

**PRE-SERVICE ENGLISH TEACHERS' CONCEPTION OF
ASSESSMENT AND THEIR FUTURE ASSESSMENT
PRACTICES IN A TURKISH CONTEXT**

**TÜRKİYE'DEKİ İNGİLİZCE ÖĞRETMEN ADAYLARININ
DEĞERLENDİRME ALGISI VE GELECEKTEKİ
DEĞERLENDİRME UYGULAMALARI**

Ramazan YETKİN

Submitted to the Graduate School of Educational Sciences of Hacettepe
University as a Partial Fulfillment to the Requirement for the Degree of Master in
English Language Teaching Program

2017

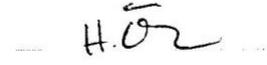
ACCEPTANCE AND APPROVAL

This is to certify that we have read this thesis, entitled "Pre-service English Teachers' Conceptions of Assessment and Their Future Assessment Practices in A Turkish Context", and that in our opinion it is fully adequate, in scope and quality, as a thesis for the Degree of Master in the Program of English Language Teaching.

Chairman Prof. Dr. Arif SARIÇOBAN



Member
(Supervisor) Asst. Prof. Dr. Hüseyin ÖZ



Member Asst. Prof. Dr. Didem KOBAN KOÇ



APPROVAL OF THE GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

This is to certify that this thesis was approved by the aforementioned examining committee members on June 15, 2017 in accordance with the relevant articles of the Rules and Regulations of Hacettepe University Graduate Education, and accepted as a Master Thesis in the program of English Language Teaching by the Board of Directors of the Graduate School of Educational Sciences on ... /..... / 2017.

Prof. Dr. Ali Ekber ŞAHİN
Director of Graduate School of Educational Sciences

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etmeniz ve kütüphane bu talebinizi yerine getirirse bile, teziniz arama motorlarının önbelleklerinde kalmaya devam edebilecektir)

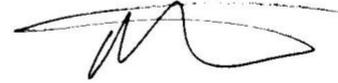
Tezimin/Raporumun 01/11/2017 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

Serbest Seçenek/Yazarın Seçimi:

15.10.2017



Ramazan YETKİN

DECLARATION OF ETHICAL CONDUCT

I have prepared this thesis in accordance with the thesis writing rules and conventions of the Graduate School of Educational Sciences of Hacettepe University, and I hereby declare that:

- All the information and documents have been obtained on the basis of academic rules,
- All audio-visual and written information and results have been presented according to the rules of scientific standards,
- In case of using other works, related studies have been cited in accordance with the scientific standards,
- All cited studies have been fully referenced,
- I did not do any distortion in the data set,
- And any part of this thesis has not been presented as any other thesis study at this or any other university.



Ramazan YETKİN

ACKNOWLEDGEMENT

First and foremost, I owe a debt of gratitude to my supervisor, Asst. Prof. Dr. Hüseyin ÖZ, for his perseverance to help, insightful comments and reflections, and unending supports during every single phase of my study.

Then, I have thanks to Prof. Dr. Arif SARIÇOBAN, and Asst. Prof. Didem KOBAN KOÇ for their guidance and help during all the processes as well as valuable feedbacks and corrections.

I also have special and sincere thanks to my colleagues but most importantly friends, Res. Asst. Arzu Kanat MUTLUOĞLU, Dilara SOMUNCU, Funda ÖLMEZ and Kadriye AKSOY, for their comments, corrections, feedbacks, helps and interests. It is clear that it would be more difficult to pass through this trek without their presence.

Additionally, I specially thank to my colleagues and fellows, Res. Asst. Zekiye ÖZER and Esra KARAKUŞ, for going through all the difficulties from the very first day with me, providing their supports and interests, and sharing my happiness and unhappiness during every stage of my study. I am also grateful to Mr. Gavin T.L. BROWN for his kindness to reply my endless e-mails and for his permission to use his inventory.

I have sincere thanks to my friends, Abdurrahman, Emre, Meyti, Ozan and Semih, for pushing me ahead in my hardest times by their encouragements and supports, and making this difficult process funnier and easier. I am also grateful to TÜBİTAK for providing me with scholarship during my MA study.

Finally, I have greatest thanks to my parents and siblings for providing me necessary support, help and love, compassion and favor during my whole life and education, and my nephews and nieces for making this world my livable for me. Last but not least, I present my deepest gratitude and appreciation to my beloved Asiye, the precious one, for supporting me day and night, presenting her endless good wishes, and being there whenever I need.

To my sister Mine,

My nephews and nieces Nisa, Yusuf, Menekse and Mete,

And, to my precious Asi.

PRE-SERVICE ENGLISH TEACHERS CONCEPTIONS OF ASSESSMENT AND THEIR FUTURE ASSESSMENT PRACTICES IN A TURKISH CONTEXT

Ramazan YETKİN

ABSTRACT

The study was conducted in order to reveal pre-service English teachers' conceptions of assessments regarding improvement, school accountability, student accountability and irrelevance as well as relations between different conceptions purposes. It also aimed to examine how participants' conceptions of assessment differs in relations to their differences of gender, years of learning English, age, grade point average and grade levels.

204 pre-service English teachers participated in the study. The data were collected using Teachers' Conceptions of Assessment Abridged Inventory (TCoA-III A) which is in 6 points Likert scale format ranging from strongly disagree to strongly agree. The obtained quantitative data were analyzed by Statistical Package for Social Sciences (SPSS 23) program.

Descriptive statistics indicated that improvement conception had the highest value among all and participants were moderately agreed that assessment is used for improvement purposes. On the contrary, conceptions of irrelevance were unearthed as having the lowest value and agreement level of all.

Then, Pearson product-moment correlation coefficient was used to investigate relations between conception levels. Correlation results indicated that improvement, school accountability and student accountability conceptions were positively and strongly correlated with each other. On the other hand, there was a negative correlation between improvement and irrelevance conceptions were found out.

A multivariate test of variance (MANOVA) was utilized to examine any effects of individual differences on participants' conceptions of assessment. Multivariate test results indicated that even if there were differences in descriptive results for each variable, grade level is the only independent variables making statistically significant difference on participants' conceptions of assessment. Then, it was

seen that even though grade level made a statistically significant difference among grade levels, results of the Bonferroni adjustment presented no significance difference when the variables considered separately.

Finally, descriptive results from each item were further interpreted with reference to previous studies on conception of assessment in the literature. It was deduced that pre-service English teachers will mostly benefit from formative assessment methods even though the tool can range. Providing feedback to their prospective students will be of high priority for conducting assessment. Secondly, it was interpreted that summative assessment would play a key role for accountability. Therefore, pre-service English teachers would use both formative and assessment tools at the same to to serve for different purposes.

Keywords: assessment, conception, conception of assessment, pre-service English teacher

Advisor: Asst. Prof. Dr. Hüseyin ÖZ, Hacettepe University, Department of Foreign Language Education, Division of English Language Teaching

İNGİLİZCE ÖĞRETMEN ADAYLARININ DEĞERLENDİRME ALGISI VE GELECEKTEKİ DEĞERLENDİRME UYGULAMALARI

Ramazan YETKİN

ÖZ

Bu çalışma İngilizce öğretmen adaylarının “Gelişim”, “Okul Sorumluluğu”, “Öğrenci Sorumluluğu”, ve “Önemsizlik” amaçlarına ilişkin değerlendirme algılarını ve farklı algılama düzeyleri arasındaki ilişkiyi ortaya çıkarmak için yürütülmüştür. Çalışma ayrıca değerlendirme algısının cinsiyet, İngilizce öğrenme yılı, yaş, not ortalaması ve sınıf seviyesi gibi değişkenlere göre nasıl etkilendiğini incelemeyi amaçlamıştır.

Bu çalışmaya 204 tane İngilizce öğretmen adayı katılmıştır. Veri TCoA- IIIA- Version 3- Abridged isimli kesinlikle katılıyorum ile kesinlikle katılmıyorum arasında altılı Likert ölçeği formatında olan envanter kullanılarak toplanmıştır. Elde edilen nicel veri SPSS 23 isimli yazılım programı kullanılarak analiz edilmiştir

Betimsel istatistik gelişim algısının en büyük değere sahip olduğunu ve katılımcıların kısmen değerlendirmenin gelişim amaçları için kullanımında hemfikir olduğunu göstermiştir. Diğer yandan, önemsizlik algısının en düşük değere ve hemfikir olma seviyesine sahip olduğu ortaya çıkarılmıştır.

Daha sonra, algılama seviyeleri arasındaki ilişkileri araştırmak için Pearson korelasyon katsayısı kullanılmıştır. Korelasyon sonuçları gelişim, okul ve öğrenci sorumluluğu algılamalarının pozitif ve güçlü bir şekilde ilintili olduğunu göstermiştir. Diğer taraftan, gelişim ve önemsizlik algılamalarının negatif bir ilişkiye sahip olduğu ortaya çıkarılmıştır.

Katılımcılarının bireysel farklılıklarının onların değerlendirme algısı üzerindeki etkilerini incelemek için çoklu varyans analizi kullanılmıştır. Çoklu varyans analizi sonuçları, betimsel istatistik sonuçlarının her değişkenin farklılık oluşturduğunu göstermesine rağmen, sınıf seviyelerinin katılımcılarının değerlendirme algılarında manidar bir fark ortaya çıkaran tek bağımsız değişken olduğunu göstermiştir. Daha sonra, sınıf seviyelerinin manidar bir fark ortaya çıkarmasına rağmen, Benferroni adaptasyonundan sonra değişkenlerin ayrı ayrı ele alındığında manidar bir fark ortaya çıkarmadıkları görülmüştür.

