with high rating scores by their students if they receive higher grades than they deserve. And lastly, prior characteristics (such as prior subject interest, student motivation etc.) hypothesis claims that students having a greater interest to the subject before assigning higher ratings to instructors. Marsh and Dunkin (1997) reported that the

effect of prior characteristics cannot be ignored because it is greater than other potential factors that might have an impact on SET results. So, Wachtel (1998) highlighted that it is critical to control this factor in some courses like English as a large number of students are enrolled.

All in all, there is not a clear definition of effective teaching and it is difficult to measure it since clarifying the determinants is quite challenging. However, there are some evaluation practices used for measurement such as self-assessment, teaching portfolios, video-based observations, peer evaluations, mentor coaching and SET. Among these evaluation practices, SET is the one which keeps its primacy. Yet, the controversy on the usefulness of SET has been going on in literature since there might be potential factors affecting the results obtained at the end of the SET process.

CHAPTER 3: METHOD

Introduction

The purpose of this study is to analyze the dimensions of an SET form used to evaluate the instructors of the preparatory school in a non-profit private university. In this chapter, the methodology of the current study is presented. First, the research design is explained. Then information about the university context and participants is provided. Lastly, instrumentation, the procedure of the data collection and data analysis are clearly defined.

Research Design

This research study was conceptualized as a quantitative approach. Quantitative research design deals with numbers and data collected in the form of scores. Statistical procedures are followed to analyze the data. It mainly focuses on collecting numerical data and generalizing it to explain a specific phenomenon (Creswell, 2009; Fraenkel et al., 2012). Creswell (2009) exemplifies the design by stating that if a research problem deals with the identification of factors that have an impact on findings, the best approach is the quantitative approach. Fowler (2013) also noted that if a questionnaire as an instrument is used to obtain statistical data about certain aspects of participants, it is quantitative in nature.

For the current study, an SET questionnaire designed by the measurement and evaluation unit of the preparatory school in a non-profit private university was used as an instrument, so the data collected are in the form of scores and numerical. The research problems deal with the dimensions of an SET form. Statistical procedures were followed for the analysis. A statistical method, principal component

analysis (PCA), was conducted using a software program to analyze the data in order to find out the underlying constructs in the scale.

Context

This research study was carried out at a preparatory school of a non-profit private university. The language taught in the preparatory school is English. Students are accepted to the programs in the university based on their scores in the university entrance exam. The medium of the education is 30% English (except a few departments). All students, except the ones in a few Turkish-medium departments, are obliged to attend the preparatory program which is made up of three 13-week study periods if they do not submit a valid exemption certificate like TOEFL IBT or if they flunk in TOEFL ITP (Institutional Proficiency Exam) administered at the beginning of the academic year. There are three main language levels called A (beginner/elementary), B (intermediate) and C (upper-intermediate) in preparatory school. Also, the school has D level (upper-intermediate), the students of which are from the department of English Language and Literature. Every one of these levels has repeat classes as well, for the ones who cannot meet the requirements of the level. The students are placed in these levels based on their scores of the preliminary exam and TOEFL ITP exam.

A level consists of two sub-levels named A-foundation (AF) (beginner) and A (elementary). Students who get a score 49 out of 100 or below in the preliminary exam or who do not take TOEFL ITP exam start as AF (beginner) level. The ones who get a score between 50 and 64 are placed in A (elementary) level. If students taking the preliminary exam get 65 (out of 100), they have the right to take TOEFL ITP. In TOEFL ITP, the ones who get 450 or below (out of 677) are placed in B (intermediate) level while those with a score between 450 and 499 study in C (upper-

intermediate) level. And students who get a score of 500 out of 677 in TOEFL ITP exam have the right to continue their undergraduate studies. (The required score for the English Language and Literature Department students, D level, is 550).

In each level, there are three courses, which are referred as *domains* in this study: main course (MC), reading and writing (RW), and listening and speaking (LS). Even though MC follows a program based upon integrated skills (grammar, reading, writing, listening and speaking), the other two courses are skill-based. MC is 15 hours for AF level while it is 10 hours for A, B, C, CR and D levels. RW is 10 hours and LS is 5 hours for all levels. In total, AF level students have 30 study hours each week and A, B, C, CR and D level students have 25.

Each course is taught by a different instructor, which means that each class has three different instructors. There are both native and Turkish instructors in the school and each instructor has to teach at least 20 hours per week (Some instructors who are in charge of administrative duties teach 10 or five hours per week.) and they are expected to work from 9 a.m. to 5 p.m. An instructor might have one class at least and four classes at most. In each course, different books and materials are covered based on the skills. Also, various softcopy and hardcopy materials are provided to the students at two learning centers called *Reading Room* and *Self-access Center*. Students might use these two centers to improve their language skills out of the class.

In the program, the performance of the students is evaluated through quizzes, midterm exams, a final exam, online assignments and portfolio tasks. At the end of the term, A and AF students are expected to get a grade 70 (out of 100), which is the average grade to become eligible to pass to the next level whereas it is 65 (out of

100) for B, C, CR and D level students. If they cannot meet this requirement, they repeat the level.

In the preparatory school, the performance of instructors is evaluated through an evaluation form designed by the measurement and evaluation unit. Evaluation is administered in the 12th week of teaching (before the final exam). A different instructor administers the evaluation and students are asked to anonymously fill out the form having a Likert-type scale to produce useful feedback to their instructors that can inform their instructional performance. The results are analyzed and shared with the instructors in a closed-envelope at the end of the academic term. Then, at the end of the academic year, an appraisal meeting with the administration is scheduled to discuss the evaluation results.

Participants

The data of the current study were obtained from all preparatory level students of 2019 – 2020 academic year Fall term in a non-profit private university. The English level of the students are beginner/elementary (AF and A level), intermediate (B level) and upper intermediate (C, C-repeat, and D level) (defined by the Common European Framework). AF, A, B, C and D level students are non-repeat groups while C-repeat (CR) level includes repeat students from the previous academic year.

In the university where this study was carried out, each student rates three instructors in the evaluation process. There are 1770 forms filled out by the students in total. 477 of these forms belong to A level, which makes up the highest percentage (26.9%). There are 465 forms from AF level, 327 from B level, 384 from C level, 99 from CR level and 18 forms from D level. Additionally, when the number of forms filled out across domains is considered, MC constitutes the highest average with 599

forms. As for RW and LS, 586 forms for RW and 585 forms for LS domain were filled out by the students (see Table 1).

At the end of each term, students are expected to rate their instructors in the university where this study was conducted, but the most crowded term in this university is the Fall term. Therefore, the current study was carried out using the data obtained in the Fall term.

Table 1

Descriptive Statistics Regarding Participants Based on English Language Levels and

Domains

Level	<i>n</i>	%
AF	465	26.3
A	477	26.9
B	327	18.5
C	384	21.7
CR	99	5.6
D	18	1
Domain		
MC	599	33.8
RW	586	33.1
LS	585	33.1
Total	1770	100

Instrumentation

The data for the current study were obtained through an SET form prepared by the measurement and evaluation unit of the preparatory school at a non-profit private university. The form consists of close-ended questions in total (see Table 2) and the language of the form is Turkish. The items are formed as a 5-point Likert scale (5 = *strongly agree*, 4 = *agree*, 3 = *neutral*, 2 = *disagree*, 1 = *strongly disagree*).

There are 22 items in the SET form. The items in the form are related to the instructor's teaching personality, instructional methods, work-ethic and

encouragement for extra-curricular studies. This SET form has been used only in the preparatory school. It has been applied in the school for many years. If feedback is provided by the instructors and deemed worthy of consideration, the items are discussed and revised in the weekly committee meetings in the Department of Foreign Languages.

An optical form is distributed with the SET form for the rating process. Each column in the optical form is for an instructor who will be rated. The first column is for the evaluation of MC instructor, the second column is for the evaluation of the RW instructor and the third column is to rate the LS instructor. Each letter in the form stands for a statement in 5-point Likert scale. Option A represents “strongly agree”, option B represents “agree”, option C represents “neutral”, option D represents “disagree” and option E represents “strongly disagree”.

Table 2

Items of the SET Form

#	Item
1	S/he is positive, lively and enthusiastic in the lesson.
2	You can easily communicate with her/him.
3	S/he has been supportive and encouraging in my learning process.
4	S/he manages the class well.
5	S/he ensures that the lessons are covered effectively and efficiently.
6	S/he knows the subject of the lesson well.
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.
8	S/he can be reached during the specified office hours.
9	S/he encourages my active participation in the lesson.
10	S/he includes pair and group work in the lesson.
11	The activities in the lessons vary.
12	S/he gets in and out of classes on time.
13	S/he comes to classes planned and prepared.
14	S/he uses the lesson time effectively till the end.
15	S/he teaches the lesson in English (except TOEFL lessons).
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).

Table 2 (cont'd)*Items of the SET Form*

20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).
22	I am glad that I took lessons from this instructor.

Note. Items of the form above were translated from its original language (Turkish) to English by the researcher.

Method of Data Collection

The data required for this research study were collected in the preparatory school of a non-profit private university. At the university, there are 13 weeks in each term and these data were collected on in the first contact hour (morning) on Thursday in the 12th week. Instructors, who were not the class's own instructors, took the evaluation forms and optical forms from the measurement and evaluation unit that morning and distributed the forms to the students in the classroom which they had been assigned to. They gave the necessary instructions on how to fill out the optical forms and the students were asked to complete the evaluation process within 15 minutes for three instructors. 70 instructors were rated in the term when this study was conducted. As it is seen in Table 3, 30 AF, 34 A, 21 B, 21 C and 9 CR level classes participated in the evaluation process. One D level class also participated in the process; however, the data were not analyzed due to the low number of participation rate. There was an average of 16 students who participated in the evaluation in AF level, 14 in B level, 16 in C level, and 11 in CR level. Regarding the domains, students from 40 MC, 41 RW and 40 LS classes rated the instructors. An average of 15 students from each domain filled out the form.

Table 3*The Number of Students across English Language Levels and Domains*

Levels	<i>n</i>	<i>M</i>	<i>SD</i>
AF	30	15.50	2.10
A	34	14.03	1.85
B	21	15.57	2.62
C	21	16.00	3.02
CR	9	11.00	1.50
Domains			
MC	40	14.65	2.97
RW	41	14.61	2.95
LS	40	14.63	3.01

Since each class had three different instructors, students anonymously rated three different instructors on the same optical form by coding on three different columns. Then, evaluation forms and the optical forms were collected and put in a closed envelope and taken to the unit. After the scores were obtained using optical mark recognition scanner, the results were analyzed on a software program.

In order to reach the data of the current study, a permission form by the academic committee of Department of the Foreign Languages where this research study was conducted was filled out by the researcher providing information on the research study and sent via e-mail to the committee members. After the situation was discussed in the weekly committee meeting of Department of the Foreign Languages, the permission was granted. The data of 2019 – 2020 Fall academic term were shared by the coordinator of measurement and evaluation unit after all the names of the instructors were deleted and the numbers were used instead.

Method of Data Analysis

The current study investigated the dimensions of an SET form designed to measure the instructional performance of the instructors teaching English in a non-profit private university. The data collected are in the form of scores and consist of

the participants' responses to the items in the SET form using a Likert scale. For all the steps in data analysis process, a software program, IBM Statistics 20, was used.

Firstly, Cronbach alpha value of the SET form was calculated on the software program mentioned above to provide an idea on reliability. The rule of thumb states that a value of .70 to .80 is an acceptable value for Cronbach's α , and higher is better (Field, 2013). Cronbach alpha value of the SET form is .93, which indicates that the scale is highly reliable.

Secondly, descriptive statistics of the quantitative data were checked in order to present the basic features and visible characteristics of the data in a simplified, meaningful way. In the analysis process, means, medians, maximum and minimum values and standard deviations were examined and interpreted.

Then, the goal of this research project was focused to decide on the statistical method for the analysis. Principal component analysis (PCA) was employed after checking the assumptions for the whole sample, for each English language level and for three domains individually. The aim was not to find out the latent constructs regarding measured variables, the goal was to reduce the number of dimensions in the form and get more interpretable summaries of variables by retaining as much information from the original measured variables as possible. As stated by Jolliffe (2002) what PCA aims is to reduce the dimensionality of a large dataset in which there are a number of interrelated variables. As it is difficult to interpret large datasets by minimizing the information at the same time, original data are transformed into the principle components (new sets of variables) in order to create new uncorrelated variables that maximize variance. So, PCA was considered to be the most appropriate statistical method for the current study.

In order to confirm that the data can be measured through PCA, it was necessary to check certain assumptions through relevant statistical analyses. Since PCA is a method based on Pearson correlation coefficients, linear relationships between all variables are assumed. So, linearity was checked first. The second assumption is on the adequacy of sampling. For reliable analysis results, a big sample size is required. For this purpose, The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy test was run. To be able to claim that the sample size is adequate, the rule of thumb stating that the ideal KMO value is supposed to be .90 or above (Kaiser, 1960) was taken into consideration. Thirdly, the data are assumed to be suitable for reduction. A good correlation of the items with each other must be observed so that the variance can be partitioned; as a result, Bartlett's test of sphericity was conducted to check this assumption. The result was regarded significant considering the criteria $p < .05$ (Hair et al., 2014). And lastly, outliers were checked. No significant outliers were observed.

Next, determining the number of factors and factor rotation were addressed. The best-known approach used to decide on the number of factors is Kaiser criterion or eigenvalue-greater-than-1.00 rule (Kaiser, 1960), which is very common due to being the default method in frequently used statistical software (e.g. SPSS), being easy to implement and simple to use. So, this approach was followed and components with eigenvalues higher than 1.00 were retained for the decision on the number of the dimensions. However, since there are contradictory views on the use of this approach in the literature, the process was supported with another component determining approach named scree test through which the eigenvalues are plotted against the components. With the use of this approach, a curved shape is created and the point at which the curve changes excessively are detected to decide on the

number of dimensions. So, the scree plot visual was created and checked to ensure the number of dimensions with cut-off points. The point where the line starts to change drastically was taken into consideration to find the number of the components (Tabachnick & Fidell, 2013). After the number of the factors were decided for the whole participants, across English language levels and three domains, they were labeled by the researcher.

As for factor rotation, varimax rotation was implemented since it has been considered as the most suitable orthogonal rotation. It is a highly suggested rotation technique to use when the dataset is explored. Also, it is the most frequently used one in psychological research studies (Fabrigar et al., 1999; Ledesma et al., 2015; Yong & Pearce, 2013).

After deciding on the number of the components that need to be retained and labeling them, another PCA was carried out by fitting the number of the components to one this time. Component matrix tables were examined to decide whether the items have strong relationships with one dominant latent factor. The rule of thumb stating that item loadings that are less than .32 are generally viewed as undesirable (Tabachnick & Fidell, 2013) was taken into consideration while looking out these component matrix tables. This analysis was conducted because the ratios between the first and the second components in all the analyses conducted before were high, meaning that the first factor was quite dominant (John et al., 2014). This strong general factor was named as *instructional effectiveness* by analyzing the wording and the content of the items by the researcher. Lastly, the results were reported and interpreted.

CHAPTER 4: RESULTS

Introduction

This study aims to analyze the factorial structure of an SET form designed to measure the instructional performance of the instructors teaching English in a non-profit private university. In this respect, the following research question was formed:

1. What are the dimensions of the SET form designed in a non-profit private university to measure instructional performance?
 - a. What are the dimensions of the SET form for all the participants in the preparatory school?
 - b. What are the dimensions of the SET form in the preparatory school across English language proficiency levels of students?
 - c. What are the dimensions of the SET form in the preparatory school across three domains (MC, RW and LS)?

In this chapter, the findings of the research questions are presented. Firstly, the results of descriptive analyses regarding the participants and the participants' answers based on the items on the scale are provided. Secondly, PCA results are reported.

Descriptive Statistics Results

In this part, the results of descriptive data analyses are presented based on the Likert-scale type items in the questionnaire. Descriptive statistics includes means, medians, maximum and minimum values, and standard deviations for the current study.

In Table 4, descriptive statistics based on English language levels are presented. The mean value of students' ratings is 4.31 in AF level, 4.51 in A level, 4.43 in B level, 4.50 in C level and 4.38 in CR level, which indicates that the highest average ($M = 4.51$) of students' ratings is in A level while the lowest average ($M = 4.38$) belongs to CR level, with no large difference across levels. When minimum and maximum values are regarded, Table 4 indicates that minimum evaluation score is in C level while the maximum score is in A level.