Son olarak, her maddeden elde edilen betimleyici sonuçlar alandaki değerlendirme algısı üzerine çalışmalara ilişkin olarak yorumlanmıştır. İngilizce öğretmen adaylarının değerlendirme araçları değişse de genellikle biçimlendirici değerlendirme yöntemleri kullanacağı sonucuna varılmıştır. Gelecekteki öğrencilerine geri dönüt sağlama değerlendirme uygulamasının önceliklerinden olacaktır. İkinci olarak, özetleyici değerlendirmenin sorumluluk için çok önemli bir rol oynayacağı değerlendirilmiştir. Bu yüzden, İngilizce Öğretmen adaylarının farklı amaçlar için hem biçimlendirici hem de özetleyici değerlendirme aygıtlarını kullanacakları sonucuna varılmıştır.

Anahtar sözcükler: değerlendirme, algı, değerlendirme algısı, İngilizce Öğretmen adayı

Danışman: Yrd. Doç. Dr. Hüseyin ÖZ, Hacettepe Üniversitesi Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Eğitimi Bilim Dalı

TABLE OF CONTENTS

ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	xi
LIST OF TABLES	xiii
ABBREVIATIONS.....	xv
1. INTRODUCTION	1
1.1. Introduction	1
1.2. Background of the Study	1
1.3. Statement of the Problem	2
1.4. Significance of the Study.....	3
1.5. Purpose of the Study.....	3
1.6. Limitations of the Study	4
1.7. Definitions	4
1.8. Conclusion	5
2. REVIEW OF LITERATURE.....	6
2.1. Introduction	6
2.2. Assessment	6
2.3. Basic Concepts of Assessment.....	7
2.3.1. Formative Assessment	7
2.3.2. Summative Assessment	8
2.3.3. Traditional Assessment	9
2.3.4. Alternative Assessment	11
2.3.5. Criterion-referenced and Norm-referenced Assessment.....	12
2.4. Principles of Assessment	13
2.4.1. Reliability	13
2.4.2. Validity.....	15
2.4.3. Practicality.....	16
2.4.4. Authenticity.....	16
2.4.5. Washback	17
2.5. Conceptions of Assessment.....	18
2.5.1 Improvement Conception.....	19
2.5.2 Student Accountability	20
2.5.3 School Accountability.....	21
2.5.4 Conception of Irrelevance.....	21
2.6. Teachers' Roles in Assessment.....	22
2.7. Research Studies Conducted on Conception of Assessment	23
2.8. Assessment Practices and Conceptions of Assessment in Turkey	24
3. METHODOLOGY	27
3.1. Introduction	27
3.2. Research Design	27
3.3. Research Questions	27
3.4. Variables	28
3.4.1. Dependent Variables	28

3.4.2. Independent Variables.....	28
3.5. Setting and Participants	29
3.5.1. Setting	29
3.5.2. Participants	29
3.6. Instrumentation	31
3.7. Data collection procedures.....	32
3.8. Data analysis procedures.....	33
4. RESULTS	35
4.1. Introduction	35
4.2. Results of Data Analysis	35
4.2.1. What are pre-service English teachers' conceptions of assessment?	35
4.2.2. How do levels of conceptions of assessment relate to each other?	38
4.2.3. Are there any significant differences in the participants' conceptions of assessment regarding different variables	39
4.2.3.1. Gender.....	39
4.2.3.2. Years of Learning English.....	41
4.2.3.3. Age	42
4.2.3.4. Grade Point Average(GPA)	44
4.2.3.5. Grade Level.....	45
5. DISCUSSION	48
5.1. Introduction	48
5.2. Summary of the Study	48
5.3. Discussion of Findings in Terms of Research Questions.....	49
5.3.1. Discussion of research question 1	49
5.3.2. Discussion of Research Question 2	50
5.3.3. Discussion of Research Question 3	51
5.3.3.1. Gender.....	51
5.3.3.2. Years of learning English.....	52
5.3.3.3. Age	53
5.3.3.4. Grade point average (GPA).....	54
5.3.3.5. Grade levels	55
5.3.4. Discussion of future assessment practices	56
6. CONCLUSION.....	59
6.1. Introduction	59
6.2. Implications for Practice	59
6.3. Suggestions for Further Research	60
6.4. Limitations of the Study	60
6.5. Conclusion	61
REFERENCES	62
APPENDICES.....	67
APPENDIX A. ETİK KOMİSYONU ONAY BİLDİRİMİ	67
APPENDIX B. ORJİNALLİK RAPORU	69
APPENDIX C. TEST OF NORMALITY- HISTOGRAM	71
APPENDIX D. TEST OF NORMALITY- Q-Q PLOT	72

APPENDIX E. DATA GATHERING INVENTORY SAMPLE.....	73
CIRRICULUM VITAE.....	74

LIST OF TABLES

Table 3.1: Demographics of Participant Pre- Service English Teachers.....	30
Table 3.2: Alpha values per level.....	32
Table 4.1: Levels of conception of assessment of TCoA-III A, Version 3- Abridged Scale (N=199).....	35
Table 4.2: Pre-service teacher's improvement level of conception of assessment.....	36
Table 4.3: Pre-service teacher's school accountability level of conception of assessment.....	36
Table 4.4: Pre-service teacher's student accountability level of conception of assessment.....	37
Table 4.5: Pre-service teacher's irrelevance level of conception of assessment.....	37
Table 4.6: Relationship between levels of conceptions of assessment	38
Table 4.7: Levene's Test of Equality of Error Variances	40
Table 4.8: Descriptive statistics of dependent variables for male and female participants.....	40
Table 4.9: Wilks' Λ for differences in conception between male (n=53) and female (n= 146) participants.....	40
Table 4.10: Levene's Test of Equality of Error Variances	41
Table 4.11: Descriptive statistics of dependent variables for participant's years of English education	41
Table 4.12: Wilks' Λ for differences in conceptions between education years; Less than 10 years (n= 26), 10 years (= 49), 11 years (n=42), 12 years (n=44), 13 years or more (n=33) of participants	42
Table 4.13: Levene's Test of Equality of Error Variances	43
Table 4.14: Descriptive statistics of dependent variables for age differences of the participants.....	43
Table 4.15: Wilks' Λ for differences in conceptions between different ages'; 20 years or less (n=97) and 21 years or more (n= 102) groups	43
Table 4.16: Levene's Test of Equality of Error Variances	44
Table 4.17: Descriptive statistics of dependent variables for grade point average (GPA) scores of the participants.....	44

Table 4.18: Wilks' Λ for differences in conceptions between high (n=48) and medium (n= 151) achievers.....	45
Table 4.19: Levene's Test of Equality of Error Variances.....	46
Table 4.20: Descriptive statistics of dependent variables for grade levels of the participants.....	46
Table 4.21: Wilks' Λ for differences in conceptions between high (n=48) and medium (n= 151) achievers.....	46
Table 4.22: MANOVA for differences in conceptions of assessment based on grade levels.....	47

ABBREVIATIONS

TCOA-III A: Teachers Conceptions of Assessment Inventory Abridged

STACC: Student Accountability

SCACC: School Accountability

IRR: Irrelevance

IMP: Improvement

N: Number of Participants

M: Mean Value

SD: Standard Deviation

P: Significance

SPSS: Statistical Package for Social Sciences

MANOVA: Multivariate Analysis of Variance

1. INTRODUCTION

1.1. Introduction

The present thesis seeks to examine conceptions of assessment among student teachers of English teachers in teacher education program and possible effects of perceived assessment conceptions on their assessment practice. Additionally, it seeks to reveal purpose and utmost reason of which an assessment practice is conducted in the classroom. In this line, this chapter began with presenting background information of the study at first. Then, problem was stated, and purpose and significance of the study were presented successively. Finally, definitions of the key terms were given in the last section of the chapter.

1.2. Background of the Study

Although conceptions of people may differ in what the assessment is, it is undisputable that it plays a pivotal role in education. Almost any educators incline to use assessment practices in some places of their teaching process usually to decide learners' successes of learning or failure. In line with these thoughts, assessment is "the process of defining, analyzing, interpreting and using information to increase students' learning and development" (Erwin, 1991, p. 15). It is used to gather necessary information in order to make decisions (Fenton, 1996).

It is a common belief that assessment is of practitioners' responsibility. "Classroom assessment requires a great deal of time and effort; teachers may spend as much as 40% of their time directly involved in assessment-related activities" (Stiggins, 1988, p. 363). Yet, policy makers, parents and pupils have a shared responsibility for assessment practices. As Danielson (2008) noted, assessment is a key for creation of education-based policies. It is used to determine how well students learn as well as to give information about the format and improvement of educational instructions and settings.

Conception of assessment is a term which seeks to reveal the purposes of conducting assessment practices. There are a number of purposes of assessment that categorized under four main purposes: improvement, school accountability, student accountability and irrelevance (Brown, 2004, p. 304). In short,

improvement conception proposes that assessment is used to improve quality and amount of learning; school accountability suggests that assessment is used to check school's performance; student accountability offers that assessment is conducted to see students' progress for learnings and finally irrelevance conceptions put forwards that assessment is of no aim and useless.

So far, very few researchers have been studied teachers' conceptions of assessment in the Turkish context. Zaimoğlu (2003) sought to reveal teachers and students' conceptions of assessment in an EFL preparatory school, it was found that improvement conception held the highest value. Besides, Vardar (2010) conducted a study in order to discover secondary school teachers' conceptions of assessment and unearthed that students' accountability kept the highest priority of all. Similarly, Yüce (2015) echoed the results of Zaimoğlu's (2003) study, in which Yüce focused on pre-service English language teachers' conceptions of assessment and revealed that they mostly used or planned to use assessment for improvement.