Table 4

Descriptive Statistics Based on English Language Levels

Levels	<i>M</i>	<i>Min.</i>	<i>Max.</i>	<i>Median</i>
AF	4.31	3.48	4.78	4.38
A	4.51	3.67	4.93	4.56
B	4.43	3.60	4.83	4.43
C	4.50	3.31	4.89	4.58
CR	4.38	3.48	4.91	4.45

Descriptive statistics based on domains are presented in Table 5. The average of the scores is 4.45 for MC, 4.49 for RW and 4.37 for LS. Across domains, mean ratings seem similar. The lowest average ($M = 4.37$) belongs to LS whereas the highest one ($M = 4.49$) is observed in RW. The lowest and the highest scores ($Min. = 3.31$, $Max. = 4.93$) are in MC domain when the maximum and minimum values are checked.

Table 5

Descriptive Statistics Based on Domains

Domains	<i>M</i>	<i>Min.</i>	<i>Max.</i>	<i>Median</i>
MC	4.45	3.31	4.93	4.51
RW	4.49	3.48	4.89	4.57
LS	4.37	3.48	4.85	4.46

Table 6 displays descriptive statistics based on the instructors. According to Table 6, the mean value is 4.45 among 70 instructors, which indicates students rated highly. The minimum score assigned to an instructor by the students is 3.31 whereas the maximum score is 4.89.

Table 6

Descriptive Statistics Based on Instructors

	<i>M</i>	<i>Min.</i>	<i>Max.</i>	<i>Median</i>
Instructors	4.45	3.31	4.89	4.51

Descriptive statistics based on item averages across domains are provided in Table 7, In MC domain, the highest average ($M = 4.77$) belongs to item 8 while item 19 has the lowest one ($M = 3.93$). In RW domain, item 13 has ($M = 4.90$) the highest average score and the lowest average score is observed in item 17 ($M = 3.95$). And lastly in LS domain, whereas item 12 has the highest average ($M = 4.78$), item 17 has the lowest one ($M = 3.79$). When the means of items in all domains are examined in order (MC, RW, LS), it is observed that item 12 ($M = 4.75$, $M = 4.77$, $M = 4.78$) and 13 ($M = 4.75$, $M = 4.90$, $M = 4.70$) have a higher average compared to other items among all domains while item 17 ($M = 3.95$, $M = 3.95$, $M = 3.79$) and 19 ($M = 3.93$, $M = 4.05$, $M = 4.83$) have a lower average. Figure 1 also displays these statistics for a more visualized presentation.

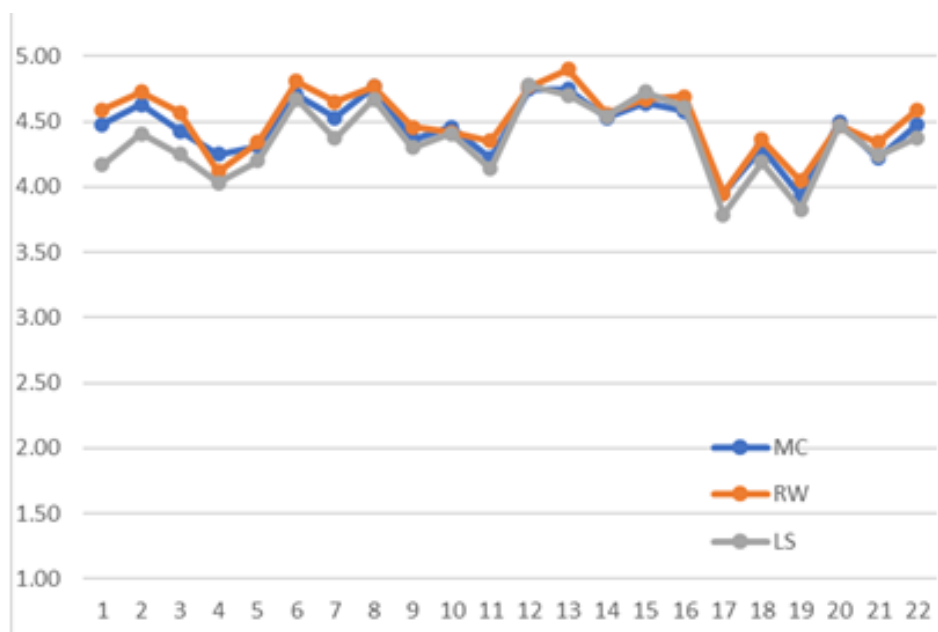
Table 7

Means of Items across Domains

Items	Domains		
	MC	RW	LS
1	4.48	4.59	4.17
2	4.63	4.73	4.41
3	4.43	4.57	4.25

Table 7 (cont'd)*Means of Items across Domains*

Items	Domains		
	MC	RW	LS
4	4.25	4.12	4.03
5	4.31	4.34	4.20
6	4.71	4.81	4.67
7	4.53	4.65	4.37
8	4.77	4.77	4.67
9	4.36	4.46	4.30
10	4.46	4.42	4.41
11	4.22	4.35	4.14
12	4.75	4.77	4.78
13	4.75	4.90	4.70
14	4.53	4.56	4.54
15	4.64	4.67	4.73
16	4.58	4.69	4.61
17	3.95	3.95	3.79
18	4.30	4.36	4.19
19	3.93	4.05	3.83
20	4.50	4.47	4.47
21	4.22	4.34	4.24
22	4.48	4.59	4.38

Figure 1*Means of Items across Domains*

Descriptive statistics based on item averages across English language levels are given in Table 8. Item 8 has the highest average ($M = 4.68$) in AF level, item 12 ($M = 4.88$) in A level, item 13 ($M = 4.73$) in B level, item 15 ($M = 4.83$) in C level and item 8 ($M = 4.75$) in CR level whereas the lowest average belongs to item 19 ($M = 3.66$) in AF level, item 17 ($M = 3.93$) in A level, item 4 ($M = 4.03$) in B level, item 17 ($M = 4.05$) in C level and item 19 ($M = 3.75$) in CR level. Once the means of items in all English language levels are considered in order (AF, A, B, C, CR), the highest means are seen in item 12 ($M = 4.63, M = 4.88, M = 4.74, M = 4.79, M = 4.72$) and item 13 ($M = 4.65, M = 4.78, M = 4.75, M = 4.82, M = 4.73$) while item 17 ($M = 3.67, M = 3.93, M = 4.04, M = 4.05, M = 3.92$) and 19 ($M = 3.66, M = 3.96, M = 4.11, M = 4.09, M = 3.73$) have the lowest means among all levels. These statistics are also provided in Figure 2 for a more visualized presentation.

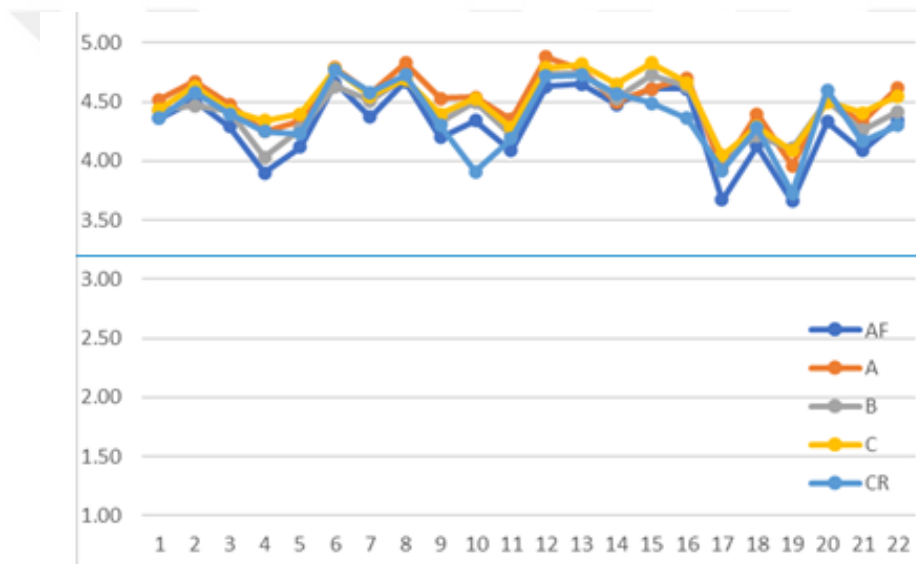
Table 8

Means of Items across English Language Levels

Items	Levels				
	AF	A	B	C	CR
1	4.36	4.52	4.43	4.44	4.36
2	4.51	4.67	4.47	4.63	4.58
3	4.29	4.48	4.42	4.43	4.40
4	3.90	4.25	4.03	4.34	4.25
5	4.12	4.34	4.27	4.40	4.23
6	4.66	4.79	4.63	4.78	4.77
7	4.37	4.57	4.51	4.55	4.58
8	4.67	4.83	4.71	4.69	4.73
9	4.20	4.53	4.33	4.40	4.30
10	4.34	4.54	4.50	4.53	3.91
11	4.09	4.35	4.22	4.28	4.19
12	4.63	4.88	4.74	4.79	4.72
13	4.65	4.78	4.75	4.82	4.73
14	4.48	4.50	4.53	4.65	4.57
15	4.61	4.61	4.73	4.83	4.49
16	4.61	4.70	4.63	4.66	4.36
17	3.67	3.93	4.04	4.05	3.92
18	4.13	4.40	4.21	4.29	4.28

Table 8 (cont'd)*Means of Items across English Language Levels*

Items	Levels				
	AF	A	B	C	CR
19	3.66	3.96	4.11	4.09	3.73
20	4.33	4.52	4.51	4.50	4.60
21	4.09	4.33	4.26	4.41	4.17
22	4.34	4.62	4.42	4.55	4.30

Figure 2*Means of Items across English Language Levels*

Principal Component Analysis (PCA) Results

The Whole Sample

Two statistical test results were checked to confirm that it is appropriate to conduct PCA for the data set: The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and Bartlett's Test of Sphericity. The value is KMO = .94, meaning that the sample size of the current study is adequate. The second statistical test is Bartlett's test of sphericity which demonstrates whether there is an adequate level of correlation between the variables for PCA. Bartlett's test result here was found to be significant with $p < .05$ meaning that there is a significant correlation between the

variables ($\chi = 18291.97$, $df = 231$, $p < .001$) and the data can be partitioned using PCA.

Next, eigenvalues were taken into consideration to determine the number of the components. As indicated in Table 9, there are four components having eigenvalues over suggested value 1.00 with 41.94%, 7.11%, 5.62%, and 4.62% explained variances, respectively. The first four components explain 59.28% of the total variance.

Table 9

The Whole Sample - Total Variance Explained Based on the Initial Eigenvalues

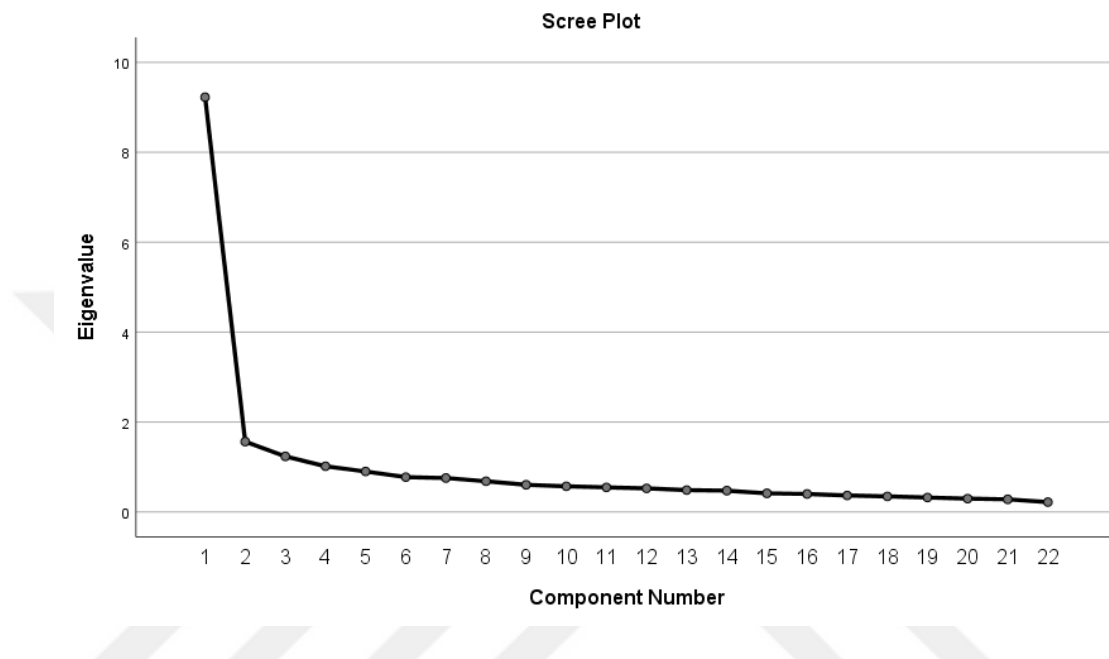
Component	Eigenvalue	% of Variance	Cumulative %
1	9.226	41.935	41.935
2	1.564	7.108	49.043
3	1.236	5.616	54.659
4	1.016	4.618	59.276
5	.900		
6	.775		
7	.754		
8	.682		
9	.604		
10	.571		
11	.547		
12	.525		
13	.486		
14	.473		
15	.414		
16	.400		
17	.367		
18	.346		
19	.320		
20	.297		
21	.280		
22	.218		

Then, the scree plot was checked to confirm the previously concluded number of the components. Figure 3 below demonstrates the scree plot of the whole sample and the ensured number of the components. The line in the plot starts to get

straight after the fourth component, which means that there are four components for the whole participants.

Figure 3

The Whole Sample - Scree Plot



As indicated in Table 10, these four components were named by the researcher. Nine items are grouped under the dimension named as *teaching philosophy and personality*. These nine items explain 41.94% of the total variance. The second component, named as *teaching methodologies*, contains six items. Although it is the second component, only 7.11% of the total variance can be explained with it. There are three items under component 3, labeled as *encouragement for extracurricular activities* and these items account for 5.62% of the variance. Lastly, component 4, which was named as *work ethic and professionalism*, has four items and these four items explain 4.62% of the variance. When the factor loadings of the items were examined, they range from .77 to .53 under component 1, from .63 to .45 under component 2, from .80 to .70 under component 3 and .72 to .50 under component 4.

Table 10*The Whole Sample – Component Labeling*

Components		Factor Loadings	Labels
Component 1			
3	S/he has been supportive and encouraging in my learning process.	.773	Teaching Philosophy and Personality
5	S/he ensures that the lessons are covered effectively and efficiently.	.769	
22	I am glad that I took lessons from this instructor.	.765	
1	S/he is positive, lively and enthusiastic in the lesson.	.752	
4	S/he manages the class well.	.737	
2	You can easily communicate with her/him.	.671	
9	S/he encourages my active participation in the lesson.	.616	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.545	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.530	
Component 2			
10	S/he includes pair and group work in the lesson.	.627	Teaching Methodologies
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.622	
15	S/he teaches the lesson in English (except TOEFL lessons).	.603	
11	The activities in the lessons vary.	.574	
14	S/he uses the lesson time effectively till the end.	.505	
15	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.451	
Component 3			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.805	Encouragement for Extracurricular Activities
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.793	

Table 10 (cont'd)*The Whole Sample – Component Labeling*

Components	Factor Loadings	Labels
Component 4		
12 S/he gets in and out of classes on time.	.718	Work Ethic and Professionalism
13 S/he comes to classes planned and prepared.	.591	
6 S/he knows the subject of the lesson well.	.518	
8 S/he can be reached during the specified office hours.	.500	

Note. Items under each component are listed based on factor loadings.

Although four components were extracted and labeled for the whole sample, the ratio of the first and the second eigenvalues indicates that a unidimensional factorial structure can be considered for the whole sample (see Table 9). The first component was observed to be the dominant and another PCA was conducted by limiting the number of the components to one. Table 11 shows the distribution of the factor loadings for the whole sample obtained through this analysis. As it is seen in Table 11, all factor loadings are higher than .32. Items 5, 22, 3, 21, 9, 4 and 1 have the highest factor loadings with greater than .70. Meanwhile, items 15, 8 and 12 are loaded on the factor with values between .42 and .50, carrying the lowest loadings. However, all the items meet the standard criteria in terms of factor loadings ($>.32$). This means that all the items have strong relationships with the single dominant component labeled as *instructional effectiveness* by analyzing the wording and the content of the items by the researcher.

Table 11*The Whole Sample - Factor Loadings*

Item	Factor Loadings
5	.787
22	.778
3	.768

Table 11 (cont'd)*The Whole Sample - Factor Loadings*

Item	Factor Loadings
21	.747
9	.746
4	.720
1	.705
7	.682
14	.671
11	.670
6	.658
20	.650
18	.638
2	.637
13	.630
19	.575
10	.551
16	.536
17	.519
15	.497
8	.484
12	.423

A Level

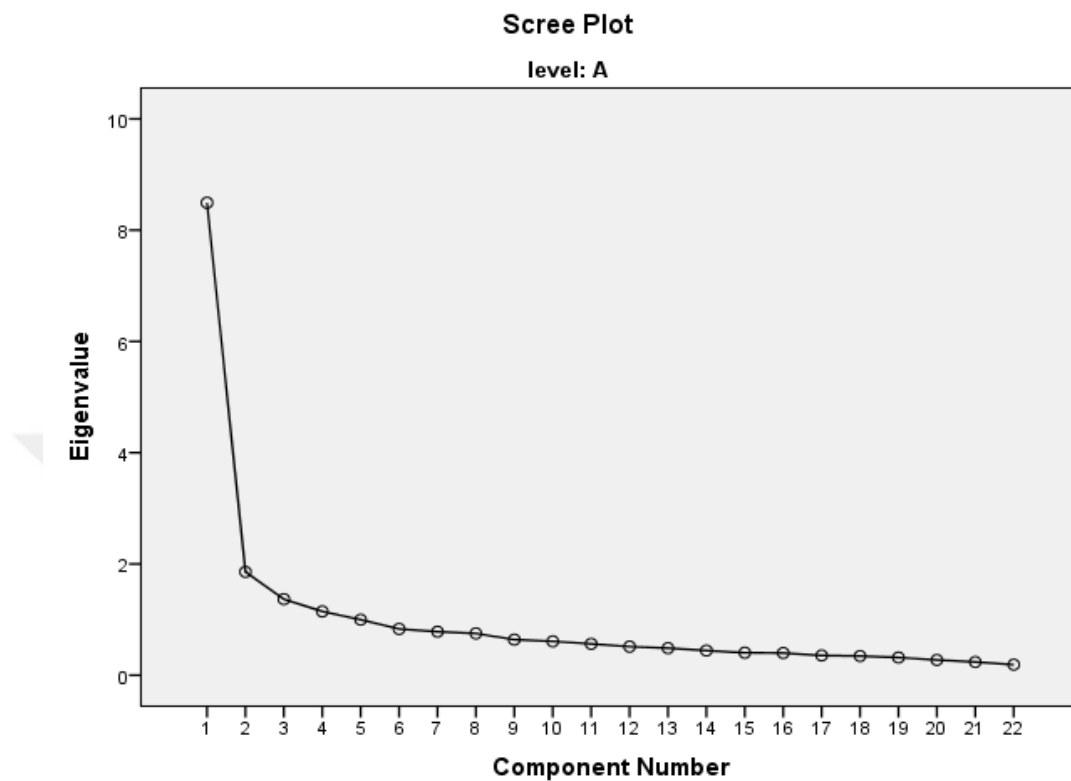
As for A level, the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was carried out first and the value is $KMO = .91$, which indicates that A level's sample size is quite enough to run PCA. Then, Bartlett's test result was found to be significant with $p < .05$, which points out a significant correlation between the variables ($\chi = 4943.882$, $df = 231$, $p < .001$).

Later on, eigenvalues were examined to decide on the number of the components. As Table 12 demonstrates, there are four components having eigenvalues over suggested value 1.00 with 38.6%, 8.44%, 6.21%, and 5.21% variance, respectively. The total amount of the variance explained by these four components is 58.47%.

Table 12*A Level - Total Variance Explained Based on the Initial Eigenvalues*

Component	Eigenvalue	% of Variance	Cumulative %
1	8.493	38.604	38.604
2	1.857	8.442	47.046
3	1.367	6.213	53.259
4	1.146	5.210	58.470
5	.997		
6	.831		
7	.783		
8	.752		
9	.639		
10	.608		
11	.564		
12	.514		
13	.486		
14	.443		
15	.404		
16	.399		
17	.354		
18	.344		
19	.317		
20	.274		
21	.238		
22	.189		

The scree plot for A level was formed to determine the number of components to be retained in PCA (see Figure 4). The plot confirms that there are four components for A level since the line begins to get straight after the fourth component.

Figure 4*A Level - Scree Plot*

As it is pointed out in Table 13, four components extracted in the analysis of A level were named by the researchers. There are 11 items under the dimension which was labeled as *teaching philosophy and personality*. These 11 items explain 38.6% of the total variance. *Encouragement for extracurricular activities* is the label given to the second component, under which five items were grouped. Yet, item 11 and 10 do not seem to fit this specified dimension. Although it is the second component, only 8.44% of the total variance can be explained with it. Four items exist under component 3, labeled as *teaching methodologies* and these items account for 6.21% of the variance. However, under this dimension, item 20 seems incompatible. Lastly, component four, which was named as *work ethic and professionalism*, contains two items which explain 5.21% of the variance. When it comes to the factor loadings of the items, they range from .80 to .47 under

component 1, from .83 to .46 under component 2, from .71 to .56 under component 3 and from .81 to .67 under component 4.

Table 13

A Level – Component Labeling

Components		Factor Loadings	Labels
Component 1			
3	S/he has been supportive and encouraging in my learning process.	.799	Teaching Philosophy and Personality
1	S/he is positive, lively and enthusiastic in the lesson.	.765	
5	S/he ensures that the lessons are covered effectively and efficiently.	.764	
22	I am glad that I took lessons from this instructor.	.756	
4	S/he manages the class well.	.702	
2	You can easily communicate with her/him.	.694	
9	S/he encourages my active participation in the lesson.	.611	
6	S/he knows the subject of the lesson well.	.597	
14	S/he uses the lesson time effectively till the end.	.546	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.533	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.474	
Component 2			
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.829	Encouragement for Extracurricular Activities
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.806	
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.694	
11	The activities in the lessons vary.	.507	
10	S/he includes pair and group work in the lesson.	.463	

Table 13 (cont'd)*A Level – Component Labeling*

Components		Factor Loadings	Labels
Component 3			
15	S/he teaches the lesson in English (except TOEFL lessons).	.711	Teaching Methodologies
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.628	
13	S/he comes to classes planned and prepared.	.590	
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.564	
Component 4			
8	S/he can be reached during the specified office hours.	.809	Work Ethic and Professionalism
12	S/he gets in and out of classes on time.	.667	

Note. Items under each component are listed based on factor loadings.

For A level, the first factor can be considered as the strong dominant factor even though four components were retained. Due to the high ratio between the first and the second components, factors were fitted to one through another PCA. The factor loadings for A level are demonstrated in Table 14 below. Items 5, 4, 3, 22, 9, 14, 1 and 21 carry the highest factor loadings with values greater than .70. Items 8 and 12 have factor loadings lower than .32. However, 20 out of 22 items carry loadings higher than .32 making the one component structure fit to the scale. These 20 items were analyzed in terms of wording and content and it was labeled as *instructional effectiveness*.

Table 14*A Level – Factor Loadings*

Item	Factor Loadings
5	.783
4	.768
3	.760
22	.725

Table 14 (cont'd)*A Level – Factor Loadings*

Item	Factor Loadings
9	.723
14	.718
1	.718
21	.713
6	.671
11	.662
13	.638
2	.610
7	.595
20	.585
18	.560
10	.555
16	.547
15	.502
17	.491
19	.468
8	.304
12	.177

AF Level

Following A level, the factorability of 22 items was examined for AF level.

The analysis started with the use of Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity. The value, KMO = .93, showed that AF level's sample size was adequate for PCA and Bartlett's test of sphericity result was observed to be significant with $p < .05$, meaning that there is substantial correlation in the data ($\chi = 5223.775$, $df = 231$, $p < .001$).

In order to decide how many components to be retained, two criteria were used: the eigenvalues and scree plot for AF level as well. The total variance explained by eigenvalues for AF level is presented in Table 15 below. According to the eigenvalues, there are four components with eigenvalues over suggested value

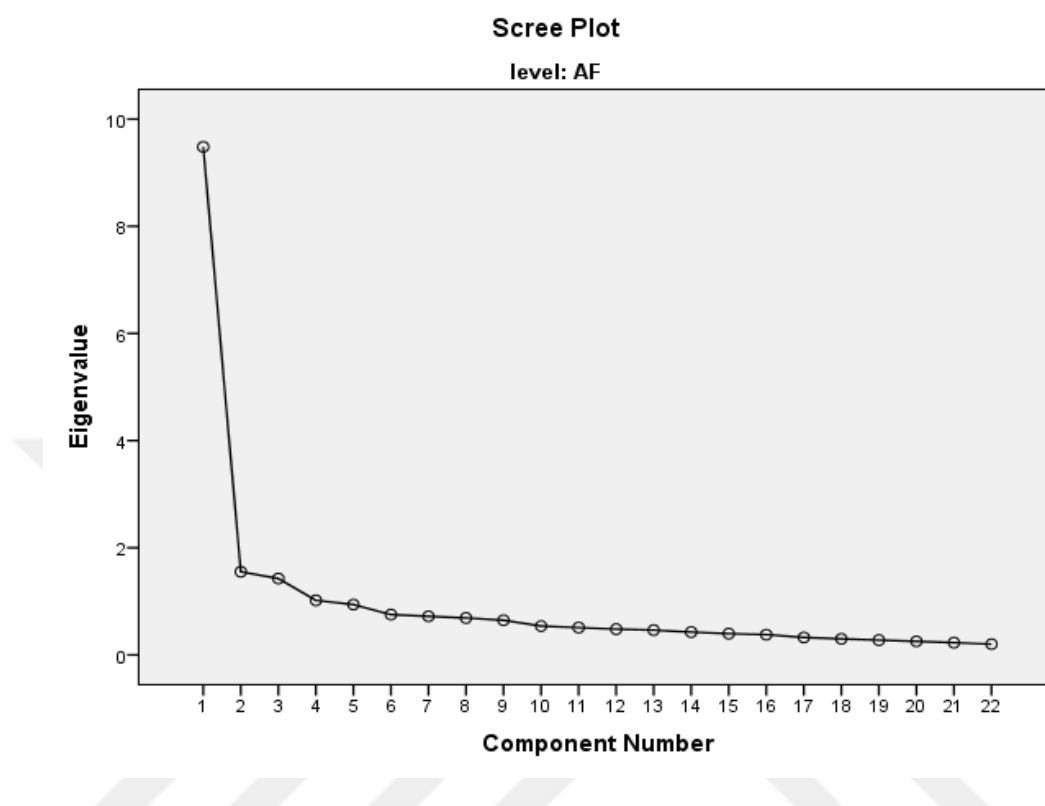
1.00, explaining 43.10%, 7.05%, 6.48% and 4.63% of the variance respectively. The total amount of variance that can be explained by these four components is 61.26%.

Table 15

AF Level - Total Variance Explained Based on the Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	9.481	43.097	43.097
2	1.551	7.051	50.148
3	1.426	6.482	56.631
4	1.018	4.626	61.256
5	.940		
6	.753		
7	.720		
8	.689		
9	.647		
10	.537		
11	.510		
12	.481		
13	.463		
14	.426		
15	.397		
16	.378		
17	.326		
18	.300		
19	.276		
20	.252		
21	.229		
22	.200		

In AF level, the scree plot was also used to determine the number of components based on the size of the eigenvalues and it was provided in Figure 5. The line starts to get straight after the fourth component, which demonstrates that there are four factors observed.

Figure 5*AF Level - Scree Plot*

Four components that were retained for AF level were named by the researcher and presented in Table 16. Eight items were grouped under the dimension labeled as *teaching philosophy and personality*. These eight items account for 43.1% of the total variance and the factor loadings of these eight items range from .80 to .52. The second component was named as *work ethic and professionalism* and six items were grouped under this dimension. Items 15 and 16 were also grouped under it, but they do not seem to fit this specified dimension. Although it is the second component, only 7.05% of the total variance can be explained with it. The factor loadings of the items here range from .73 to .46. Component 3, labeled as *encouragement for extracurricular activities*, has four items. These items account for 6.5% of the variance and the range of the factor loadings is from .83 to .46. Lastly, there are four items grouped under component 4 and it was named as *teaching*

methodologies. Although three of these items seem to fit this dimension, item 20 seems incompatible here. Items under component 4 explain 4.63% of the variance and the factor loadings of these four items range from .77 to .50.

Table 16

AF Level – Component Labeling

Components		Factor Loadings	Labels
Component 1			
22	I am glad that I took lessons from this instructor.	.797	Teaching Philosophy and Personality
3	S/he has been supportive and encouraging in my learning process.	.788	
1	S/he is positive, lively and enthusiastic in the lesson.	.752	
5	S/he ensures that the lessons are covered effectively and efficiently.	.730	
4	S/he manages the class well.	.713	
2	You can easily communicate with her/him.	.660	
9	S/he encourages my active participation in the lesson.	.613	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.515	
Component 2			
13	S/he comes to classes planned and prepared.	.730	Work Ethic and Professionalism
6	S/he knows the subject of the lesson well.	.673	
12	S/he gets in and out of classes on time.	.653	
15	S/he teaches the lesson in English (except TOEFL lessons).	.582	
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.572	
8	S/he can be reached during the specified office hours.	.456	
Component 3			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.828	Encouragement for Extracurricular Activities
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.749	

Table 16 (cont'd)*AF Level – Component Labeling*

Components		Factor Loadings	Labels
Component 3			
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.715	Encouragement for Extracurricular Activities
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.462	
Component 4			
10	S/he includes pair and group work in the lesson.	.770	Teaching Methodologies
11	The activities in the lessons vary.	.611	
14	S/he uses the lesson time effectively till the end.	.527	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.491	

Note. Items under each component are listed based on factor loadings.

Regarding the ratio between the first two components in AF level analysis, it was observed that the first component is the strong general factor and the factorial structure can be considered unidimensional. So, PCA was repeated by fitting the factors into one. Table 17 shows the factor loadings of the items for one factor structure. As it can be clearly seen in Table 17, items 9, 5, 3, 22, 7, 11, 4, 21 and 1 carry the highest loadings with values greater than .70. All the items' factor loadings are higher than .32 which is the rule of thumb making the relationship strong for the whole scale. Only item 12 carries a loading lower than .50. When an in-depth analysis of the items was considered, this sole strong component was labeled as *instructional effectiveness*.

Table 17*AF Level – Factor Loadings*

Item	Factor Loadings
9	.767
5	.760
3	.750
22	.744
7	.735
11	.718
4	.716
21	.716
1	.707
14	.668
6	.668
20	.651
2	.642
13	.631
18	.626
19	.622
15	.597
8	.553
16	.551
17	.518
10	.512
12	.467

B Level

Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity were run to confirm whether B level data are suitable to conduct PCA. The sample size was adequate to conduct PCA when the value (KMO = .90) was looked out. Additionally, Bartlett's test of sphericity result was significant with $p < .05$, indicating that there is significant correlation between variables ($\chi = 3896.468$, $df = 231$, $p < .001$).