1.3. Statement of the Problem

Assessment is a crucial and key part of education, but practicing of assessment is demanding. As Stiggins (1988) reports classroom assessment necessitates almost half time of teachers to prepare and conduct. Even though assessment is a common practice, it doesn't have fixed rules and borders in general so that practitioners may benefit from it, that's why, it becomes a demanding task. However, teachers are not taught or ready for such a task (Stiggins, 1988). In this case, teachers' beliefs and practices play a key role for application of assessment techniques. According to Pajares (1992), beliefs and acts are so interconnected that beliefs of teacher candidates will likely to affect their application and practice in their real classrooms. In order to make assessment more meaningful, useful, and applicable, it is eminent to reveal teacher candidates' conceptions about assessment and provide them with necessary training about purposes of assessment.

Griffiths, Gore and Ladwig (2006) found out that practitioner's beliefs are even far greater and effective than their school experience and context on their

preferences. So, its important to uncover what they believe in order to shape their understanding according to the educational policies and needs.

Even though increasing number of studies are being conducted on conceptions of assessment recently, a few of them have been carried out in Turkey so far. Therefore, researching pre-service English teachers' conceptions of assessment in current setting likely to reveal beliefs, procedures, assessment practices and curriculum as well as contribute to literature.

1.4. Significance of the Study

Assessment practices are so commonly used at any level of education, though conceptions of practitioners have been ignored or less analyzed so far. This study will contribute to the conception of assessment literature by examining and revealing pre-service English teacher's conceptions of assessment in Turkish context. Studying teacher candidates' conceptions will help us to understand assessment practices, students' approaches to assessment and teacher training, along with giving some important clues about overall assessment procedures in Turkish educational context.

1.5. Purpose of the Study

Brown (2008) suggests that people's beliefs and the rules of their social environment appear to be important in determining their type of behavior and practices. Beliefs and conceptions of people play an important role in the implementation and assessment process of teaching and learning environment. Wiggins and McTighe (1998) put forward that effective teachers pioneer quality of teaching by creating a good design as well as planning the lesson like an assessor prior to implementation.

Every teacher uses their own way of assessing to students learning outcomes based on their thoughts and perceptions about teaching, learning, assessing and this shapes students' performance outcomes. Hence, focusing inclusively teachers' beliefs during their training and professional development seems to be of high importance (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997).

Therefore, the purpose of this study to reveal pre-service English teachers' conceptions of assessment and their possible effects on their prospective real class practices. It also aims to explain any possible effect of variables such as years of English education, grade, success, and gender on their beliefs about assessment practices. In order to conduct the research, the following research questions were formulated:

1. What are pre-service English teachers' conceptions of assessment?
2. What is the relationship among conceptions of assessment of the pre-service teachers?
3. Are there any significant differences in the participants' conceptions of assessment by;
 - a. Grade level
 - b. Gender
 - c. Academic achievement (GPA)
 - d. Years of Learning experience
 - e. Age

1.6. Limitations of the Study

In this thesis, listed reasons would be seen as the limitations of the study especially with generalizability of the results.

1. The data were collected and analyzed by using quantitative methods. Absence of any qualitative method could be a limitation.
2. Of all the participants were from the same setting and absence of participants from different settings could be a limitation for generalization.
3. Participants' possible future assessment applications are withdrawn from their answers to survey items. An interview with students would be more effective to make inference.

1.7. Definitions

Assessment; it is "an ongoing process aimed to improve student learning" (Jandra, 2011, p. 2). Erwin (1991) comes up with a more detailed explanation as

“the process of defining, analyzing, interpreting, and using information to increase students’ learning and development” (p.15).

Conception; it is a window by which someone sees, views, interprets and understands their thoughts of world (Pratt, 1992). Conceptions is a more general term being made up of beliefs, concepts, wishes, preferences, meanings, thoughts and so. (Thompson, 1992).

Pre-service teacher; it is defined as “a pre-service teacher is a student teacher who has not yet undertaken any teaching and completed his training to be a teacher” (Yüce, 2015, p. 7)

Conception of assessment; it is a term used to reveal people’s conceptions of purposes of assessment use. It is categorized under four main conceptions: “improvement, school accountability, student accountability and irrelevance” (Brown, 2004, p. 304).

1.8. Conclusion

The present chapter was designed to give an overall idea about the content, aim and structure of the thesis. In this chapter, some background information was given firstly. Then problem(s) was stated and purpose and significance of study- why such a study was conducted- was tried to be explained. Then, some relevant and key definitions were provided and the chapter was concluded by presenting limitations of the study.

2. REVIEW OF LITERATURE

2.1. Introduction

The aim of the study is to investigate pre-service English teacher's conceptions of assessment and their relative applications in their real classroom teaching. Accordingly, the chapter is made up of relevant literature. Firstly, it gives insight information about background of assessment and its types; then, it focuses more specifically on the notion of "conception of assessment" and refers to its four main dimensions.

2.2. Assessment

Assessment plays a pivotal role in the process of language learning and teaching. It not only gives information to teachers about how effective their teaching is, but also to students about how well they learn, understand and internalize related topics. Accordingly, teachers could judge and renew- if necessary- their methods and related materials, and students could take a different look into their way of studying. According to Black and William (1998), "assessment refers to all those activities undertaken by teachers, and by the students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 2). In short, "Assessment involves making assumptions about what exists, what it is like and how we might know about it" (Knight, 2002, p. 279) According to Gonzales (2003), assessment is "a systematic gathering of information about students' performance that enables teachers to monitor their learning" (p.89). To help students learning, Harlen (2005) proposes that "the students, the ones who do the learning, have information about where they are in their learning, what steps they need to take and how to take them" (p. 215).

Assessment can serve many different purposes. According to Trotter (2006), "assessment can be used to provide motivation. Strategies for modifying the assessment system that can influence students' approaches include integrating assessment into the learning process so that what is assessed is the total learning experience" (p. 508). Assessment is also used for making decisions. Harlen (2005) in her study puts forward that;

All assessment in the context of education involves making decisions about what is relevant evidence for a particular purpose, how to collect the evidence, how to interpret it and how to communicate it to intended users. Such decisions follow from the purpose of conducting the assessment. These purposes include helping learning, summarizing achievements at a certain time, monitoring levels of achievement, and research (p. 207).

Then, assessment also plays an effective role in educational reform. Cheng (1999) brings up some reasons of assessment roles in educational reform as “first, assessment results are relied upon to document the need for change. Second, assessments are seen as critical agents of reform. Third, assessment results are used to demonstrate that change has or has not occurred” (p. 254).

Different assessment types could serve for different purposes. Badders (2000) underlines that “different kinds of information must be gathered about students by using different types of assessment. The types of assessments that are used will measure a variety of aspects of students learning, conceptual development, and skill acquisition and application” (p. 2).

2.3. Basic Concepts of Assessment

2.3.1. Formative Assessment

Formative assessment is generally known as assessment for learning. Moss and Brookhart (2010) define it as “an active process that partners the teacher and the student to continuously and systematically gather evidence of learning with the express goal of improving student achievement” (p. 6). In addition to promoting students’ learning, formative assessment, according to Brown (2004), has the purpose of assessing students during the process through which they form their skills and competencies and helping to make this growing process permanent. “The effectiveness of formative assessment depends on whether students actually perceive the gap between where they currently are and where they should be; and then if they do, what they are willing to do about closing it.” (Biggs, 1997, p.104); hence, any information “...would be called formative if it were used to help learning and teaching” (Harlen, 2005, p.208)

Formative assessment also helps teachers to promote their professional development. Any teacher, in order to help their students to improve and sense,

should be aware and have the knowledge of formative assessment techniques. Baird (2011) claims that “formative assessment is very much to do with teacher practice and its implementation has been seen as a form of professional development” (p. 344). Teachers should always be ready and update their beliefs and knowledge about assessment in order to tackle with new or unforeseen challenges. New practices of assessment could bring new challenges to teachers existing competencies, knowledge and beliefs that they already form about the aims and purposes of assessment (Muñoz, Palacin & Escobar, 2012).

The notion ‘feedback’ is of high importance in formative assessment classes. Black and William (2006) argue that in order to have a substantial improvement for learning in the classroom, appliance of formative assessment may be the sole way, by which interactive feedback could be given, so this shapes and effects quality of learning and relevant pedagogy alike. By this way, “assessment (formative) is specifically intended to provide feedback on performance to improve and accelerate learning.” (Sadler, 1997, p. 77) According to Harlen and James (1997), formative assessment supplies both teachers and students with necessary competencies and understanding to plan the next step. They state that “the judgment of a piece of work, and what is feedback to the pupil, will depend on the pupil...” (Harlen & James, 1997, p.370), but in this sense, students need to get instructions related to interpretation of feedback and building connections between the feedback and their production (Sadler, 1998).

2.3.2. Summative Assessment

Summative assessment, mostly referred to assessment of learning, is defined as “assessment which counts towards, or constitutes a final grade for, a module or course or here a pass is required for progression by the student” (Bloxham & Boyd, 2007, p.236). According to Harlen & James (1997), the aim of summative assessment is based on reporting the results to each interested party including parents, teachers, students themselves as well as school governors and boards.

Harlen (2005) categorizes uses of summative assessments into two groups; internal and external. Internal usage of summative assessment is made up of “regular grading for recordkeeping, informing decisions about courses to follow where there are options within the school, and reporting to parents and to the

students themselves” (p. 208). By the help of feedback, all three parties (teachers, students and parents) could be aware of needs and progress. External uses of assessment are comprised of “certification by examination bodies or for vocational qualifications, selection for employment or for further or higher education, monitoring the school’s performance and school accountability” (p. 208), so that acquired information could be used for making decision primarily about students’ improvement as well as teachers and schools.

In his study, which is about students’ perceptions of continuous summative assessment, what Trotter (2006) found out is that even if it is time consuming and requires hard-working, it results in improvement for students and, this eliminates additional work. Along the same line, Harlen and James (1997) states that “it (summative assessment) has an important role in the overall educational progress of pupils” (p. 370), through which teachers could draw inferences about their students’ progress as well as their own way of teaching.