In B level analysis, to decide how many components to retain, eigenvalues were taken into account first. According to the results indicated in Table 18, the first five components have eigenvalues greater than 1.00, meaning that there are five

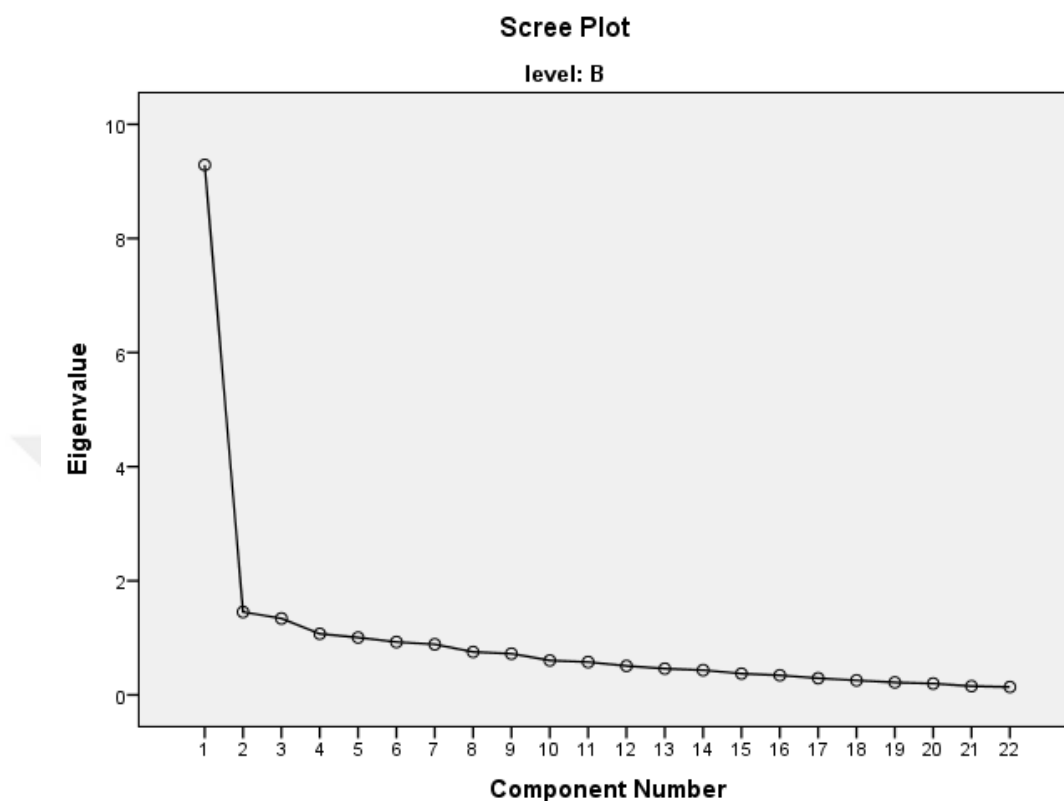
components in the dataset. 64.36% of the variance is accounted for by the first five components.

Table 18

B Level - Total Variance Explained Based on the Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	9.288	42.218	42.218
2	1.453	6.605	48.823
3	1.340	6.092	54.914
4	1.071	4.867	59.782
5	1.006	4.573	64.355
6	.927		
7	.886		
8	.753		
9	.722		
10	.604		
11	.575		
12	.508		
13	.460		
14	.433		
15	.374		
16	.344		
17	.293		
18	.254		
19	.219		
20	.199		
21	.153		
22	.138		

Next, for B level, the scree plot was checked (see Figure 6). The scree plot shows that the eigenvalues start to form a straight line after the fifth component, meaning that five components are supposed to be retained for the analysis.

Figure 6*B Level - Scree Plot*

After deciding the number of the components in B level, these components were named by the researcher. As it is shown in Table 19, there are nine items under the first component named as *teaching philosophy and personality*. 42.22% of the total variance can be explained with these nine items. The factor loadings of the items under component 1 range from .77 to .54. The second component was labeled as *encouragement for extracurricular activities*. Three items exist under this dimension and they explain 6.61% of the variance. The factor loadings of these three items range from .78 to .70. *Teaching methodologies*, the label given to component 3, has three items, which account for 6.09% of the variance and the factor loadings of these three items range from .79 to .56. Under component 4, there are four items grouped and it was named as *guidance*. However, items 2 and 8 do not seem much related to this dimension although items 3 and 7 seem to be associated with

guidance. 4.87% of the variance is accounted for these four items under component 4 and the factor loadings of the items range from .74 to .45. And lastly, component 5 was labeled as *work ethic and professionalism* with three items which explain 4.57% of the variance and these three items' factor loadings range from .70 to .55.

Table 19

B Level – Component Labeling

Components		Factor Loadings	Labels
Component 1			
5	S/he ensures that the lessons are covered effectively and efficiently.	.770	Teaching Philosophy and Personality
22	I am glad that I took lessons from this instructor.	.755	
4	S/he manages the class well.	.729	
1	S/he is positive, lively and enthusiastic in the lesson.	.702	
6	S/he knows the subject of the lesson well.	.637	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.600	
11	The activities in the lessons vary.	.595	
14	S/he uses the lesson time effectively till the end.	.569	
9	S/he encourages my active participation in the lesson.	.542	
Component 2			
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.783	Encouragement for Extracurricular Activities
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.760	
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.695	
Component 3			
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.793	Teaching Methodologies
10	S/he includes pair and group work in the lesson.	.610	
15	S/he teaches the lesson in English (except TOEFL lessons).	.561	

Table 19 (cont'd)*B Level – Component Labeling*

Components		Factor Loadings	Labels
Component 4			
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.737	Guidance
12	You can easily communicate with her/him.	.710	
3	S/he has been supportive and encouraging in my learning process.	.639	
8	S/he can be reached during the specified office hours.	.452	
Component 5			
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.701	Work Ethic and Professionalism
12	S/he gets in and out of classes on time.	.624	
13	S/he comes to classes planned and prepared.	.553	

Note. Items under each component are listed based on factor loadings.

For B level, the largest amount of variance (42.22%) is explained with the first component, which makes it the dominant one. Since the second component accounts for only 6.61% of the variance, the factorial pattern can be considered unidimensional. So, another PCA was conducted by fitting the items into one factor structure. As it can be seen in Table 20 below, all the items carry loadings higher than .32 providing strong relationships between the items and the single dominant component. Items 22, 5 and 21 show the greatest loadings with values higher than .80. Yet, items 15, 8, 12, 16 and 17 carry loadings between .36 and .50. Even the lowest loaded items have a strong relationship with the dominant component. This single dominant component and the items loaded were analyzed and leded the researcher to name it as *instructional effectiveness*.

Table 20*B level – Factor Loadings*

Item	Factor Loadings
22	.847
5	.840
21	.805
4	.782
9	.773
3	.767
14	.679
6	.677
11	.666
1	.664
18	.664
10	.638
20	.625
7	.608
2	.607
13	.598
19	.546
17	.495
16	.486
12	.452
8	.412
15	.357

C Level

As in the data analysis of all other English language levels, in C level, Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy result was focused on first to see whether C level's sample size was enough to run PCA. The value (KMO = .91) points out that C level has an adequate sample size to run PCA. When the Bartlett's test of sphericity result was considered, it was found significant with $p < .05$, underlining the notable correlation between variables ($\chi = 4300.948$, $df = 231$, $p < .001$).

The next step was to determine how many factors to retain for C level. The eigenvalues and scree plot were used for this purpose as well. Considering the

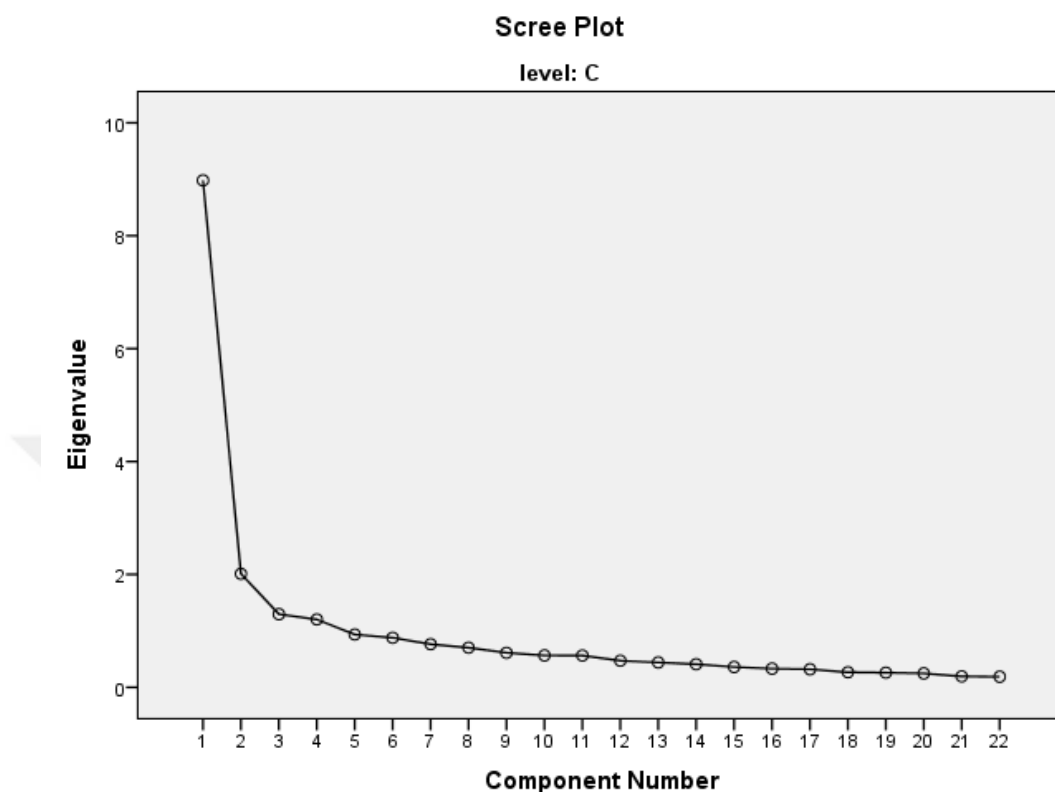
eigenvalues in Table 21, there are four components with eigenvalues over the suggested value 1.00, explaining 40.82%, 9.14%, 5.89% and 5.47% of the variance respectively. These four factors account for 61.32% of the total variance.

Table 21

C Level - Total Variance Explained Based on the Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	8.981	40.823	40.823
2	2.011	9.142	49.965
3	1.296	5.892	55.857
4	1.203	5.468	61.324
5	.935		
6	.878		
7	.764		
8	.703		
9	.612		
10	.564		
11	.563		
12	.471		
13	.440		
14	.410		
15	.360		
16	.330		
17	.320		
18	.268		
19	.259		
20	.247		
21	.194		
22	.188		

Figure 7 below presents the scree plot for C level. It points out that there are four components to be retained for C level since eigenvalues start to form a straight line after the fourth principal component.

Figure 7*C Level - Scree Plot*

Four components were extracted in C level and they were named by the researcher (see Table 22). 12 items were grouped under component 1 named as *teaching philosophy and personality* while there are six items under component 2 labeled as *encouragement for extracurricular activities*. The items in the first component explain 40.82% of the total variance and 9.14% of the variance can be explained with the items in component 2. Six items were grouped under component 2, yet items 10, 11 and 14 do not seem to fit this dimension since they do not seem to be associated with *encouragement for extracurricular activities*. *Teaching methodologies*, the label given to component 3, has three items, which account for 5.89% of the variance. And lastly, only one item is grouped under component 4, which was named as *work ethic and professionalism*. This item explains 5.47% of

the variance. When the factor loadings of the items are regarded, they range from .83 to .48 under component 1, from .80 to .43 under component 2, from .71 to .51 under component 3 and the factor loading of the only item is .75 under component 5.

Table 22

C Level – Component Labeling

Components		Factor Loadings	Labels
Component 1			
5	S/he ensures that the lessons are covered effectively and efficiently.	.831	Teaching Philosophy and Personality
3	S/he has been supportive and encouraging in my learning process.	.783	
4	S/he manages the class well.	.760	
1	S/he is positive, lively and enthusiastic in the lesson.	.753	
22	I am glad that I took lessons from this instructor.	.746	
9	S/he encourages my active participation in the lesson.	.665	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.663	
2	You can easily communicate with her/him.	.663	
6	S/he knows the subject of the lesson well.	.617	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.575	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.574	
8	S/he can be reached during the specified office hours.	.480	
Component 2			
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.800	Encouragement for Extracurricular Activities
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.747	
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.719	

Table 22 (cont'd)*C Level – Component Labeling*

Components		Factor Loadings	Labels
Component 2			
11	The activities in the lessons vary.	.698	Encouragement for Extracurricular Activities
14	S/he uses the lesson time effectively till the end.	.452	
10	S/he includes pair and group work in the lesson.	.434	
Component 3			
15	S/he teaches the lesson in English (except TOEFL lessons).	.705	Teaching Methodologies
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.695	
13	S/he comes to classes planned and prepared.	.513	
Component 4			
12	S/he gets in and out of classes on time.	.746	Work Ethic and Professionalism

Note. Items under each component are listed based on factor loadings.

When the ratio between the eigenvalues of the first and the second components is focused on in C level, a unidimensional structure can be considered since the first component explains a lot larger amount of variance in the dataset when compared with the other ones. That is, the first component seems to be dominant, so PCA was repeated by fitting the items into one factor structure. Table 23 demonstrates the factor loadings for C level in one factor structure. As evident from Table 23, all the items loaded have a value greater than .32 indicating a strong relationship between the items and the dominant component. Items 3, 5, 22, 21, 7, 1 and 9 have the highest factor loadings with values greater than .70 and the factor loading values of the items 12 and 15 are below .50. This single dominant component was labeled as *instructional effectiveness* by the researcher.

Table 23*C Level – Factor Loadings*

Item	Factor Loadings
3	.802
5	.790
22	.778
21	.753
7	.734
1	.713
9	.701
20	.672
18	.645
2	.642
4	.629
14	.620
19	.612
6	.606
13	.604
11	.597
17	.552
16	.548
10	.531
8	.523
15	.403
12	.398

CR Level

The same procedures were followed for CR level to check the suitability of the data to run PCA. Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy and Bartlett's test of sphericity were run. Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy value (KMO = .87) showed that CR level's sample size is adequate for PCA. Also, Bartlett's test of sphericity result was observed to be significant with $p < .05$, meaning variables are strongly correlated with each other ($\chi = 1844.581$, $df = 231$, $p < .001$).

When the number of the components that need to be retained is regarded, eigenvalues and scree plot were examined. As it is seen in Table 24, five components

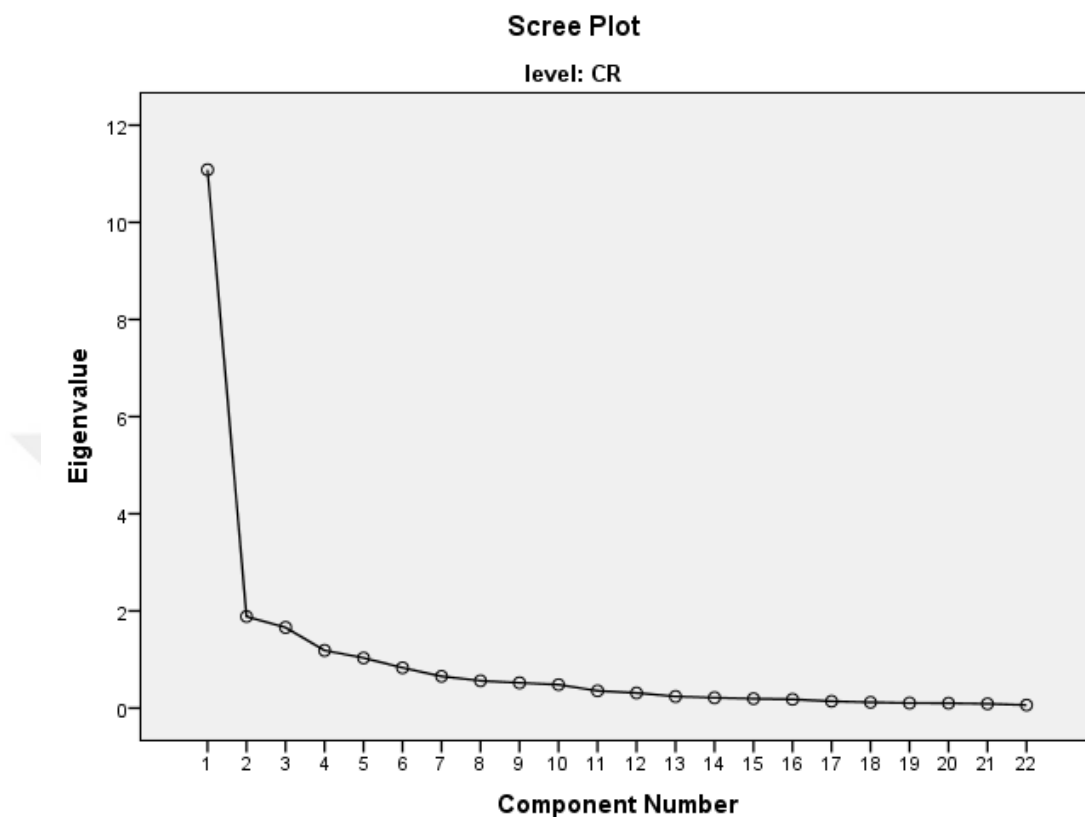
exist for CR level. The first component seems to be highly dominant since it accounts for 50.38% of the total variance. The second one explains 8.57%, the third one 7.55%, the fourth one 5.39% and the fifth one 4.69% of the total variance. 76.58% of the total variance is explained by these five principal components.