“Summative assessment methods are typically paper and pencil measures such as quizzes, tests, exams, essays or projects that form a portion of a student’s final grade” (Volante, Beckett, Reid & Drake, 2010, p. 3). According to Harlen and James (1997), the characteristics of summative assessments include; it is applied at certain times when success needed to be revealed, it focuses on students’ progress, different performance outcomes could be used for the same purposes since they based on same criteria, it should include a reliable and valid method, it should include procedures for quality insurance, and it should be evidence-based.

2.3.3. Traditional Assessment

Traditional assessment, also referred as paper-pencil assessment, is by far the most used assessment type in many educational settings. It includes a wide variety of test types including open-ended, short answer, true-false and the like as its evaluation tool (Çalışkan & Kaşıkçı, 2010). According to Abbott (2012), “traditional assessments generally test an individuals’ ability to recall or apply knowledge within specific time limits - do our exams entice students to engage with subject matter, or compel them to simply grapple with it?” (p. 36); namely, they aim to uncover subject areas that students have some degree of problems (Slater, Ryan & Samson, 1997). In order to find out the the reasons behind wide

and frequent choice and use of traditional tests, Çalışkan and Kaşıkçı (2010) puts forward in their study that;

It was found that social studies teachers always prefer to use multiple choice tests in the assessment and evaluation process, besides which they usually use open-ended, short answer and true-false tests. The reason why teachers widely apply these traditional tools could be their sense of self-adequacy in preparing, applying and evaluating these tools, familiarity with the use of these tools and the assumption that these tools measure the knowledge of the students accurately (4155).

Abbott (2012) comes up with three dimensions of traditional assessment theme by which students effective learning can be accelerated, these are: take-home exams, oral examination and group examinations. In order to promote students deeper learning and provide them with necessary precautions to hamper possible learning or understanding breakdowns, preceding traditional assessment test types should be benefited and applied conveniently.

Brown and Abeywicrama (2012), in their book, list features of traditional assessments: (a) standardized exams, (b) timed, multiple choice format, (c) decontextualized test items, (d) scores suffice for feedback, (e) norm-referenced scores, (f) focused on discrete items, (g) summative, orient to product, (h) non-interactive performance, (i) fosters intrinsic motivation. Similarly, Anderson (1998) believes that traditional assessment has “philosophical beliefs and theoretical assumptions” and he itemizes these features as follows;

1. assumes knowledge has universal meaning,
2. treats learning as a passive process,
3. separates process from product,
4. focuses on mastering discrete,
5. focuses on mastering discrete, isolated bits of information,
6. assumes the purpose of assessment is to document learning,
7. believes that cognitive abilities are separated from affective and conative abilities,
8. views assessment as objective, value-free, and neutral,
9. embraces a hierarchical model of power and control (8-9).

2.3.4. Alternative Assessment

Alternative assessment, which also known as performance- based or authentic assessment (Hancock, 1994), provides new opportunities to all the parties of the school context- teacher, students, parents and school a like- besides traditional approaches to assessment manner. Alternatively, it brings new ways to students' performance demonstration over time rather than paper-pencil exams and its pressure over students with time limits. Alternative assessment states that there had better to be new tools for collecting students' achievements, and similarly new processes to diagnose students' achievement outcomes to look for each students' unique favors (Corcoran, Dershimer &Tichenor, 2004). Likewise, Krajcik, Czerniak and Berger (1999, cited in Corcoran et al, 2004) notes that alternative assessment has both high validity and reliability, tolerates cultural differences, assesses understanding thoroughly, and stays close with cognitive learning techniques.

Although many think that alternative assessment techniques take redundant time and bring extra burden (Şahin &Karaman, 2013), it offers new and variety of formats by which students are able to show their capabilities over subject matters and different skills (Yıldırım, 2004). Supportively, "alternative testing offers a both the teacher the opportunity not to compare levels and knowledge but to follow a students' evolution individually and in time" (Chirimbru, 2013, p. 93).

Alternative assessment, since its authentic feature, prepares students for real life, so students can make use of what they learn in the class out of class through conceptualizing and internalizing. Hamayan (1995) highlights that "alternative assessment... can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom" (p.213), and honor students to develop and make use of their own thoughts out of their experiences (Corcoran et al, 2004).

There are many characteristics and strategies of alternative assessment (Buck, 1999 cited in Corcoran et al, 2004; Corcoran et al, 2004; Frank &Barzilai, 2004; Herman, Aschbacher &Winters, 1992), they include;

1. Alternative assessment should include both qualitative and quantitative measurements.
2. It aims to measure real word based meaningful activities.

3. It includes higher order thinking skills.
4. Students should perform a tasks and conceptualize from their own experiences.
5. It focuses in product- improvement- and uses different formats for assessing students' achievement.
6. It is not a one-time process, instead it extent evaluation process over and different times.

2.3.5. Criterion-referenced and Norm-referenced Assessment

Criterion referenced assessment is an approach to testing “in which the learner is assessed purely in terms of his/her ability in the subject, irrespective of the ability of his/her peers” (Verhelst, Van Avermaet, Takala, Figureas, & North, 2009, p. 184). Therefore, CR assessment places students according to their scores out of some pre-designated criteria. Instead of comparing students to each other, it compares to students according to their scores in the learning objectives, and places them by looking at their achievement on the specific learning objectives (Kim, Lee, Chung, & Bong, 2010).

According to CR based assessment, the focal point should be what the students have already accomplished instead of the amount of their achievements. Moreover, students' scores should stem from their performances in certain criteria and objectives in a crystal clear manner, instead of depending upon other students' performances (Airasian & Madaus, 1972 as cited in Tyler & Wolf, 1974). Knight (2001), in his book, presents some advantages of CR assessment as follows;

1. Assessment criteria clearly identify what is valued in a curriculum.
2. In criterion-referenced curricula, teachers know exactly what they should teach.
3. Level descriptors make it clear to learners what they have to show in order to get a particular mark.
4. Level descriptors make it possible to give learners feedback which identifies what they need to do in order to get better marks.

5. Level descriptors can be used to make assessment feed out informative, identifying exactly what learners have achieved.
6. It is possible to make judgments about the quality and quantity of learning (p.19)

Norm referenced assessment, on the contrary, divides students (as successful or unsuccessful) regarding their placements compared to their friends, not based on any pre-designated criteria or learning objectives which value their performance instead (Airasian & Madaus, 1972 as cited in Tyler & Wolf, 1974). Kim et al. (2010) reveals that;

Norm-referenced assessment, compares each student's performance to the performance of others in the same reference group. Students' scores are largely determined by the relative superiority or inferiority of their performance compared to those of other students, regardless of how much of the specific learning objectives they successfully mastered (142).

According to Bond (1996), NR assessment is used in order to order students from high to low achievers, and their performance results are regulated at NRT's by comparing to a large group of students with similar levels. Knight (2001) proposes that use NR assessment makes "reasonable to reward" students because it provides us with necessary data to compare to students each other, and order students into ranks so that we can reveal who is first and last achievers instead of comparing them according to performances over learning objectives. In this line, Bond (1996) exemplifies NR assessment as "if a student receives a percentile rank score on the total test of 34, this means that he or she performed as well or better than 34% of the students in the norm group" (p.2).

2.4. Principles of Assessment

2.4.1. Reliability

Reliability has long been seen as one of the key factors of any assessment tool in order to make any assessment process reliable, dependable and consistent. Many definitions have been uttered so far, but its importance for any assessment device has preserve its valuable role without any change. According to Stanley (1964), reliability means "consistency or stability of a measurement" (p. 150). After delivering the test to the same test-takers in different times, but without no language practice between times, reliability makes the test results sure that they will be very close (Heaton, 1988) Similarly, Brown and Abeywicrama (2010)

asserts that any reliable test type should be coherent and trustworthy, and when you deliver the same test to same or similar students, the outcomes should share common results.

There seems to be four important factors affecting reliability; test takers, scoring process, administration matter and the assessment tool. In order to assert a test as reliable, the test should yield similar results on two or more applications, it should have clear rubric for assessment, test items and guidance should be clear, and it should share common ordinance for scoring and evaluation process (Brown & Abeywicrama, 2010).

The reliability concept includes four main reliability types, these are student related reliability, rater reliability, test administration reliability and test reliability.

Student related reliability could occur because of any psychical or psychological problem of test takers such as anxiety and illness, so the test taker could not get his/her exact score outcome during that test application (Brown and Abeywicrama, 2010).

Rater reliability occurs when there are salient similarities or differences in test takers' scores because of the different scorers. It has two types: inter rater reliability and intra rater reliability. Inter rater reliability means that more than one scorer has provided similar results after scoring the same test. On the contrary, intra rater reliability means one scorer, especially classroom teachers, always yields different results due to unclear rubric and direction for scoring or labeling students as good student and bad student and the like (Brown& Abeywicrama, 2010).

Test administration reliability is seen when the conditions during test administration has an adverse effect on test- takers. Examples can be noisy streets, bad lighting situation, too cold or hot classrooms, unsuitable chairs and desk and the like (Brown and Abeywicrama, 2010).

Test reliability refers to content and composition of the test itself and its items. According to test reliability concept, multiple choice test items should be evenly challenging, distractors should be relevant and well-created, test items should be well-designed and distributed. Likewise, essay type as open ended tests or

subjective tests should have well-designed scoring rubrics and objective evaluation criteria and the like (Brown and Abeywicrama, 2010).