Table 24

CR Level - Total Variance Explained Based on the Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	11.083	50.377	50.377
2	1.886	8.571	58.948
3	1.660	7.545	66.493
4	1.186	5.393	71.886
5	1.032	4.691	76.577
6	.831		
7	.654		
8	.563		
9	.520		
10	.482		
11	.354		
12	.312		
13	.239		
14	.213		
15	.194		
16	.180		
17	.140		
18	.118		
19	.104		
20	.100		
21	.088		
22	.062		

After checking the eigenvalues to determine the number of factors, a scree plot was formed (see Figure 8). According to the plot, there are five factors that need to be kept as the line on the plot starts to get straight after the fifth component.

Figure 8*CR Level - Scree Plot*

Five components in CR level were named by the researcher and presented in Table 25. As it is shown in Table 25, seven items were grouped under component 1 and it was labeled as *teaching philosophy and personality*. 50.38% of the variance can be explained with these seven items. Under component 2, labeled as *teaching methodologies*, five items exist. However, among these items, items 1 and 4 do not seem to be related to this dimension and it doesn't make sense to observe them under this component. The items in the second component explain 8.57% of the total variance. *Guidance*, the label given to component 3, has four items, which account for 7.55% of the variance. When the content and the wording of the items under component 3 were considered, items 6 and 7 do not seem to be as relevant to this dimension as the other ones grouped under it. Four items were grouped under

component 4 named as *work ethic and professionalism* and 5.39% of the total variance is accounted for these four items. And lastly, component 5, labeled as *encouragement for extracurricular activities*, contains two items and these items explain 4.69% of the variance. When the factor loadings of the items were regarded, they range from .72 to .49 under component 1, from .80 to .60 under component 2, from .80 to .60 under component 3, from .84 to .60 under component 4 and from .89 to .72 under component 5.

Table 25*CR Level – Component Labeling*

Components		Factor Loadings	Labels
Component 1			
2	You can easily communicate with her/him.	.724	Teaching Philosophy and Personality
3	S/he has been supportive and encouraging in my learning process.	.664	
22	I am glad that I took lessons from this instructor.	.644	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.613	
14	S/he uses the lesson time effectively till the end.	.603	
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.591	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.486	
Component 2			
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.801	Teaching Methodologies
11	The activities in the lessons vary.	.750	
13	S/he comes to classes planned and prepared.	.673	
15	S/he teaches the lesson in English (except TOEFL lessons).	.622	
10	S/he includes pair and group work in the lesson.	.601	

Table 25 (cont'd)*CR Level – Component Labeling*

Components		Factor Loadings	Labels
Component 3			
4	S/he manages the class well.	.800	Guidance
5	S/he ensures that the lessons are covered effectively and efficiently.	.793	
1	S/he is positive, lively and enthusiastic in the lesson.	.629	
9	S/he encourages my active participation in the lesson.	.595	
Component 4			
12	S/he gets in and out of classes on time.	.836	Work Ethic and Professionalism
6	S/he knows the subject of the lesson well.	.797	
8	S/he can be reached during the specified office hours.	.635	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.597	
Component 5			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.889	Encouragement for
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.724	Extracurricular Activities

Note. Items under each component are listed based on factor loadings.

For CR Level, although five components were retained and named, the ratio between the first two components was taken into consideration because the first component was seen to be quite dominant. So, another PCA was conducted by fitting the items into one factor structure. As Table 26 shows, all the items are loaded to this dominant component with values between .51 and .83 proving a strong relationship between the items and the dominant component. 22 out of 22 items meet the standard regarding the factor loading which is higher than .32. The highest loaded items are 22, 20 and 3 with values greater than .80 and the lowest loaded items are 19, 12, 10 and 17 with values between .60 and .51. However, even the lowest ones are greater

than .50. Considering the wording and the content of the items, the single dominant component was labeled as *instructional effectiveness* by the researcher.

Table 26

CR Level - Factor Loadings

Item	Factor Loadings
22	.829
20	.821
3	.813
18	.785
2	.767
21	.766
14	.757
5	.757
9	.747
1	.741
7	.735
13	.722
15	.713
6	.706
11	.675
8	.655
4	.652
16	.613
19	.598
12	.583
10	.549
17	.513

MC Domain

The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was carried out for MC domain first and the value KMO = .94 shows that the sample size for MC is adequate to run PCA. Then, a significant correlation between the variables was observed with $p < .05$ through Bartlett's test result ($\chi^2 = 6591.695$, $df = 231$, $p < .001$).

Later on, the eigenvalues were taken into account with the aim of determining the number of the components. As it is seen in Table 27, four components have

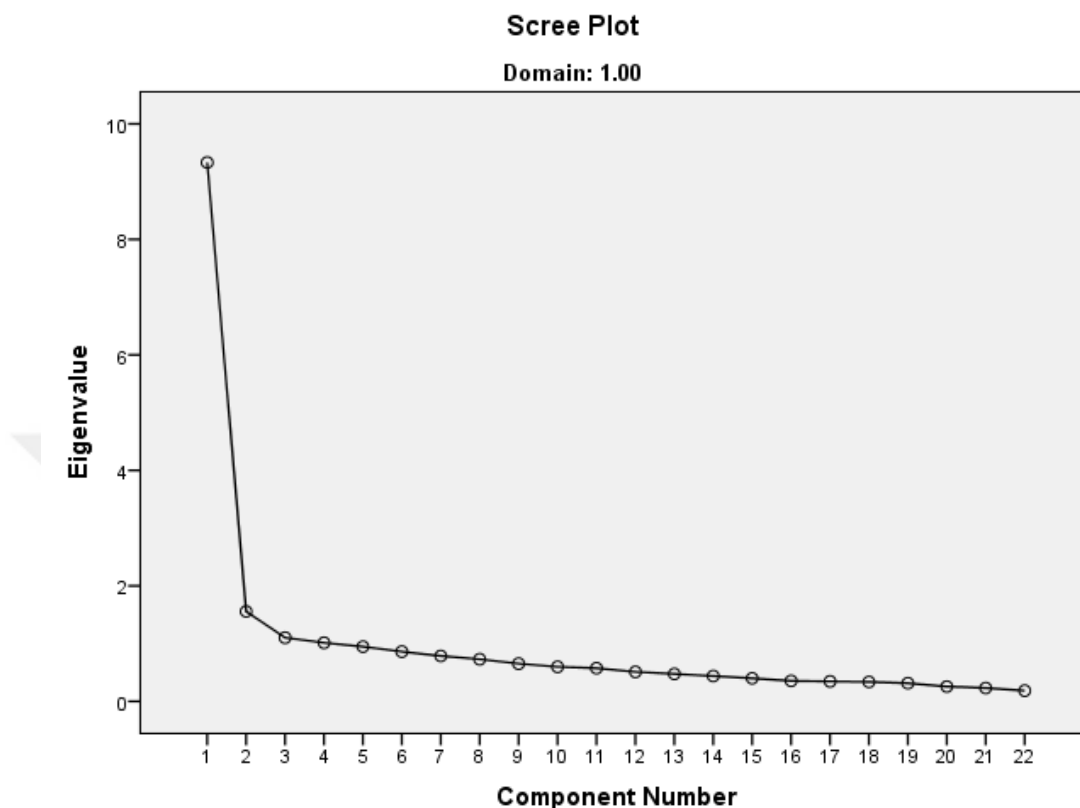
eigenvalues over the suggested value, which means that there are four components in MC domain. 59.12% of the variance in the dataset is explained by these four components. The first component accounts for a large portion of the total variance with 42.42%, proving that the unidimensional factorial structure can be considered for the dataset.

Table 27

MC Domain - Total Variance Explained Based on the Initial Eigenvalues

Component	Eigenvalue	% of Variance	Cumulative %
1	9.331	42.415	42.415
2	1.557	7.077	49.492
3	1.102	5.011	54.503
4	1.015	4.614	59.117
5	.947		
6	.861		
7	.785		
8	.730		
9	.652		
10	.598		
11	.575		
12	.511		
13	.475		
14	.439		
15	.399		
16	.355		
17	.346		
18	.336		
19	.314		
20	.257		
21	.231		
22	.185		

A scree plot was formed to decide the number of the components for MC domain. Since the curve begins to flatten at the fourth component line on the plot, there are four components to be retained for the analysis, which can be seen in Figure 9 below.

Figure 9*MC Domain - Scree Plot*

After four components were retained for MC domain, they were labeled. As it is indicated in Table 28, 11 items were grouped under component 1 and 42.42% of the variance is accounted for these items. Component 1 was labeled as *teaching philosophy and personality* by the researcher regarding the content and the wording of these items. The factor loadings under this dimension range from .82 to .52. Under component 2, which was labeled as *teaching methodologies*, there are five items and these items explain 7.08% of the variance. The factor loadings under this component range from .73 to .42. Three items exist under component 3 named as *encouragement for extracurricular activities* and 5.01% of the variance is explained with these three items. The range of the factor loadings under this component is from .81 to .63. Lastly, component 4 contains three items which explain 4.61% of the variance and

this component was named as *work ethic and professionalism*. The factor loadings of the items range from .74 to .43 under component 4.

Table 28

MC Domain – Component Labeling

Components		Factor Loadings	Labels
Component 1			
5	S/he ensures that the lessons are covered effectively and efficiently.	.823	Teaching Philosophy and Personality
22	I am glad that I took lessons from this instructor.	.801	
3	S/he has been supportive and encouraging in my learning process.	.795	
1	S/he is positive, lively and enthusiastic in the lesson.	.757	
4	S/he manages the class well.	.719	
2	You can easily communicate with her/him.	.636	
9	S/he encourages my active participation in the lesson.	.605	
14	S/he uses the lesson time effectively till the end.	.583	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.558	
6	S/he knows the subject of the lesson well.	.555	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.518	
Component 2			
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.732	Teaching Methodologies
10	S/he includes pair and group work in the lesson.	.644	
11	The activities in the lessons vary.	.570	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.480	
15	S/he teaches the lesson in English (except TOEFL lessons).	.422	
Component 3			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.805	Encouragement for Extracurricular Activities

Table 28 (cont'd)*MC Domain – Component Labeling*

Components		Factor Loadings	Labels
Component 3			
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.782	Encouragement for Extracurricular Activities
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.627	
Component 4			
12	S/he gets in and out of classes on time.	.742	Work Ethic and Professionalism
13	S/he comes to classes planned and prepared.	.598	
8	S/he can be reached during the specified office hours.	.428	

Note. Items under each component are listed based on factor loadings.

As for MC domain, the dominance of the first component was observed and the ratio between the first and the second components is high. So, PCA was repeated by fitting the components into one. Table 29 shows the factor loadings of the items for one factor structure. As it can be clearly seen in Table 29, items 5, 22, 3, 21, 1, 9, 7 and 4 carry the highest loadings with values greater than .70. All the items' factor loadings are higher than .32. Only item 12 carries a loading lower than .40. When an in-depth analysis of the items was considered, this sole strong component was labeled as *instructional effectiveness* for MC domain.

Table 29*MC Domain – Factor Loadings*

Item	Factor Loadings
5	.816
22	.814
3	.789
21	.768
1	.754

Table 29 (cont'd)*MC Domain – Factor Loadings*

Item	Factor Loadings
9	.753
7	.717
4	.713
11	.665
14	.660
6	.652
18	.646
2	.644
13	.624
20	.604
19	.581
16	.547
15	.536
17	.501
10	.498
8	.437
12	.363

RW Domain

As for RW domain, the sample size adequacy was tested through The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and the sample size was found adequate with the value KMO = .92 to run PCA. Then, the correlation between the variables was found to be significant with $p < .05$ when the Bartlett's test result ($\chi = 5748.625$, $df = 231$, $p < .001$) was checked.

Next, it was found out that there are four principal components for RW domain as the first four components in Table 30 have eigenvalues over suggested value 1.00. These four components explain %57.44 of the total variance.

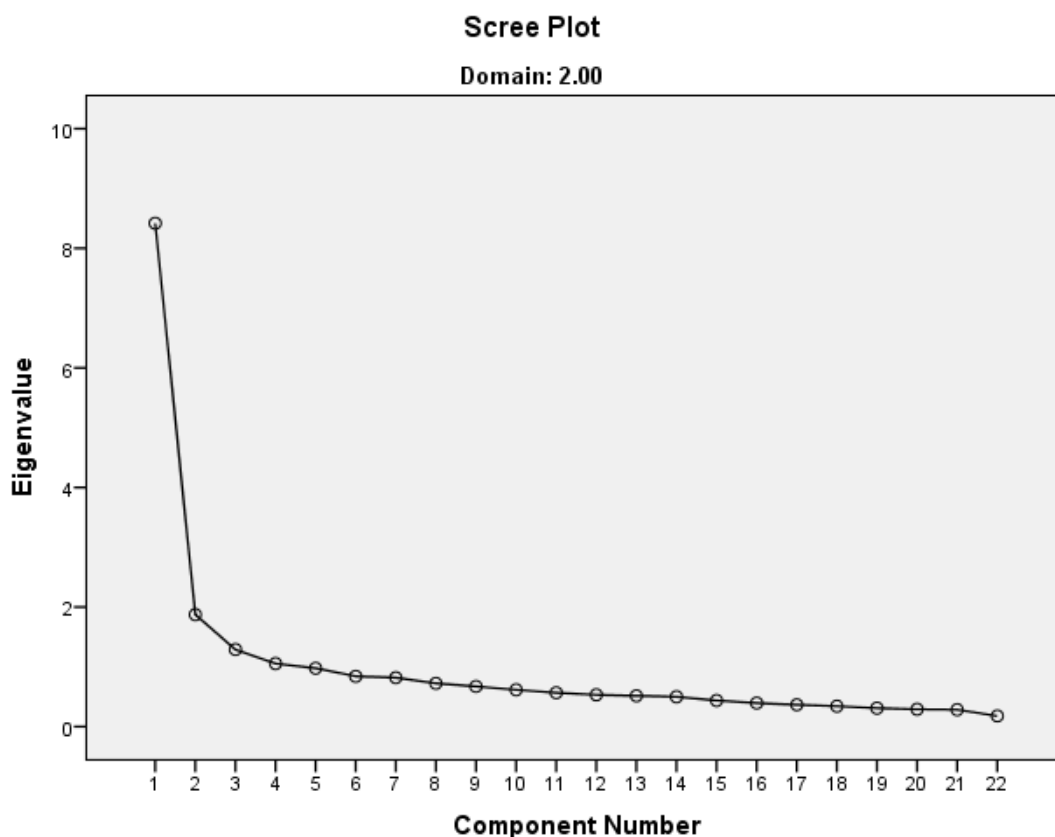
Table 30*RW Domain - Total Variance Explained Based on the Initial Eigenvalues*

Component	Eigenvalue	% of Variance	Cumulative %
1	8.419	38.269	38.269

Table 30 (cont'd)*RW Domain - Total Variance Explained Based on the Initial Eigenvalues*

Component	Eigenvalue	% of Variance	Cumulative %
2	1.873	8.516	46.785
3	1.290	5.864	52.649
4	1.054	4.792	57.441
5	.977		
6	.842		
7	.819		
8	.724		
9	.672		
10	.616		
11	.567		
12	.532		
13	.516		
14	.499		
15	.436		
16	.394		
17	.366		
18	.342		
19	.309		
20	.291		
21	.284		
22	.178		

The confirmation of the number of components was made through a scree plot for RW domain (see figure 10). The plot demonstrates that there are four components to be retained for the analysis since the line starts to get straight after the fourth component.

Figure 10*RW Domain - Scree Plot*

For RW domain, four components extracted were labeled by the researcher. It was presented in Table 31 below. There are 10 items grouping under component 1 and 38.27% of the variance is accounted for these items. The factor loadings of these 10 items range from .78 to .44. This component was named as *teaching philosophy and personality*. *Work ethic and professionalism* is the label given to component 2, which has five items. These five items explain 8.52% of the variance and the factor loadings of them range from .74 to .46. Five items are grouped under component 3 named as *teaching methodologies* and 5.86% of the variance is explained with these three items. The factor loadings under this dimension range from .68 to .47. Lastly, component 4 has three items which explain 4.79% of the variance and this

component was named as *encouragement for extracurricular activities*. The factor loadings of these three items range from .80 to .74.