2.4.2. Validity

Validity is by far the most important feature any test should carry. Any valid test already counted as reliable but the opposite is not possible all the time, that's why looking for "validity" criteria hold a pivotal role for any assessment tool. Validity, concisely, means "usefulness (of a test) for a given purpose, especially for predicting an outcome (Stanley, 1964, p. 150). According to Brown and Abeywicrama (2010), a valid test should measure what is intended beforehand, should discriminate irrelevant variables, should focus on performance of test-takers and includes performance as a criterion, provide beneficial outcomes regarding test takers capabilities and should be backed up by relevant construct and theories. According to Heaton (1988), any valid test should also assess "particular skills" that is looked for.

The concept of validity includes five main types: content validity, criterion validity, construct validity, consequential validity, and face validity.

Content validity encompasses that the test should represent the course content as well as making apparent that course goals and aims should overlap with test items (Heaton, 1988).

Criterion validity connotes the extent to which the result of the tests evokes the pre-determined criteria of the test (Brown & Abeywicrama, 2010). It has two types: predictive and concurrent validity. Predictive validity is basically aimed to measure possible future successes of test-takers instead of current situation (i.e. placement tests), whereas concurrent validity looks for result "in respect of the particular criterion used" (Heaton, 1988, p. 161), and requires to see some other performance outcomes besides the assessment (Brown& Abeywicrama, 2010).

Construct validity denotes that if the test stems from a theory, then scores or results of the test should include and associate to characteristics of that theoretical framework (Stanley, 1964). For example, if the test is constructed to measure linguistic proficiency, it should not only test some linguistic features such as accuracy and fluency as well as showing some relevance to other proficiency tests.

Consequential validity deals with all the results of the test such as how well the test measured pre-designated criteria, its effects on students (i.e. preparation, wash back), and any social outcome of the test purposeful or not (Brown & Abeywicrama, 2010).

Face validity entails that pupils conceive assessment as beneficial, equal and appropriate way for improvement (Gronlund, 1998). According to Heaton (1988), when the test looks good enough to administrators, teachers, test takers and the like, then it could be inferred that it has the face validity (Heaton, 1988).

2.4.3. Practicality

Practicality of assessment which is another benchmark (principle) of assessment should be kept in mind and ensured before any assessment application take place, means that whether the assessment tools and process are proper and applicable to the context regarding time, management, cost and the like. According to Brown and Abeywicrama (2010), practicality in/of assessment “refers to the logistical, down-to earth, administrative issues involved in making, giving, and scoring an assessment instrument” (p. 26). They exemplify and explain practicality as if a test is taking five hours of test-takers or if takes five minutes of test-takers to complete but several hours of examiner to evaluate, then the test is impractical. If a test meets following criteria: (a) cost effective, (b) can be completed within suitable time limit, (c) has open and crystal clear directions for application, (d) fitted into available resources, (e) effective benefits from human resources, (f) and, regarding time and effort for both preparation and evaluation processes, then the test can be considered as practical (Brown & Abeywicrama, 2010).

2.4.4. Authenticity

Authenticity of assessment refers to relevance of assessment tools or contents in to the real or authentic world, the use or inclusion of tasks, language etc. from the authentic environment. Brown and Abeywicrama (2010) puts forward that in order to say this test is authentic, then the test task should be included and presented in the real world, because “there is often a gap between what we require of students in assessment tasks and what occurs in the world of work” (Boud, 1990, p. 101).

Authenticity of assessment plays a key role in educational reform and raising learners who meet current social and informational standards. Gulikers, Bastiaens, Kieschner and Kester (2006) stated that there is a shift from standardized assessments to performance-based assessment, therefore authenticity plays a significant role in this process. Maclellan (2004) claimed that when learners conceive a need to figure out the materials in order to achieve that goal or task, then they will have deeper learning. Similarly, if learners get assessment as real and authentic, the assessment task will be valued (Palmer, 2004). In her study that aimed to investigate academic's perceptions of authenticity in assessment task, Maclellan (2004) concluded that "assessment should focus on real world problems and have some meaning to real world audience" (p. 19).

2.4.5. Washback

Washback is generally defined as "the effect of testing on teaching and learning" (Hughes, 2003 as cited in Brown & Abeywicrama, 2010, p. 37). According to Brown and Hudson (1998) effects of washback could be either harmful or helpful to educational process. They claimed that if the test procedures don't meet the goals, aims and objectives of the curriculum, then the test can create negative washback. On the contrary, if the test meets the standards, objectives and aims of the curriculum, then the assessment will result in positive washback effect. In this respect, Brown and Abeywicrama (2010) differentiate washback from impact on assessment as washback effect can be "both promotion or inhibition of learning" (p. 37)

Brown and Abeywicrama (2010) stated that "washback can have a number of positive manifestations, ranging from the benefit of preparing and reviewing for a test to the learning that accrues from feedback on one's performance" (p. 38). In this line, Green (2013), in his review study on washback on language assessment, revealed some effects and benefits of washback effect and studying washback as follows:

1. The identification of needs in relation to communication between test providers and other stakeholders is one likely outcome of researching washback.

2. Better understanding of how washback occurs in teaching and learning processes can help to inform targeted intervention.
3. Research evidence can be a powerful tool for encouraging participants to reconsider their current practices.
4. Washback research has given us some new insights into how tests are used and how they are accommodated in a wide range of educational settings.
5. It is very clear that washback, like other forms of evidence in our field, has to be considered in relation to specific contexts of test use (p. 48-49).

2.5. Conception of Assessment

Conceptions play an important role on shaping peoples' ideas, behaviors and the way they act. Brown, Hui, Yu and Kennedy (2011) defines conception as "ecologically rational representations of the thought and practice traditions an individual experience within a culture" (308). Brown and Hirschfeld (2008) is interested in the effect of conception on education by asserting that "...conceptions have an impact on their educational experiences and learning" (3).

The conceptions or beliefs that teachers hold play a pivotal role on teaching and learning process. Teachers, as the leading and mediating figure in the classroom, guide and inform the class according to their beliefs. Harris and Brown (2009) indicates that "teachers' conceptions of assessment are important as they shape their usage of assessment practices" (p. 365). Similarly, "teachers are a key factor in turning assessment information and processes into improved learning. Thus, it is important to understand "what teachers think about assessment and how they make use of it" (Brown, Kennedy, Chan & Yu, p. 348). Teachers' techniques for assessing student's outcomes vary according to their view of language, assessment, learning and teaching (Moiinvaziri, 2015), so it is important to give a great attention to their beliefs (about assessment) in order to understand their practices well and look for new reforms on assessment practices, if necessary, (Brown, Lake & Matters, 2011), since they are the key figure not only for learning process but also for interpretation and implementation of assessment results into learning process (Azis, 2012).

Culture is another benchmark influencing conceptions and educational policies. According to Brown, Lake and Matters (2011), “differences in culture or society lead not only to differences in policy but also to differences in conceptions of corresponding practices and processes” (p. 211). That’s why teachers’ conceptions of assessment not only affect their practices in the classroom but also show social and cultural differences of teachers (Brown, Hui, Yu & Kennedy, 2011). For this reason, its important to put a clear emphasize on teachers’ beliefs and practices of assessment for educational polices are implemented and practices through those teachers (Brown et al., 2011).

Much research on teachers’ and students’ conceptions of assessment has been conducted so far (Azis, 2012; Brown 2002, 2004, 2006, 2008; Moioinvaziri, 2015). Brown (2002) has been uniquely studied teacher’s conceptions of assessment and purposes of assessment for learning and teaching processes. He argued and identified the purposes of assessment under four major purposes;

1. assessment is for improving quality of teaching and learning,
2. assessment is for making student’s learning outcomes accountable,
3. assessment is for accounting teachers and schools, and
4. assessment is for no purpose, useless.

Then, four major purposes of assessment will be identified and explained in detail in the following part.

2.5.1 Improvement Conception

Any act of teaching aims to improve students learning, as assessment does similarly. Assessment provides students with what they have learnt and which path they should follow next, so it aims to assist students with enhanced learning opportunities for their “provide support for future learning” (Hornby, 2003). According to Brown (2002), “the major premise of this conception is that assessment informs the improvement of students’ own learning and improves the quality of teaching (27).

Assessment should provide students with improved learning results as well as give opportunity to certify their learning outcomes (Brown et al., 2009); hence, “assessment needs to be understood or used in ways that contribute to the

improvement of teaching and learning” (240). Likewise, any assessment method regardless of its formal or informal basis, should enhanced teachers teaching efficacy and should aid students to boost their individual learnings (Harris & Brown, 2009).

2.5.2 Student Accountability

Assessment has long been understood and used as either assessment of learning (summative) or assessment for learning (formative), hence the primary and major premise of assessment has become checking students learning outcomes and their future learnings. Similarly, use of assessment for the purpose of accountability of student’s improvement is common. According to Brown (2002), students’ accountability through assessment means that “the students are individually accountable for their learning through their performance on assessment” (p. 40). Additionally, it places students into certain groups considering their qualification in a class (Brown, 2004), ratify students’ learnings and make students be sure what parts have been learned and what parts should be learned and mastered flowingly (Brown, Hui, Yu, & Kennedy, 2011). In a nutshell, student accountability means how assessment is used to check students’ performance based on pre-established criteria. (Moiinvaziri, 2015).

Brown (2002) asserts that “student accountability is largely about high stakes consequences such as graduation or selection or being publicly reported on as earning a certain grade, level, or score” (p. 41). This is mainly seen as allocating grades to students, evaluating their performance outcomes and placing accordingly into groups based on pre-determined criteria, and also giving some qualification examination for either graduation or passing to higher level of education (Brown, 2004). Motivating and encouraging learners to take part in self-learning and grading them accordingly is one of the most important aspect of accounting students’ own learnings (Brown, 2002).

Even though students aware that assessment improve learning and assess how well schools are doing, their belief over the use of assessment for making students accountable is undisputed (Brown & Hirschfeld, 2008). In an another study about the use of assessment for student accountability purpose, it was concluded that it is not an astonishment to find out preferred assessment methods by which they

boost their grades, more generously, and scaling up their learning. (Brown & Hirschfeld, 2007).

2.5.3 School Accountability

Accountability and credibility of a school relatively have an important role in education process. Families, inclusively in Turkey, are eager to see their students in schools that credit higher successes in high stake national examinations. Hence, using assessment for the purpose of evaluating the performances of schools is of high importance. According to Moiiinvaziri (2015), school accountability means “the use of assessment to see how well teachers or schools are doing in relation to the established standards” (p. 76).

Brown (2002) puts forward mainly two provisions of school accountability use: one is indication of quality of instruction in a school, and the other is the improvement of quality of education. Similarly, school accountability might be a precursor to improve the quality of educational principles by which students enhance their ability to get better qualification and grasp perception of their achievements (Brown, 2004).

2.5.4 Conception of Irrelevance

The notion of ‘irrelevance’ means that assessment has no consistent place and no benefits in educational context, and students, teachers and all shareholders are affected adversely when applied. Brown (2008) states that assessment, mostly known as assessing students’ performances formally, has no valid place in classroom use. The conception of irrelevance stems from the view that the process of outer checks of students’ performances are not precise, accurate, clear and concerned to teacher’s capabilities to help and improve students learning (Brown, Lake & Matters, 2011). In his study, Brown (2002) asserts that;

The premise of the fourth conception of assessment is that assessment, usually understood as a formal, organized process of evaluating student performance, has no legitimate place within teaching and learning. Teachers’ knowledge of students based on long relationship and their understanding of curriculum and pedagogy preclude the need to carry out any kind of assessment beyond the intuitive in-the-head process that occurs automatically as teachers interact with students (43).

Assessment is rejected for its thought that it reduced time allocated for instruction (Smith, 1991). Moreover, he also included that testing programs cause limitation on time for instructions, bound teachers’ abilities to teach the course content and

benefit from different approaches and materials that are not related to the testing format and tight curricular opportunities and manners of instructions.

“Beliefs about the emotional impact of testing on young children generate feelings of anxiety and guilt among teachers” (Smith, 1991, p. 9). In the same line, assessment is also rejected or appeared to be irrelevant because students consider it as being wicked and vain (Brown et al., 2008) Brown et al. (2008) in their study asserts that regardless of their grades, students consider assessment as being unequal, poor and unrelated for themselves. Teachers also, to some extent, are affected by limitations of testing based- classes when dealing with their own teachings methods and related curricula. According Brown (2002), assessment has a destructive power on teachers’ autonomy and their personal professionalism for the unique purpose of teaching, hence teachers intuitive reasoning should be considered and used instead of assessing students’ performance formally (Harris & Brown, 2009).

2.6. Teachers’ Roles in Assessment

Teachers role in the classroom keeps a pivotal role since they deal with a range of issues from teaching to assessment and the like. Both success and failure of teachers during the process of teaching mostly stems from the fact how they use their roles, responsibilities and power as a teacher (Sünbül, 1996). According to her, besides providing students with necessary information teaching, assessment also falls into teachers’ area of roles and responsibilities. In his study, Heritage (2007) counts knowledge of assessment as one of the four critical elements of any teachers’ knowledge.

Formative assessment, mostly referred as assessment for learning, provides students with necessary feedbacks. According to Heritage (2007), “effective feedback from teachers provides ...how they (students) can move forward” (p. 142), an also “it is seen that formative assessment feedback is essential to encourage the kind of ‘deep’ learning desired by tutors” (Higgins, Hartley & Skelton, p. 53). Teachers feedbacks plays a significant function in students’ motivation and their sense of self-sufficiency which has a greater influence on learning (Heritage, 2007). Besides mere feedback of formative assessment through teachers, higher order skills of students such as monitoring, planning or

evaluating their own works is also shaped and designed by teachers' knowledge of meta-cognitive strategy (Song & Koh, 2010)

In a study with 35 Iranian teachers from different secondary schools, Saad, Sardareh and Ambarwati (2013) unearthed that students are eager to accept pivotal role of teachers and their beliefs in assessment despite adverse effect of top-down managerial process to assessment. Sünbül (1996) puts forward that if classroom teachers carry out following roles and responsibilities effectively, they can promote deeper learning and raise successful students. These roles and responsibilities are;

1. fostering evaluation tools fitting students aims to attain their objectives,
2. applying assessment tools,
3. grading and,
4. assessing relevant assessment program.

2.7. Research Studies Conducted on Conception of Assessment

Many studies have been implemented to reveal different purposes of assessment in different cultures and contexts (Azis, 2015; Brown et al. 2009; Moinvaziri, 2015; Peterson & Irving, 2008). Moinvaziri (2015) applied a questionnaire to 147 university students in Iranian context. The results showed that most of them thought that assessment was used for the aim of improving quality of teaching and learning.

Azis (2015) investigated the conceptions of assessment of 107 English junior high teachers in Indonesian context. In his mixed method study, participants were given a questionnaire and semi-structured interviews. The results indicated that participants believed that the aim of the assessment was to improve teachers' teaching and students learning. It also unearthed that they were willing to use practices of assessment to help and improve their own classroom teaching.

In Hong Kong context, almost 300 teachers from primary and secondary schools were given Teacher's Conception of Assessment inventory and Practices of Assessment inventory. The results were strongly and clearly related to use of assessment to improve teaching. It was seen that Hong Kong teachers believed to

improve their students learning outcomes by using assessment practices (Brown et al., 2009).

In an another study, Azis (2012) reviewed many studies which were conducted on teachers' conception and practices of assessment. After close examination of studies from six different countries, it was concluded that assessment and learning are interrelated and it provide students with learning improvement.

Peterson and Irving (2008) had a study on 41 of 8 and 9 grades of students in New Zealand context. Students were divided into five focus groups each including 6 to 10 students. The study was an exploratory study and aimed to explore students' conceptions on purposes of assessment and feedback. Definition, purpose and personnel response were the three key parts of assessment and feedback addressed in the focus groups. Students asserted that any kinds of assessment had a following purpose, and the main purpose of assessment was supplying feedback to students that was benefited to coach students to improve their learning.

Brown and Michaelides (2011) revealed that "conceptions of assessment were positively correlated with the improvement purpose, suggesting that in both jurisdictions, teachers believe that good schools improve learning" (p. 321). Invariably, it is inferred that, classroom assessment gets students, teachers and schools to be accountable for what they carry out (Brown & Hirschfeld, 2007). In Hong Kong, not only school administrators but also parents believes that education in good schools result in much better grade outcomes in examinations (Brown et al, 2011). Brown (2004) conducted a study with 525 teachers and manager in New Zealand context. He sought four main purposes of assessment with 50 item (COA- III). He concluded that participants agreed with school accountability conception and besides irrelevance, all three purposes are positively related.

2.8. Assessment Practices and Conception of Assessment in Turkey

Assessment conceptions, policies and practices plays a significant role in Turkey since high stakes tests are required not only to be accepted to a higher education institution or to be employed into any state-hold job position and the like. In a study with 242 teachers from different fields in 2012, Gelbal and Kelecioğlu unveiled that

most of the teachers would rather to use traditional method over others to identify their students' levels of achievements and needs. Additionally, teachers felt more secure and qualified with traditional assessment practices. Similarly, Birgin and Baki (2009) investigated assessment preferences of randomly selected 975 primary school teachers from different settings. They revealed similar results with Gelbal and Kelecioğlu's work: teachers are most proficient with traditional assessment techniques, but not alternative assessment. They proposed that teachers had better to be provided with required in-service training for alternative assessment practices for new curriculum and educational reform demands so.

72 pre-service teachers from different educational fields were conducted by Tatar and Murat (2011) to find out their beliefs over assessment needs and practices. 14 different metaphors were used to unveil their assessment preferences (diagnostic, summative and formative). Even if perceptions toward formative and summative were equal, participants had by far the most opted for diagnostic assessment. They asserted that it was vital to determine students' needs just before teaching to start so that possible instruction should be shaped and focused according to their poor sides.

Vardar (2010) sought for revealing participants' conceptions of assessment in Turkish context under for main purpose: improvement, student accountability, school accountability and irrelevance. She unlocked that the highest score for students' accountability and the lowest was for irrelevance. Student accountability may be due to competitive nature of Turkish education system and irrelevance conception might originate due to the important place of high stake testing in mainstream education.

Zaimoğlu (2013) sought to bring into open teachers' conceptions of assessment based on different criteria such as gender, years of education, undergraduate institution that they graduated. According to statistical results, improvement conception had the highest value and irrelevance had the lowest generally. It was found out that gender and education level played an important role for school accountability whereas their undergraduate institutions accounted for improvement. It was also unveiled that teachers believed and were aware that assessment played a key role for not only the quality of instruction but also improvement of students learning in the classroom.

Pre- service English language teachers' conception of assessment was also studied by Yüce (2015). She also found out that pre-service English language teachers mostly agreed with improvement conception. They also believed that school accountability played a second importance for effective learning results but most teachers saw "irrelevance" as something bad. They also insisted that assessment outcomes should be reliable, objective and non-contradictory.

3. METHODOLOGY

3.1. Introduction

This study aims to unveil pre-service English teachers' conceptions of assessment in Turkish context and their tendency to use assessment for any purposes including student improvement, school or student accountability and assessment as irrelevance. Accordingly, this chapter was designed and organized to present research design, setting, participants and instrumentation, research questions, procedures for data collection and analysis.