Table 31

RW Domain – Component Labeling

Components		Factor Loadings	Labels
Component 1			
4	S/he manages the class well.	.784	Teaching Philosophy and Personality
22	I am glad that I took lessons from this instructor.	.761	
5	S/he ensures that the lessons are covered effectively and efficiently.	.760	
1	S/he is positive, lively and enthusiastic in the lesson.	.722	
3	S/he has been supportive and encouraging in my learning process.	.712	
2	You can easily communicate with her/him.	.589	
9	S/he encourages my active participation in the lesson.	.582	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.465	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.441	
Component 2			
13	S/he comes to classes planned and prepared.	.740	Work Ethic and Professionalism
12	S/he gets in and out of classes on time.	.626	
15	S/he teaches the lesson in English (except TOEFL lessons).	.563	
6	S/he knows the subject of the lesson well.	.528	
8	S/he can be reached during the specified office hours.	.459	
Component 3			
11	The activities in the lessons vary.	.683	Teaching Methodologies
10	S/he includes pair and group work in the lesson.	.671	
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.577	
14	S/he uses the lesson time effectively till the end.	.478	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.469	

Table 31 (cont'd)*RW Domain – Component Labeling*

Components		Factor Loadings	Labels
Component 4			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.804	
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.793	Encouragement for Extracurricular Activities
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.736	

Note. Items under each component are listed based on factor loadings.

When the ratio between the first and the second components' eigenvalues is looked out for RW domain, it seems that the first component has dominance. A large portion of the total variance (39.27%) is explained by the first component, which demonstrates that the factorial structure can be considered unidimensional. So, another PCA was conducted by fitting the items into one factor structure. As it can be seen from Table 32 below, all the items carry loadings higher than .32 by indicating strong relationships between the items and the single dominant component. Items 5, 4, 22, 21, 14 and 9 show the greatest loadings with values higher than .70. Items 15, 19, 17 and 12 carry loadings between .458 and .499, which make them the lowest loaded items. However, even the lowest loaded items have a strong relationship with the dominant component. This single dominant component and the items loaded were analyzed and leaded the researcher to label it as *instructional effectiveness* in RW domain.

Table 32*RW Domain – Factor Loadings*

Item	Factor Loadings
5	.788
4	.745
22	.734
21	.720
14	.704
9	.700
3	.693
20	.676
11	.635
7	.630
6	.612
13	.594
1	.591
18	.589
2	.556
10	.526
16	.515
8	.509
15	.499
19	.490
17	.480
12	.458

LS Domain

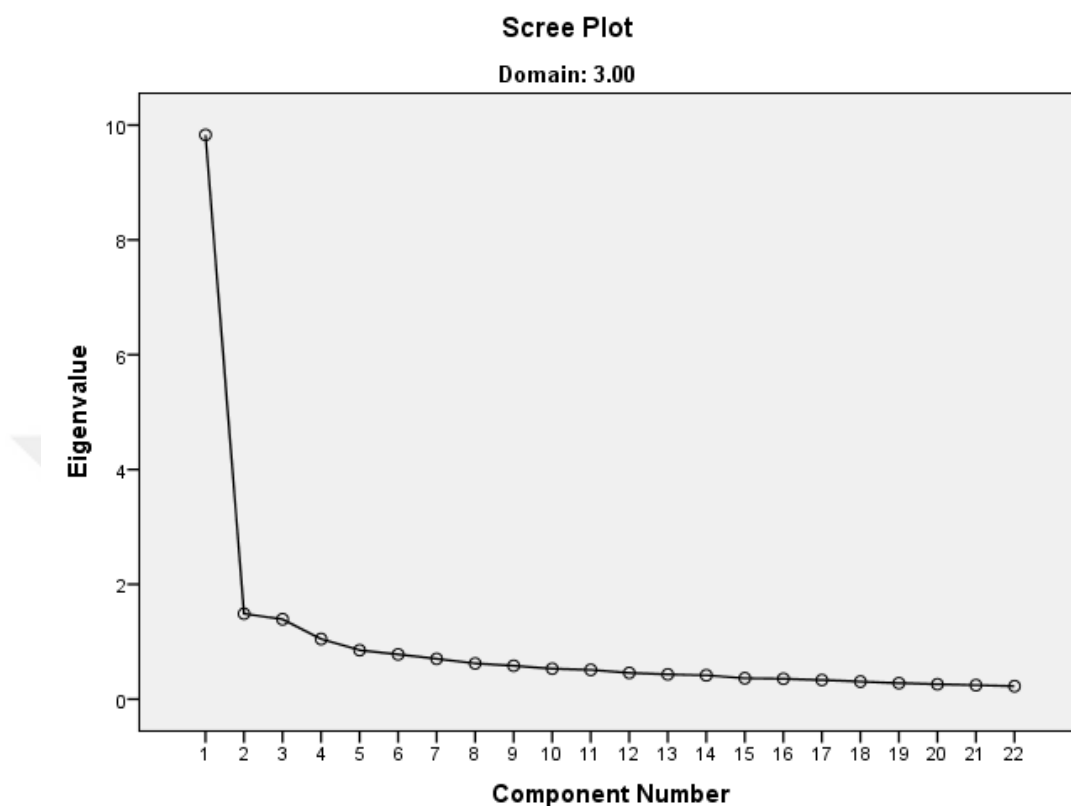
Firstly, The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was conducted for LS domain and the sample size was found to be adequate with the value $KMO = .94$. Secondly, Bartlett's Test of Sphericity result was checked and it was observed that there is a significant correlation between the variables with $p < .05$ ($\chi = 6693.102$, $df = 231$, $p < .001$).

Following the steps above, the eigenvalues were examined to decide on the number of components for LS domain. There are four components for LS domain, which is presented in Table 33 below. These four components account for 62.52% of the overall variance.

Table 33*LS Domain - Total Variance Explained Based on the Initial Eigenvalues*

Component	Eigenvalue	% of Variance	Cumulative %
1	9.830	44.681	44.681
2	1.487	6.761	51.442
3	1.391	6.323	57.764
4	1.047	4.758	62.522
5	.853		
6	.780		
7	.702		
8	.623		
9	.581		
10	.531		
11	.509		
12	.458		
13	.430		
14	.415		
15	.363		
16	.355		
17	.333		
18	.306		
19	.279		
20	.259		
21	.244		
22	.224		

As it can be seen in Figure 11, a scree plot for LS domain was formed. The number of factors was decided through it as well. On the plot, there are four components since the line starts to get straight after the fourth component.

Figure 11*LS Domain - Scree Plot*

Four components retained for LS domain were named by the researcher. As it is indicated in Table 34, the first component was named as *teaching philosophy and personality*, the second one as *teaching methodologies*, the third component as *encouragement for extracurricular activities* and the fourth one as *work ethic and professionalism*. Nine items were grouped under component 1 and 44.68% of the variance is explained with these items. The factor loadings under component one range from .77 to .54. Under component 2, six items exist and these items explain 6.76% of the variance. The factor loadings of these six items range from .75 to .46. Four items, with which 6.32% of the variance is explained, are grouped under component 3. The range of the factor loadings of these items is from .83 to .49. Lastly, component 4 has three items which explain 4.76% of the variance and the factor loadings under this dimension range from .65 to .43.

Table 34*LS Domain – Component Labeling*

Components		Factor Loadings	Labels
Component 1			
5	S/he ensures that the lessons are covered effectively and efficiently.	.770	Teaching Philosophy and Personality
4	S/he manages the class well.	.765	
1	S/he is positive, lively and enthusiastic in the lesson.	.757	
3	S/he has been supportive and encouraging in my learning process.	.757	
22	I am glad that I took lessons from this instructor.	.713	
9	S/he encourages my active participation in the lesson.	.676	
2	You can easily communicate with her/him.	.666	
21	S/he helps me develop my critical thinking skills in various activities (reasoning, analysis, evaluation, etc.).	.589	
7	Whenever I need, s/he recommends studies that may be useful to me and makes the necessary guidance.	.542	
Component 2			
15	S/he teaches the lesson in English (except TOEFL lessons).	.750	Teaching Methodologies
13	S/he comes to classes planned and prepared.	.663	
16	S/he uses visual tools (board, picture, video, projection, internet etc.) effectively.	.582	
10	S/he includes pair and group work in the lesson.	.541	
14	S/he uses the lesson time effectively till the end.	.504	
6	S/he knows the subject of the lesson well.	.455	
Component 3			
17	S/he encourages me to use “Self-Access Center” and “Reading Room”.	.833	Encouragement for Extracurricular Activities
19	S/he encourages participation in extra-curricular activities (drama club, story club etc.).	.801	
18	S/he recommends visual and printed publications (film-series, documentaries, newspapers, magazines, books, online resources, etc.) to support my learning process.	.704	
11	The activities in the lessons vary.	.489	

Table 34 (cont'd)*LS Domain – Component Labeling*

Components		Factor Loadings	Labels
Component 4			
8	S/he can be reached during the specified office hours.	.652	Work Ethic and Professionalism
12	S/he gets in and out of classes on time.	.648	
20	S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.).	.428	

Note. Items under each component are listed based on factor loadings.

The first component captures a large amount of the overall variance (44.68%) in the observed variables in LS domain and the ratio between the eigenvalues of the first and the second components is quite high. So, the factorial structure for LS domain can be considered unidimensional. Another PCA was carried out by fitting the items into one factor structure. As it is indicated in Table 35, all the items are loaded to this single dominant component with the values between .47 and .80 proving a strong relationship between the items and the dominant factor. 22 out of 22 items meet the standard considering the factor loading which is higher than .32. The highest loaded items are 3, 9, 22, 5, 21, 1 and 4 with values greater than .70 and the lowest loaded items are 12 (.47) and 15 (.50). Considering the wording and the content of the items, the single dominant component was labeled as *instructional effectiveness* by the researcher for LS domain.

Table 35*LS Domain – Factor Loadings*

Item	Factor Loadings
3	.795
9	.775
22	.770
5	.769

Table 35 (cont'd)*LS Domain – Factor Loadings*

Item	Factor Loadings
21	.755
1	.729
4	.720
11	.694
20	.693
6	.692
7	.681
2	.673
14	.670
18	.663
13	.656
19	.631
10	.625
17	.559
16	.545
8	.503
15	.498
12	.469

The results of the descriptive statistics and PCA were presented for all the participants, across English language levels and three domains above. Regarding the first PCA results, four components were extracted for the whole participants, English language levels and three domains except for B level and CR level. There are five components in B and CR levels. These components were named by the researcher as *teaching philosophy and personality, teaching methodologies, work ethic and professionalism, encouragement for extracurricular activities and guidance*. Then, it was observed that the ratio of the eigenvalues between the first two components in each analysis was quite high. So, another PCA was conducted by fitting the number of the factors to one and the factorial pattern was considered unidimensional. This single dominant factor was labeled by the researcher as *instructional effectiveness*. The results of the second PCA showed that while certain items have a strong

relationship with the single dominant component, the relationship of some others is weak, which is discussed in the following chapter.



CHAPTER 5: DISCUSSION

Introduction

The current study aimed to analyze the factorial structure of an SET form designed and used by the preparatory school of a non-profit private university. Considering this aim, a quantitative analysis method, Principal Component Analysis (PCA), was utilized to analyze the data obtained. In this chapter, an overview of the study was presented first. Then, major findings of this research study were discussed. Following this, implications for practice and implications for further research were presented. At the end of the chapter, limitations of the present study are provided.

Overview of the Study

The factorial structure of an SET was analyzed for the whole participants ($N = 1770$), across English language levels and for three domains, separately. The following research question with sub-questions guided the current study:

1. What are the dimensions of the SET form designed in a non-profit private university to measure instructional performance?
 - a. What are the dimensions of the SET form for all the participants in the preparatory school?
 - b. What are the dimensions of the SET form in the preparatory school across English language proficiency levels of students?
 - c. What are the dimensions of the SET form in the preparatory school across three domains (MC, RW and LS)?

The data of the study were collected from all preparatory level students of 2019 – 2020 academic year Fall term in a non-profit private university. The language level of the students is beginner/elementary (AF and A level), intermediate (B level)

and upper intermediate (C, C-repeat and D level). The students from AF, A, B, C and CR levels are included in the study; however, D level students are excluded due to the low participation number. At the university, to collect the data from the students, an evaluation form having 22 items with a Likert scale was used. The data were gathered in the first contact hour (morning) on Thursday a week before the last week and a different instructor was assigned to each class to conduct the evaluation process. The students were asked to fill out the SET form within 15 minutes anonymously and rate their three different instructors individually.

The researcher utilized Principal Component Analysis (PCA) through a software program to analyze the data and to interpret the results. Firstly, descriptive statistics of the data were checked to present the basic features and visible characteristics of the data in a simplified way. Then, PCA was run after checking the assumptions for the whole sample, for each English language level and for three domains individually. Next, determining the number of factors and factor rotation were addressed. After deciding on the number of the factors that need to be retained, another PCA was carried out by fitting the number of the factors to one this time, which was necessary to find out whether the items have strong relationships with one dominant latent factor. After that, the findings were reported. Finally, the researcher interpreted the findings, which are provided in the following section.

Discussion of Major Findings

The findings of this study are presented below in response to each sub-research question of the study.

Findings Regarding the Whole Sample

When the results of the study are considered for the whole participants, namely the students, the factorial structure of the form was observed as

unidimensional. That is, all of the items of the scale cohere around a single dominant latent dimension. The ratio between the first and the second factors in the analysis was considered and it was concluded that there is a general dominant factor. This single latent dimension was named as *instructional effectiveness* by the researcher. One dimensional factorial structure of the form indicates that the responses of the students who filled out the form are mostly related to instructional effectiveness and are not affected by other potential factors. The items aiming to measure instructional effectiveness are mostly perceived in the same way by the students.

When the factor loading values for the whole group are considered, items 5 (effective and efficient lesson delivery), 22 (being glad to take the course from this instructor), 3 (supportive and encouraging attitude), 21 (guiding for critical thinking skills improvement), 9 (encouraging for active participation), 4 (managing the class) and 1 (being positive and lively) have the highest values, which indicates that the students mostly associated these items with instructional effectiveness. When the content of all these items are checked, they might gather around the concept of teaching personality of the instructor. The reason why these items have the highest loadings could be explained by the 10 criteria of effective instruction defined by Wotruba and Wright (1975) in literature, which were also provided in Chapter 2. That is, item 5, the most related one to instructional effectiveness, is about covering the lesson effectively and efficiently. This item might be related to the criteria of explaining abstract concepts clearly, clarifying concepts with examples, facts or comparisons, creating a motivating atmosphere, and encouraging students to think for themselves. Item 22 is about being glad to take the class from this instructor, which might be related to all of the criteria provided as it consists of a general concept. Also, this item directly asks whether the students are pleased with the

instruction in class, which is directly associated with instructional effectiveness.

Items 3 and 9 are on the instructor's being supportive and encouraging and it is about the criterion of creating an encouraging atmosphere in class. Item 21 is on the instructor's guidance to improve critical thinking skills and the criterion of encouraging students to think for themselves may be associated with it. Item 4 is about classroom management that may be about criteria of fair evaluation and open communication. Lastly, item 1 is related to being energetic and lively in class, which may be supported by the criterion of enjoying while teaching.

When it comes to the items with low factor loadings, items 12 (getting in and out of classes on time), 8 (availability within office hours) and 15 (teaching in English) do have weaker relationships with the single latent dominant dimension. It means that the students found these items less related to instructional effectiveness. Item 12, the lowest loaded one (.423), is about whether the instructor gets in and out of classes on time. Item 8 is on the instructor's being available during the specified office hours and item 15 is related to using English in class. Item 12 (getting in and out of classes on time) and 8 (availability within office hours) might be perceived as an institutional rule, so the students might not have given enough importance to these items, and they might not have associated them with being an effective instructor. In addition, students may not prefer the instructor to get in and out of class on time because the students may come to school from long distances and be late. Item 15 is on using the target foreign language in class. According to the findings of a study by Öz and Karaazmak (2019), the perceptions of the students on using mother tongue in foreign language classes is positive even though it is not preferable by the instructors in foreign language classes. The reason why item 15 (teaching in English) has a low

loading might be related to this. The majority of the students might not have thought that it is a necessary criterion for instructional effectiveness.

Findings Regarding the English Language Levels

The factorial structure, unidimensionality is desired in most of the instruments where there is a strong dominant factor surrounding all the items in the form (Calderón et al., 2019). The factorial structure of the form for each level was looked out and a unidimensional pattern was observed when checking the ratio of the eigenvalues between the first and the second components. It was concluded by the researcher that there is one strong latent dominant dimension, named as *instructional effectiveness*, and the relationship of the items with it is powerful. Regardless of the students' level of English, most of the items can be gathered around a single dominant dimension. In other words, in all levels, the students understood that the items created to measure instructional effectiveness were perceived as instructional effectiveness related and they responded accordingly.