3.2. Research Design

The research is conducted by applying quantitative research procedures. Pekrun, Goetz, Titz and Perry (2002) state that "quantitative measures are needed for more rigorous tests of hypotheses" (p. 94). They also assert that quantitative assessment works more properly and precisely when we need clear understanding of cause and effect relation. Even though this study doesn't aim to reveal a causal relation, quantitative method will be a good tool to examine how different variables may influence participants' views on assessment.

In this study, survey method has been applied to collect teacher candidates' conceptions of assessment. Survey is defined as a technique by which necessary information is collected by asking questions to a sample. Similarly, survey study research is gathering data from a sample of population to confirm present conditions according to different variables (Fraenkel, Wallen & Hyun, 1993). In this respect, a TCOA-III A- Version 3- Abridged scale including 27 items was utilized to collect data in the current study.

Cross sectional survey was used to collect demographic data from participants. Due to design of the instrument and time limits of the study, cross sectional survey was preferred over longitudinal survey design.

3.3. Research Questions

The main purpose of the current study is to explore pre-service English as a foreign language (EFL) teachers' levels of conceptions of assessment, and why they believe and use assessment out of four purposes of conceptions of

assessment. The present study also seeks to find out the effects of such independent variables such as grade, success, age, gender and years of English learning on the participants' understanding of assessment conceptions. To this end, the following research questions were articulated to guide the present study:

1. What are pre-service English teachers' conceptions of assessment?
2. How participants' conceptions of assessments relate to each other?
3. Are there any significant differences in the participants' conceptions of assessment regarding;
 - a. Gender
 - b. Years of Learning English
 - c. Age
 - d. Grade Point Average (GPA)
 - e. Grade Level

3.4. Variables

3.4.1. Dependent Variables

Conception of assessment: Conception of assessment is the main dependent variable of the study and it includes four levels (subscale): improvement, school accountability, student accountability and irrelevance. Each level tries to assess how pre-service English teachers conceive assessment. The higher level of mean scores for each levels indicates that the higher pre-service English teachers have agreement on that conception level. Additionally, each level (subscale) stands for different (dependent) variable which has interval level of measurement.

3.4.2. Independent Variables

Age: Age is one of the independent variables of the study through which it is aimed to see whether age has any effect on pre-service teachers' conceptions of assessment or not. It is a categorical variable with nominal scale. In the study, age is divided into two subcategories: twenty or less and twenty-one or more.

Gender: Gender is an independent variable by which it is aimed to examine any possible effect of gender difference on participants' conception of assessment. Gender is a categorical variable which has nominal scale.

Grade point average (GPA): Grade point average is an independent variable which is used to investigate whether overall success or failure has any effect on participants' conceptions of assessment. GPA is a categorical variable with ordinal scale.

Years of learning English: Years of learning English is another independent variable which asks for how many years the participants have spent learning English and how English learning background affects their conceptions of assessment. This is a continuous variable and it has ratio level of measurement. In this study, years of learning English has divided into five groups as less than 10 years, 10 years, 11 years, 12 years and 13 and more.

Grade: Grade is the last independent variable which ask student their grade levels (second, third or four) and seeks to reveal how different levels affects their conceptions. Grade is a categorical variable which has nominal scale.

3.5. Setting and Participants

3.5.1. Setting

This study was conducted at Hacettepe University in Ankara, Turkey. Second, third and fourth grade students of English Language Teaching Department participated in the study. Hacettepe University, is a state-hold university, is one of the oldest and prestigious universities of Turkey. Its graduates –of inclusively Faculty of Education- have always played a significant role in mainstream education (primary, secondary or university level) and acted as a role model. English Language Teaching Department has a long history and its thousands of graduates have always played an effective role in any level of mainstream education. For this purpose, the thought that finding out these teacher candidates conceptions about assessment and their purposes of using assessment will likely to reveal some important clues not only for today's understandings but also for future applications, since beliefs can affect one's behaviors to a high degree.

3.5.2. Participants

204 pre-service English language teaching department students who were studying at Hacettepe University participated in the study. The female participants outnumbered male participants; namely, 55 of them were male and 149 of them were female, due to usual female dominance in faculties of education in Turkey.

Participants were selected by both convenience sampling – the researcher had an easy access to them – and also by purposeful sampling to provide that all the participants had taken “Assessment and Evaluation” course already. Participants were 2nd (sophomore), 3rd (junior), and 4th (senior) grade students and their age’s ranged from 18 to 25 utmost. All the students were taken “Assessment and Evaluation” course just or before 2015- 2016 spring term. Participants had at least five years of English learning background and more, and a few of them had more than fifteen years of experience. After applying test of normality to the data, five outliers (histogram and q-q plot results) were deleted in order to consolidate normal distribution of the data.

Table 3.1: Demographics of Participant Pre- Service English Teachers

Variables	<i>n</i>
Age	
18	3
19	36
20	61
21	48
22	38
23	15
24	2
25	1
Gender	
Female	149
Male	55
Grade	
Sophomore	90
Junior	74
Senior	40
Years of English Education	
Less than 10 years	27
10 years	49
11 years	44
12 years	48
13 years or more	36
GPA	
3.01 – 4.00	156
2.00 – 3.00	48
Total	204

3.6. Instrumentation

To gather the data, an instrument called “Teachers’ conceptions of assessment inventory--Abridged (TCoA-III A Version 3-Abridged)” were used and applied to collect data for the study. This inventory includes 27 items, which was the shorter version of original “Teacher Conception of Assessment” inventory that was developed and used by Brown (2001, 2003). The inventory was in Likert scale format ranging from 1 (strongly disagree) to 6 (strongly agree). Participants were asked to prefer one out of six (strongly disagree, mostly disagree, slightly agree, moderately agree, mostly agree and strongly agree) options and to respond to each item separately. The higher value they responded to an item means that the higher they agreed to this specific statement or level regarding their assessment conceptions. By the way, participation was voluntary and each participant was given a “Voluntary Participation Form” before delivering the inventory. The inventory was in hand-out format and it was given just before the planned course started. Instructors were informed at least one day in advance. Similarly, participants were delivered necessary information including aims of the study, time allocation, voluntariness and the like. They, participants, were also made sure about confidentiality issue and they were informed that a copy of the study results would be delivered to them if they preferred to have. The data were collected in April and May, 2016 and each student-teacher filled out once; namely, cross-sectional survey method was used.

Reliability analysis was also performed for the scale. As stated beforehand, TCoA-III A-Version 3-Abridged Inventory includes 4 conceptions levels including a total of 27 items. These levels are improvement conception (12 items), school accountability conception (3 items), student accountability conception (3 items) and irrelevance conception (9 items). The inventory has a 6-point Likert scale format ranging from strongly disagree to strongly agree. All the essential validity and reliability procedures were already checked (Brown, 2007). The alpha values computed with the data for this study are presented in the Table 2.

Table 3.2: Alpha values per level

Purposes	Alpha
Improvement	.87
School Accountability	.61
Student Accountability	.48
Irrelevance	.52
Total	.83

For the reliability of the inventory for this study, Cronbach's alpha coefficients were calculated as 0.83 for the inventory in total, for the first level (improvement) .87; for the second level (school accountability) .61; for the third level (student accountability) .48; and for the fourth level (irrelevance) .52. Even though some of the levels' values indicated slightly lower reliability value, overall value indicated a satisfactory level of reliability.

3.7. Data collection procedures

The data was collected during April and May of 2015- 2016 academic year's spring semester at Hacettepe University. 204 of English Language Teaching department students from 2nd, 3rd and 4th grades participated. Before collecting the data, the owner of the scale was informed about the aim of the study and he was asked for a permission to use the scale. After the permission was granted via email, which includes necessary permission of use, conditions of use and rules of citation, The Ethical Committee of Institute of Educational Sciences of Hacettepe University was delivered required documents including scale and its permission, voluntary participation form, and form of ethical committee permission authorization of the thesis study and was asked to collect data and carry out the thesis. After all the permissions were granted and authorization was taken, the data started to be collected at Faculty of Education. The data was collected during normal class time and they were given "Teachers conception of assessment inventory Abridged (TCoA-III A- 3 Abridged" inventory and a "Voluntary participation form" together. Before delivering survey and voluntariness form, students were provided with aims of the study, concise information about the forms, timing of the surveying, and confidentiality of their returns. They were also made sure that a copy of study's results section would be provided to them if they

were interested in the study much deeper. The class teachers were also informed about the study and data collection at least one day in advance, and necessary permissions were taken in order to use a short while before their normal class started. 3 different sections from 2nd and 3rd grades and 2 different sections from 4th grades students were included in the study. The time for the collection of data lasted from 15 to 20 minutes. After the collection of the data, teacher and students of each sections were informed about confidentiality once again and appreciated.

3.8. Data analysis procedures

The data was entered to Statistical Package for the Social Sciences (SPSS 23) software program in order to check and reveal frequencies and descriptive results out of data. Before proceeding to descriptive statistics, the data was investigated for missing values and no missing values were detected. Then, the data was explored in order to see distribution of data (parametric or non-parametric data), because of the fact that distribution of data (normal or non-normal distribution) leads to totally different analysis methods. Even though test of normality results showed non-normal distribution according to Kolmogorov- Smirnov results (Sig = .052, Sig- IMP= .005, Sig-STCCA= .000, Sig-SCCCA= .000, Sig-IRR= .005) due to size of the sample, histogram and q-q plot results (please see appendix) clearly indicated that the data was normally distributed. In order to consolidate test of normality results, 5 outliers out of 204 participants were deleted. Then, reliability analysis was performed for the scale. Cronbach's alpha coefficients was computed as 0.83 for the inventory. This result demonstrated that the inventory and its items had a satisfactory level of reliability.