When each level was analyzed and interpreted individually, considering the factor loading values given in tables in Chapter 4, a similar factorial pattern was observed as evidenced by the same questions have similar factor loadings in AF, A, B and C levels whereas the table of CR level is slightly different. Item 5 (effective and efficient lesson delivery) has the highest factor loading (.783) in A level, it is the second highest loaded factor in AF (.76), B (.84), C (.79) levels. This item is related to the instructor's covering the class effectively and efficiently. Baliyan and Moorad (2018) defined effective and efficient classroom performance with certain dimensions like implementing effective teaching methods, having classroom management skills, assessing fairly, and giving clear feedback. As item 5 (effective and efficient lesson delivery) has a broad framework including many concepts

related to effective instruction in it, it might represent these criteria and most of the students from all levels might have responded to this item with these criteria in their minds. Item 22 (being glad to take the course from this instructor), 3 (supportive and encouraging attitude) and 21 (guiding for critical thinking skills improvement) are the other ones having high loadings in A, AF, B and C levels. This means that all students, regardless of their English level, directly associate the development of critical thinking skills (item 21), being supportive and encouraging (item 3) with instructional performance. Item 22 (being glad to take the course from this instructor) is totally related to instructional effectiveness as the wording of this item indirectly asks the students whether the instruction in class is effective or not. Item 3 (supportive and encouraging attitude) is the other highly loaded one and the reason behind it might be that students value instructor support a lot in their learning process. A study by Barnes and Lock (2013) investigated English language learners' perceptions of effective foreign language teachers. The findings of the study showed that students value concepts like supportive classroom atmosphere or encouraging participation a lot. Item 21 is on the development of critical thinking skills in class. The students perceived this item as associated with instructional effectiveness, the reason of which may be explained by a study conducted by Lailiyah & Wediyantoro (2021). The study investigated the perceptions of the English learners at one of the universities on critical thinking skills. The findings showed that the students have positive attitudes toward the improvement of critical thinking skills in foreign language classes. This might be the reason why most of the students in the current study found item 21 (guiding for critical thinking skills improvement) highly related to instructional effectiveness.

However, the factor loadings table is a bit different in CR level (see Table 20). Item 5 (effective and efficient lesson delivery) is not one of the top five most loaded items in the ranking of the factor loadings of CR level. The highest loaded one is item 22 (.829) (being glad to take the course from this instructor), which is a general statement about being glad to take the course from the instructor and the one which indirectly asks instructional effectiveness. Item 20 (applying the classroom rules) is the second highly loaded one (.821) and it is on the classroom rules. Repeat level students are generally the ones who do not care about the classroom rules, so item 20 (applying the classroom rules) seems meaningless here. As these students are repeat-level students, being demotivated might be an issue to be addressed here. Chen and Hoshower (2003) stated that students' expectations might have an impact on their motivation which is necessary to provide meaningful feedback. Or what they care more may be encouragement and support rather than how the lesson is taught, which might be the reason why item 3 (supportive and encouraging attitude) has a high factor loading (.813) as well.

When the lowest loaded items are regarded for English language levels, it is obvious in all levels that item 12 (getting in and out of classes on time) is the least loaded one. The students might have thought that this item is about being punctual and obeying the institution's rules, not necessary to be an effective instructor. Additionally, the students may not have seen instructor's punctuality as a positive feature considering the possibility of being late for the lesson. While item 12 (getting in and out of classes on time) has the lowest or the second lowest loading in other levels, it ranks the third lowest one in CR level. The lowest one in CR is item 17 (.513), which is about the instructor's encouraging students to use learning centers like "Self-Access Center" and "Reading Room". Considering the highest factor

loadings and the lowest ones, the factor loading order of this level is slightly different from the others. This might be explained by their perception on feedback. According to Chen and Hoshower (2003), students' perception of SET is related to believing that their feedback is valuable. If they do not think so, they show less motivation to participate or give meaningless feedback. Since CR level includes the students from the previous academic year and might have less motivation, they might have thought that nothing changed even though they had shared their views on instructional effectiveness last year. So, they might not have given their full attention while rating. The other items having less factor loadings in all levels are item 8 (availability within office hours), 10 (including pair and group work in the lesson), 15 (teaching in English) and 19 (encouraging participation in extra-curricular activities). Item 8 is about whether they can reach their instructor during the office hour. This item might not have been perceived as related to instructional effectiveness since most of the students may reach their instructors within the working hours all the time as it is a non-profit private university and the instructors are expected to be in their offices from 9 a.m. to 5 p.m. Item 10 is on the instructor's including pair and group work in the lesson and this item has a much lower loading in AF and CR levels when compared to the other ones. The reason behind this may be explained by the findings of a study by Koç (2018). The findings revealed that even though group work and pair work activities are quite beneficial for the learning process, students' tendency to use their mother tongue during these activities make them undesirable. Since AF students have the lowest language proficiency, they tend to communicate with their mother tongue in such activities, which causes them to find these activities useless. CR level students are generally proficiency exam oriented, they might find these activities waste of time and prefer not to use English

during these activities. Item 15, which is on teaching the lesson in English, might again have been perceived differently as indicated above. The students find it difficult to use English in class and do not relate it with effective instruction. Lastly, item 19 is about the instructor's encouraging them for extra-curricular activities. Wong and Leung (2018) stated that students might not prefer to take part in extracurricular activities due to lack of time, not finding them useful and not being interested. The reason why the students participated in the current study find extracurricular activities less related to instructional effectiveness might be the same.

Findings Regarding Three Domains

The structure of the form was analyzed and a unidimensional structure was observed for all domains when the ratio between the eigenvalues of the first two components was considered in each domain. This means that regardless of the domain, the students responded to most of the items by considering a single dimension, which was named as instructional effectiveness by the researcher. However, when the factor loadings of each domain were examined, some differences were detected in the ranking of the highest factor loadings and the lowest factor loadings in LS domain whereas the results seem similar in MC and RW.

In MC and RW domain, item 5 (effective and efficient lesson delivery) is again the one with the highest factor loading and the item is about covering the lesson effectively and efficiently. As also explained above, this item's framework is a pretty broad in terms of instructional effectiveness. In the literature, effective instruction is defined as a concept guiding student to gain certain skills, knowledge, and attitudes (Reiser & Dick, 1996). So, the majority of the students in MC domain might have formed their views regarding this perspective. Item 22 (being glad to take the course from this instructor) is the second highest one in MC and the third highest

one in RW. The item is related to whether the students are glad to take the course from the instructor, which also means that whether they are pleased with the instruction in the class. Thus, the majority of the students might have perceived it as an item that is highly related to instructional effectiveness. When the lowest loaded items were looked out in these two domains, item 12 (getting in and out of classes on time) is the one having the lowest factor loading in MC and RW. It demonstrates that the students in both domains might have thought that the instructor's getting in and out of classes on time is less associated with instructional effectiveness since this item is related to it. Obeying the institution's rules might not be an indicator of the effectiveness of their learning process. While item 8 (availability within the office hours) has the lowest loading in MC domain's factor loadings table (see Table 22), the second lowest loaded item in RW domain's factor loadings table (see Table 24) is item 17. Item 8 is about the availability of the instructor during the specified office hours. As highlighted above as well, the reason behind it could be that the students might reach their instructors whenever they want within the working hours since it is a private university and the instructors are to be in their offices from 9 a.m. to 5 p.m. So, this might have been seen as a criterion which is less related to the instructional effectiveness. Also, most of the students do not prefer to contact with their instructor outside of the classroom, so they might not have responded this item in a meaningful way. Item 17 is on the instructor's encouraging them to use "Self-access Center" and "Reading Room". The students might not have associated instructional effectiveness with being encouraged to use these centers due to some possibilities. These two centers may not be designed to appeal the students' goals, the resources in these centers may not be up-to-date or they may not provide adequate facilities for the

needs of the students. Having a library at the university where the students can reach a large number of resources may also be another possibility.

In LS domain, the ranking of the highest factors differs from that of the other domains. Item 5 (effective and efficient lesson delivery) is not the factor which has the highest loading, the highest loading belongs to item 3 (supportive and encouraging attitude). Rezalou & Yağiz (2021) noted in their study that speaking is a difficult skill to develop and students might feel uncomfortable and reluctant while speaking in a foreign language due to lack of vocabulary, lack of self-confidence and not being interested in the conversation's topic. The reason why this question is at the top of the list could be related to it. Since the students have the most difficulty in speaking skill while learning English, they might expect a lot more support from their instructors and might associate instructional effectiveness with this need. Then, item 9 (encouraging for active participation) comes second in LS domain's factor loadings table (see Table 26). Item 9 is about the instructor's encouraging the students for active in-class participation. It is also about the need for encouragement. The students might have felt that they needed a pushing power to be able to participate in classroom activities which aim to improve their speaking skill. As they were hesitant to use the target foreign language in class, they might have perceived this item as related to the instructional effectiveness. And the last highly loaded item is 22 (being glad to take the course from this instructor) which is the common one in many domains and English language levels. It asks whether the students are pleased to take the course from the instructor, which means that it indirectly asks if the instruction is effective. So, this item might have been substantially associated with instructional effectiveness by most of the students.

When the items with low factor loadings are regarded in LS domain, items 12 (getting in and out of classes on time), 15 (teaching in English) and 8 (availability within the office hours) are the ones which are found to be less related to instructional effectiveness by the students. These items are the ones which have low loadings in almost all English language levels and the other domains. The reasons why they were found to be less related to the instructional effectiveness might be the same. Item 12 is about the instructor's punctuality. The instructor's getting in and out of classes on time was found to be less associated with instructional effectiveness since the institution's rules might not be an indicator of the effectiveness of their learning process. Also, there is the possibility of being late for the students, so they might not have thought that this rule is to their advantage. Item 15 is on teaching the lesson in English. Since the students find it challenging to use English in class, they might not have found it as an aspect of effective instruction. Item 8 is about availability of the instructor during the office hours. Most of the students may contact with their instructors within the working hours all the time as it is a private university and the instructors are in their offices from 9 a.m. to 5 p.m. So, this item might not have been perceived as related to instructional effectiveness.

Implications for Practice

Regarding the results of the current study, certain items in the SET form designed and used in the preparatory school of the university where this study was conducted have low factor loadings, which means that these items are found less related to instructional performance by the students. These items are 12 (getting in and out of the classes on time), 15 (teaching in English), 8 (availability within office hours) and 17 (encouraging for "Self-access center" and "Reading Room"). They are common for the whole participants, across English language levels and across three

domains. These items can be reviewed and revised by the measurement and evaluation unit for a more refined evaluation process. After reviewing and revising the items with low factor loadings in the SET form, the same analysis can be run again and the results can be compared to ensure that all the items in the form measure what it aims to measure.

Different dimensions might form with certain items in the form when the responses of the students are considered, the rationale behind the SET process might be explained to the newly enrolled students.

Other schools and institutions can also benefit from the results of this study. An in-depth information of SET forms across different groups. As this study pointed out, although SET forms show unidimensional factorial structure, items having stronger relationships with the factor (instructional effectiveness) may differ across such groups. Thus, the examination of the factorial structure across subgroups in addition to whole group may provide additional information. Based on such information, instructors and other related parties may hear students' voices better.

Implications for Further Research

Based on the results of the present study, the following implications for further research could be raised:

- This research was conducted with the data obtained from the responses given by the students to the SET form in one academic term. The research can be developed with the data obtained from the whole academic year to be able to reach more reliable judgments.
- This study examined the factorial structure of the SET forms using different grouping variables. A further invariance analysis may provide additional insight about the factorial structure of SET forms.

- This study assumed that instructors' performance was similar across different levels and domains. Since instructors were recruited based on several stages by the institution, this assumption seems plausible.

Limitations

- In this study, there might be dependency among students within the same classes, which might have been resulted in Type I inflation. Due to the data structure, dependency was not controlled in this study in the 2019-2020 academic year Fall term, which limits the results to that specific academic term only.
- The data of the study were collected only from one higher education institution.

REFERENCES

- Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, 2001(109), 59-87.
<https://doi.org/10.1002/ir.4>
- Airasian, P. W., & Gullickson, A. (1994). Teacher self-assessment: Potential and barriers. *Kappa Delta Pi Record*, 31(1), 6-10.
<https://doi.org/10.1080/00228958.1994.10531885>
- Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of Teacher Evaluation* (pp. 110–145). Sage.
- Aleamoni, L. M. (1987). Typical faculty concerns about student evaluation of teaching. *New Directions for Teaching and Learning*, 1987(31), 25-31.
<https://doi.org/10.1002/tl.37219873105>
- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Allan, J., & Clarke, K. (2007). Nurturing supportive learning environments in higher education through the teaching of study skills: To embed or not to embed? *International Journal of Teaching and Learning in Higher Education*, 19(1), 64-76.
- Alok, K. (2011). Student evaluation of teaching: An instrument and a development process. *International Journal of Teaching and Learning in Higher Education*, 23(2), 226-235.
- Andersen, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *PS: Political Science & Politics*, 30(2), 216-219.
- Antony, J., Karamperidis, S., Antony, F., & Cudney, E. A. (2019). Understanding and evaluating teaching effectiveness in the UK higher education sector using

- experimental design: A case study. *International Journal of Quality & Reliability Management*, 36(2), 202-216.
- Aregbeyen, O. (2010). Students' perceptions of effective teaching and effective lecturer characteristics at the University of Ibadan, Nigeria. *Pakistan Journal of Social Sciences*, 7(2), 62-69.
- Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48(3), 193-210.
- Baliyan, S. P., & Moorad, F. R. (2018). Teaching effectiveness in private higher education institutions in Botswana: Analysis of students' perceptions. *International Journal of Higher Education*, 7(3), 143-155. <https://doi.org/10.5430/ijhe.v7n3p143>
- Bangert, A. W. (2006). The development of an instrument for assessing online teaching effectiveness. *Journal of Educational Computing Research*, 35(3), 227-244.
- Barnes, B. D., & Lock, G. (2013). Student perceptions of effective foreign language teachers: A quantitative investigation from a Korean university. *Australian Journal of Teacher Education*, 38(2), 19-36. <http://dx.doi.org/10.14221/ajte.2013v38n2.2>
- Barth, M. M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84(1), 40-46. <https://doi.org/10.3200/JOEB.84.1.40-46>
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles*, 43(5), 407-417. <https://doi.org/10.1023/A:1026655528055>
- Beck, R. L. (1990). Evaluation of teaching/programs. *NACTA Journal*, 7(4), 549-569. <https://doi.org/10.21449/ijate.725701>
- Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In M. Paulsen (Ed.), *Higher education: Handbook of theory and research* (pp. 279-326). Springer.

- Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter?. *Assessment & Evaluation in Higher Education*, 30(6), 593-601. <https://doi.org/10.1080/02602930500260688>
- Beran, T., & Violato, C. (2009). Student ratings of teaching effectiveness: Student engagement and course characteristics. *Canadian Journal of Higher Education*, 39(1), 1-13.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62.
- Blackhart, G. C., Peruche, B. M., DeWall, C. N., & Joiner, T. E. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology*, 33(1), 37-39.
- Blunt, A. (1991). The effects of anonymity and manipulated grades on student ratings of instructors. *Community College Review*, 18(4), 48-54.
<https://doi.org/10.1177/009155219101800409>
- Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. Jossey-Bass Higher and Adult Education Series. Jossey-Bass.
- Canelos, J. (1985). Teaching and course evaluation procedures: A literature review of current research. *Journal of Instructional Psychology*, 12(4), 187.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. Jossey-Bass Higher and Adult Education Series. Jossey-Bass.
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518.
- Centra, J. A. (2009). *Differences in responses to the student instructional report: Is it bias*. Educational Testing Service.

- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71(1), 17-33.
- Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71-88. <https://doi.org/10.1080/02602930301683>
- Chen, C. Y., Wang, S. Y., & Yang, Y. F. (2017). A study of the correlation of the improvement of teaching evaluation scores based on student performance grades. *International Journal of Higher Education*, 6(2), 162-168.
- Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of marketing education*, 31(1), 16-30. <https://doi.org/10.1177/027347530832408>
- Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education*, 5(1), 52-65. <https://doi.org/10.5465/amle.2006.20388384>
- Clayson, D. E., & Haley, D. A. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Review*, 21(2), 101-112.
- Constantinou, C., & Wijnen-Meijer, M. (2022). Student evaluations of teaching and the development of a comprehensive measure of teaching effectiveness for medical schools. *BMC Medical Education*, 22(1), 1-14.
- Cranton, P. A., & Smith, R. A. (1986). A new look at the effect of course characteristics on student ratings of instruction. *American Educational Research Journal*, 23(1), 117-128.
- Creswell, J. W. (2009). *Research design: Qualitative and mixed methods approaches*. Sage Publications.

- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. *Learning Policy Institute*.
[https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/17357/46%20Effective Teacher Professional Development REPORT.pdf](https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/17357/46%20Effective%20Teacher%20Professional%20Development%20REPORT.pdf)
- Dechesne, M. A. (1995). Methods for improving teaching using student feedback techniques. *ASp. la revue du GERAS*, (7-10), 231-240.
<https://doi.org/10.4000/asp.3896>
- Dewi, I. (2021). A mentoring-coaching to improve teacher pedagogic competence: An action research. *Journal of Education, Teaching and Learning*, 6(1), 1-6.
- Ding, C., & Sherman, H. (2006). Teaching effectiveness and student achievement: Examining the relationship. *Educational Research Quarterly*, 29(4), 40-51.
- Doolittle, P. (1994). Teacher portfolio assessment. *Practical Assessment, Research, and Evaluation*, 4(1), 1.
- Eröz-Tuğa, B. (2013). Reflective feedback sessions using video recordings. *ELT Journal*, 67(2), 175-183. <https://doi.org/10.1093/elt/ccs081>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272. <https://doi.org/10.1037/1082-989X.4.3.272>
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9(3), 199-242.
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10(2), 149-172.
- Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, 21(1), 45-116.

- Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in higher education*, 26(3), 227-298.
- Feldman, K. A. (1993). College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34(2), 151-211.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Springer.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage Publications.
- Fowler Jr, F. J. (2013). *Survey research methods*. Sage Publications.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.
- Galbraith, C. S., Merrill, G. B., & Kline, D. M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*, 53(3), 353-374.
- Gonyea, N. E., Osick, M., & Bradley, E. G. (2018). An Evaluation of the relationship between instructor appearance and college student evaluations of teaching. *Journal of Research in Education*, 28(2), 66-92.
- Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, 29(4), 410-425.
- Harrison, P. D., Douglas, D. K., & Burdsal, C. A. (2004). The relative merits of different types of overall evaluations of teaching effectiveness. *Research in Higher Education*, 45(3), 311-323.

- Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. *College Teaching*, 49(1), 26-31.
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1).
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1).
<https://doi.org/10.1080/2331186X.2017.1304016>
- Hoyt, D. P., & Lee, E. J. (2002). Basic data for the revised IDEA system. *IDEA Technical Report*, 12.
- Huston, T., & Weaver, C. L. (2008). Peer coaching: Professional development for experienced faculty. *Innovative Higher Education*, 33(1), 5-20.
- Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *The Journal of Economic Education*, 36(1), 29-42.
- John, M. T., Reißmann, D. R., Feuerstahler, L., Waller, N., Baba, K., Larsson, P., ... & Rener-Sitar, K. (2014). Exploratory factor analysis of the oral health impact profile. *Journal of Oral Rehabilitation*, 41(9), 635-643.
- Jones, A. M., & Airasian, P. W. (1996). Through their eyes: Teacher self-assessment. *Mid-Western Educational Researcher*, 9(1), 12-16.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- Kalender, İ. (2011). Contaminating factors in university students' evaluation of instructors. *Eğitim ve Bilim*, 36(162), 56-65.

- Kalender, İ. (2015). Measurement invariance of student evaluation of teaching across groups defined by course-related variables. *International Online Journal of Educational Sciences*, 7(4), 69-79.
- Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching?. *Assessment & Evaluation in Higher Education*, 27(5), 411-425.
- Kinney, D. P., & Smith, S. P. (1992). Age and teaching performance. *The Journal of Higher Education*, 63(3), 282-302.
- Koç, E. (2018). Exploring collaborative learning with a focus on group activities in EFL classrooms. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 19(3), 582-597.
- Koh, H. C., & Tan, T. M. (1997). Empirical investigation of the factors affecting SET results. *International Journal of Educational Management*, 11, 170-178.
- Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review*, 18(1), 59-63. [https://doi.org/10.1016/S0272-7757\(98\)00004-1](https://doi.org/10.1016/S0272-7757(98)00004-1)
- Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*, 2001(109), 9-25. <https://doi.org/10.1002/ir.1>
- Lailiyah, M., & Wediyantoro, P. L. (2021). Critical thinking in second language learning: students' attitudes and beliefs. *International Journal of Language Education*, 5(3), 180-192.
- Ledesma, R. D., Valero-Mora, P., & Macbeth, G. (2015). The scree test and the number of factors: A dynamic graphics approach. *The Spanish Journal of Psychology*, 18. <https://doi.org/10.1017/sjp.2015.13>
- Lei, S. A. (2010). Classroom physical design influencing student learning and evaluations of college instructors: A review of literature. *Education*, 131(1).

- Lengeling, M. M. (1996). The use of portfolios for teacher evaluation. *Mextesol Journal*, 20(3).
- Lin, W. Y. (1992). Is class size a bias to student ratings of university faculty? *Chinese University of Education Journal*, 20(1), 49-53.
- Linden, S., Meij, J., & McKenney, S. (2022). Teacher video coaching, from design features to student impacts: A systematic literature review. *Review of Educational Research*, 92(1), 114-165.
- Liu, O. L. (2012). Student evaluation of instruction: In the new paradigm of distance education. *Research in Higher Education*, 53(4), 471-486.
- Lueck, T. L., Endres, K. L., & Caplan, R. E. (1993). The interaction effects of gender on teaching evaluations. *The Journalism Educator*, 48(3), 46-54.
- Mace, J. P. (1997). Valuing the individual student: Using single-subject design as a tool for evaluating classroom teaching performance. *Journal of Social Work Education*, 33(2), 361-374.
- Macfadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (2016). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment & Evaluation in Higher Education*, 41(6), 821-839.
- MacNeill, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303.
- McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly*, 88(3), 868-881.
- Marks, R. B. (2000). Determinants of student evaluations of global measures of instructor and course value. *Journal of Marketing Education*, 22(2), 108-119.
- Martin, L. L. (2016). Gender, teaching evaluations, and professional success in political science. *PS: Political Science & Politics*, 49(2), 313-319.

- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754. <https://doi.org/10.1037/0022-0663.76.5.707>
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253-388. [https://doi.org/10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)
- Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on students' evaluations of teaching. *American Educational Research Journal*, 38(1), 183-212. <https://doi.org/10.3102/00028312038001183>
- Marsh, H. W. (2007). *Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness*. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Springer, Dordrecht.
- Marsh, H. W., & Dunkin, M. J. (1997). *Students' evaluations of university teaching: A multidimensional perspective*. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 241-320). New York: Agathon Press.
- Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education*, 73(5), 603-641.
- Marsh, H. W., & Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. *American Educational Research Journal*, 18(1), 103-112. <https://doi.org/10.3102/00028312018001103>

- Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217-251. <https://doi.org/10.2307/1163195>
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187–1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?. *Journal of Educational Psychology*, 92(1), 202-228. <https://doi.org/10.1037/0022-0663.92.1.202>
- McGartland Rubio, D., Berg-Weger, M., & Tebb, S. S. (2001). Using structural equation modeling to test for multidimensionality. *Structural Equation Modeling*, 8(4), 613-626.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 65(6), 384-397. <https://doi.org/10.2307/40248725>
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218–1225. <https://doi.org/10.1037/0003-066X.52.11.1218>
- Mercado, L. A., & Baecher, L. (2014). Video-based self-observation as a component of developmental teacher evaluation. *Global Education Review*, 1(3), 63-77.
- Merritt, D. J. (2008). Bias, the brain, and student evaluations of teaching. *St. John's Law Review*, (82)1, 235-288.
- Miguel, C. V., Moreira, C., Alves, M. A., Campos, J. B. L. M., Glassey, J., Schaer, E., ... & Madeira, L. M. (2019). Developing a framework for assessing teaching effectiveness in higher education. *Education for Chemical Engineers*, 29, 21-28. <https://doi.org/10.1016/j.ece.2019.06.001>

- Miles, P., & House, D. (2015). The tail wagging the dog; An overdue examination of student teaching evaluations. *International Journal of Higher Education*, 4(2), 116-126.
- Miller, J., & Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology*, 283-298.
- Murdoch, G. (2000). Introducing a teacher-supportive evaluation system. *ELT Journal*, 54(1), 54-64.
- Murray, H. G. (2007). Low-inference teaching behaviors and college teaching effectiveness: Recent developments and controversies. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 145–200). Springer.
- Naidoo, R. (2016). Higher education is trapped in a competition fetish. *University World News*, 409(16).
- Narayanan, A., Sawaya III, W. J., & Johnson, M. D. (2014). Analysis of differences in nonteaching factors influencing student evaluation of teaching between engineering and business classrooms. *Decision Sciences Journal of Innovative Education*, 12(3), 233-265.
- Nargundkar, S., & Shrikhande, M. (2014). Norming of student evaluations of instruction: Impact of noninstructional factors. *Decision Sciences Journal of Innovative Education*, 12(1), 55-72.
- Neath, I. (1996). How to improve your teaching evaluations without improving your teaching. *Psychological Reports*, 78(3), 1363-1372.
<https://doi.org/10.2466/pr0.1996.78.3c.136>

- Orsini, J. L. (1988). Halo effects in student evaluations of faculty: A case application. *Journal of Marketing Education*, 10(2), 38-45.
<https://doi.org/10.1177/027347538801000208>
- Ory, J. C. (2000). Teaching evaluation: Past, present, and future. *New Directions for Teaching and Learning*, 83, 13-18.
- Oxford University Press. (n.d.). Bias. In *Oxford learner's dictionaries*. Retrieved May 10, 2022, from <https://www.oxfordlearnersdictionaries.com/definition/english/bias>
- Öz, H., & Karaazmak, F. (2019). L2 learners' perceptions of using L1 in EFL classrooms. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, (42), 213-222.
<https://doi.org/10.21497/sefad.675180>
- Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8(3), 399-411.
- Pepe, J., & Wang, M. (2012). What instructor qualities do students reward? *College Student Journal*, 46(3), 603-614.
- Peterson, R. L., Berenson, M. L., Misra, R. B., & Radosevich, D. J. (2008). An evaluation of factors regarding students' assessment of faculty in a business school. *Decision Sciences Journal of Innovative Education*, 6(2), 375-402.
- Pourjamal Ghouyjah, H., Sotoudehnama, E., & Faghih, E. (2018). On the impact of teaching experience on EFL instructors' self-assessment of their instructional effectiveness. *Iranian Journal of English for Academic Purposes*, 6(1), 47-70.
- Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). Routledge.
- Ray, B., Babb, J., & Wooten, C. A. (2018). Rethinking SETs: Retuning student evaluations of teaching for student agency. *Composition Studies*, 46(1), 34-194.

- Renaud, R. D., & Murray, H. G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education*, 46(8), 929-953.
- Rezalou, A., & Yağiz, O. (2021). EFL students' perceptions and attitudes toward using communicative activities in ELT classroom. *Shanlax International Journal of Education*, 9(2), 112-124.
- Sayavedra, M. (2014). Teacher evaluation. *ORTESOL Journal*, 31, 1-9.
- Scott, C. S. (1977). Student ratings and instructor-defined extenuating circumstances. *Journal of Educational Psychology*, 69(6), 744-747.
<https://doi.org/10.1037/0022-0663.69.6.744>
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499.
- Seldin, P. (1989). Using student feedback to improve teaching. *New Directions for Teaching and Learning*, 26(11), 89-97. <https://doi.org/10.1002/tl.37219893711>
- Seldin, P. (1993). The use and abuse of student ratings of professors. *Chronicle of Higher Education*, 39(46).
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26(11), 1329-1342.
- Solomon, D. (1966). Teacher behavior dimensions, course characteristics, and student evaluations of teachers. *American Educational Research Journal*, 3(1), 35-47.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.

- Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42(4), 276-294.
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson.
- Tennant, G., & Khamis, T. (2017). Student evaluation of teaching: Bringing principles into practice. *Journal of Higher Education in Africa/Revue de l'enseignement supérieur en Afrique*, 15(1), 89-104.
- Theall, M., Abrami, P. C., & Mets, L. A. (2001). *The student ratings debate: Are they valid? How can we best use them?* (1st ed.). Jossey-Bass.
- Usher, A. (2009, May). Ten years back and ten years forward: Developments and trends in higher education in Europe region. *UNESCO Forum on Higher Education in the Europe Region: Access, Values, Quality and Competitiveness* (pp. 75-125).
- Young, S., & Shaw, D. G. (1999). Profiles of effective college and university teachers. *The Journal of Higher Education*, 70(6), 670-686.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-212.
- Weinberg, B. A., Fleisher, B. M., & Hashimoto, M. (2007). *Evaluating methods for evaluating instruction: The case of higher education*. National Bureau of Economic Research. <https://www.nber.org/papers/w12844>.
- Wilson, K. L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the course experience questionnaire. *Studies in Higher Education*, 22(1), 33-53.
- Wimberly, R. C., Faulkner, G. L., & Moxley, R. L. (1978). Dimensions of teacher effectiveness. *Teaching Sociology*, 6(1), 7-20.

- Wong, H., & Leung, S. (2018). How do tertiary education students perceive co-curricular activities under the new education system?. *International Education Studies*, 11(2), 83-96.
- Wotruba, T. R., & Wright, P. L. (1975). How to develop a teacher-rating instrument: A research approach. *The Journal of Higher Education*, 46(6), 653-663.
- Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78(6), 313-317.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9(2), 79-94.
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12(1), 55-76.
- Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, 37(2), 227-235.
- Zuehlke, E. (2008). *College enrollment rate increases, but financial challenges bring uncertainty*. Population Reference Bureau. <https://www.prb.org/resources/college-enrollment-rate-increases-but-financial-challenges-bring-uncertainty/>

APPENDIX

Bu anket formu X üniversitesi tarafından öğretim görevlileri ile ilgili düşüncelerinizi ortaya çıkarmak amacıyla hazırlanmıştır. Anket maddelerini yanıtlarken, maddede bahsedilen ifadeye ne derece katıldığınızı aşağıdaki sıralamaya göre işaretleyiniz. Anketi doldururken Ad Soyad bilgilerinizi paylaşmak zorunlu değildir.

A: Kesinlikle katılıyorum.
 B: Katılıyorum.
 C: Kararsızım
 D: Katılmıyorum.
 E: Kesinlikle katılmıyorum.

2019-2020 YILI 1. DÖNEM ÖĞRETMEN DEĞERLENDİRME ANKETİ	
1	Derslerde pozitif, canlı ve heveslidir.
2	Kendisi ile rahatlıkla iletişim kurulabilir.
3	Öğrenme sürecimde destekleyici ve teşvik edici olmuştur.
4	Sınıfa hâkimdir.
5	Derslerin etkili ve verimli bir biçimde yürütülmesini sağlar.
6	Derslerin konusuna hâkimdir.
7	İhtiyaç duyduğumda bana yararlı olabilecek çalışmalar tavsiye eder ve gerekli yönlendirmeleri yapar.
8	Belirtilen ofis saatlerinde kendisine ulaşılabilir.
9	Derse aktif katılımımı teşvik eder.
10	İkili ve grup çalışmalarına yer verir.
11	Derslerde yapılan etkinlikler çeşitlilik gösterir.
12	Derslere zamanında girip çıkar.
13	Derslere planlı ve hazırlıklı gelir.
14	Ders süresini sonuna kadar etkili kullanır.
15	Dersi İngilizce işlemeye özen gösterir (TOEFL dersleri hariç).
16	Görsel araçları (tahta, resim, video, projeksiyon, internet vb.) etkin kullanır.
17	Bireysel Çalışma Merkezi (Self-Access Center) ve okuma odasını (Reading Room) kullanmamı teşvik eder.
18	Öğrenme sürecimi destekleyecek görsel ve basılı yayınları (film-dizi, belgesel, gazete, dergi, kitap, online kaynaklar vb.) tavsiye eder.
19	Program dışı etkinliklere (örn. Drama, Story Club vb.) katılımı teşvik eder.
20	Sınıf kurallarını (ders giriş-çıkış saatleri, cep telefonu kullanımı, ders materyalleri vb.) titizlikle uygular.
21	Çeşitli etkinliklerde (akıl yürütme, analiz, değerlendirme vb.) eleştirel düşünme becerimi geliştirmeme katkı sağlar.
22	Bu öğretim elemanından ders aldığım için memnunum.