After test of normality was conducted and reliability analysis was computed, the data was subjected to descriptive statistics. Mean values for each item and each subscale (improvement, school accountability, student accountability and irrelevance) were calculated and interpreted. Higher mean value for each item or subscale indicated that participants had higher level of agreement with that specific conception or vice versa. After descriptive statistics were computed and mean values were interpreted for general conceptions of assessment values and for each dependent variables (improvement, school accountability, student accountability and irrelevance), the data were investigated by using Pearson product-moment correlation coefficient in order to investigate relations (strong,

medium, and small) among each dependent variable separately and to reveal the direction of the relation (positively or negatively correlated). In order to ensure assumptions of normality and linearity, preliminary analysis was conducted before the data was computed. Then the output data results were interpreted.

This analysis steps were followed by Multivariate Analysis of Variance test, because there was more than one dependent variable in this cases. Therefore, Multivariate Analysis of Variance (MANOVA) was preferred over Independent sample-t test and Analysis of Variance (ANOVA) test because the latter two tests were required multiple statistical analysis which might cause low reliability of the results. In such a case, the probability of facing Type 1 error, finding significant differences after multiple analysis although there was no statistically significant difference in reality, might become powerful. Before proceeding to analysis of data according to MANOVA test, the data was investigated to reveal whether the data met all the assumptions of MANOVA or not. Firstly, outliers were checked and five outliers were excluded from out of 204 participants to ensure normality. Secondly, The Mahalanobis distance were calculated and it was seen that it provided multivariate normality ($MD = 15.86$). Thirdly, assumption of linearity was satisfied according to linearity analysis. Then, the assumption of multicollinearity and singularity were satisfied according to correlation between dependent variables since there were correlation up around .8 according to Pallant (2010). Followingly, Box's Test of Equality of Covariance were performed to check whether the data violates the assumption of homegenity of variance-coveriance matrices, and also Levene's Test of Equality of Error Variance were applied to inspect whether the data violates the assumption of equality of variance or not. It is known that if the *Sig* value is larger than .001, then it means that there is no violation from the assumption of homogeneity of variance-coveriance matrices. After all assumptions were met, the data was subjected to Manova test. All the assumption was investigated for each dependent variable before their Multivariate test results and Wilks' Lambda values were taken into consideration. If the dependent variable met all the assumption, then Multivariate test's results and Wilks' Lambdas' were calculated, checked and interpreted.

4. RESULTS

4.1. Introduction

The chapter is designed to present analysis of the data and research findings. In order to compute, analyze and explore data for further investigation and interpretation, descriptive statistics, correlation and multivariate analysis tests were performed successively.

4.2. Results of Data Analysis

4.2.1. What are pre-service English teachers' conceptions of assessment?

The question tries to investigate and determine what the purposes of pre-service English teachers are in order to perform assessment. Namely, it seeks to reveal their conceptions of assessment and its levels/ purposes (improvement, school accountability, student accountability and irrelevance). The table presents descriptive statistics for each component of Teacher Conceptions of Assessment Abridged Scale (TCoA- IIIA- Version 3- Abridged). The scale includes values from 1 (minimum) to 6 (maximum) for each response.

Table 4.1: Levels of conception of assessment of TCoA-III A, Version 3- Abridged Scale (N=199)

Conception of Assessment Purposes	N	M	SD
<i>Improvement</i>	199	4.24	.70
<i>School Accountability</i>	199	4.02	.75
<i>Student Accountability</i>	199	3.75	.94
<i>Irrelevance</i>	199	3.58	.55

As shown in the table, four levels of conceptions of assessment are included and presented in the TCoA- IIIA Scale. Improvement conception (M= 4.24, SD= .70) has the highest rank and agreement level among all variables and is followed by student accountability (M= 4.02, SD=.75). Improvement and student's accountability conceptions have a moderate agreement level among all variables. Conception of irrelevance (M=3.58, SD=.55) holds the lowest mean value of all variables and is considered around a moderate disagreement level among all the variables.

Table 4.2: Pre-service teacher's improvement level of conception of assessment

Item	Improvement Conception	N	M	SD
IMP3	Assessment is a way to determine how much students have learned from teaching.	199	4.41	1.24
IMP4	Assessment provides feedback to students about their performance.	199	4.75	1.10
IMP5	Assessment is integrated with teaching practice.	199	4.32	1.09
IMP6	Assessment results are trustworthy.	199	3.59	1.12
IMP12	Assessment establishes what students have learned.	199	4.23	1.06
IMP13	Assessment feeds back to students their learning needs.	199	4.53	.90
IMP14	Assessment information modifies ongoing teaching of students.	199	4.25	.98
IMP15	Assessment results are consistent.	199	3.51	1.15
IMP21	Assessment measures students 'higher order thinking skills.	199	3.37	1.16
IMP22	Assessment helps students improve their learning.	199	4.35	1.06
IMP23	Assessment allows different students to get different instruction.	199	3.99	1.20
IMP24	Assessment results can be depended on.	199	4.06	1.18

As seen in the table, pre-service English teachers highly agree with the statement “Assessment provides feedback to students about their performance” (M =4.75, SD=10) among improvement conceptions. It can be inferred from the mean values that assessment acts to provide feedback (formative) to learners as Brown (2003) stated formative nature of improvement purpose of assessment. This “feedback” nature of assessment is also backed up by the following statement; “Assessment feeds back to students their learning needs” which is second in rank (M=4.53, SD=.90). It can be inferred from the results that students “mostly and moderately agree” with feedback part of assessment to improve their learning.

Table 4.3: Pre-service teacher's school accountability level of conception of assessment

Item	School Accountability Conception	N	M	SD
SCACC1	Assessment provides information on how well schools are doing.	199	4.21	1.21
SCACC10	Assessment is an accurate indicator of a school's quality.	199	3.49	1.30
SCACC19	Assessment is a good way to evaluate a school.	199	3.56	1.26

The above table shows that pre-service English teachers mostly agree with the statement “Assessment provides information on how well schools are doing” (M=4.21, SD=1.21). The mean values demonstrate that pre-service English teachers “moderately agree” on that assessment provides enough information about the current situations of schools running (whether doing well or not). Secondly, even though pre-service teachers are slightly above a moderate disagreement level, it can be deduced from the table that assessment could also be used in order to check and assess schools’ performances (M=3.56, SD=1.26).

Table 4.4: Pre-service teacher’s student accountability level of conception of assessment

Item	Student Accountability Conception	N	M	SD
STACC2	Assessment places students into categories.	199	4.0754	1.09
STACC11	Assessment is assigning a grade or level to student work.	199	4.0653	1.01
STACC20	Assessment determines if students meet qualifications standards.	199	3.9296	1.10

As indicated in the table, pre-service English teachers agree mostly with the statement “Assessment places students into categories” within the student accountability conception (M=4.07, SD1.09). Namely, assessment is used to group students into different levels such as high, medium and low achievers. Similarly, they also “moderately agree” on that assessment is used to grade students’ performance (M= 4.06, SD=1.01). Therefore, it can be concluded that pre-service teachers agree on the (required) roles of assessment in categorization and evaluation of their performances.

Table 4.5: Pre-service teacher’s irrelevance level of conception of assessment

Item	Irrelevance Conception	N	M	SD
IRR7	Assessment forces teachers to teach in a way against their beliefs.	199	3.14	1.34
IRR8	Teachers conduct assessments but make little use of the results.	199	3.72	1.32
IRR9	Assessment results should be treated cautiously because of measurement error.	199	4.83	1.07
IRR16	Assessment is unfair to students.	199	2.97	1.33
IRR17	Assessment results are filed & ignored.	199	3.12	1.26
IRR18	Teachers should take into account the error and imprecision in all assessment.	199	4.56	1.12

IRR25	Assessment interferes with teaching.	199	3.64	1.27
IRR26	Assessment has little impact on teaching.	199	2.75	1.26
IRR27	Assessment is an imprecise process.	199	3.51	1.12

As presented in the Table 7, pre-service teachers mostly agree with the statement “Assessment results should be treated cautiously because of measurement error.” of irrelevance conception (M=83, SD= 1.07). Similarly, pre-service teachers also highly (second in the rank) agree with statement “Teachers should take into account the error and imprecision in all assessment” (M= 4.56, SD= 1.12). It is interesting to see that even though pre-service teachers, in general “moderately agree” with other levels of conceptions as shown in the Tables 5, 6 and 7, they also “mostly agree” on that assessment processes (measurement, errors, imprecisions etc.) should seriously be taken into account to benefit from it; otherwise, it could be seen as irrelevant to teaching and learning process.

4.2.2. How do levels of conceptions of assessment relate to each other?

The question “How do levels of conceptions of assessment relate to each other” was asked to investigate the relations between each levels of the dependent variable and the direction of correlation (positive or negative). In order to interpret the relationships, the following table was presented.

Table 4.6: Relationship between levels of conceptions of assessment

Inventory	1	2	3	4
1. Improvement	-			
2. School Accountability	.694**	-		
3. Student Accountability	.554**	.591**	-	
4. Irrelevance	-.146*	-.090	.047	-

** $p < 0.01$ level (2-tailed).

* $p < 0.05$ level (2-tailed).

The relationships among different levels of conceptions of assessment was investigated by using Pearson product-moment correlation coefficient. In order to provide insurance to assumptions of normality and linearity, preliminary analyses were performed. There were strong, positive correlations between improvement and school accountability levels, $r = .69$, $n = 199$, $p < .05$ with 48, 23% variance of

the coefficient of determination, and between improvement and student accountability conceptions, $r = .55$, $n = 199$, $p < .05$ with a 30, 64% variance of the coefficient of determination. There was also a strong, positive correlation between school accountability and student accountability, $r = .59$, $n = 199$, $p < .05$ with 34,92% variance of the coefficient of determination. Improvement and irrelevance conceptions were negatively correlated with a small degree of relationship, $r = -.14$, $n = 199$, $p < 0.5$ with a -2.13% variance of the coefficient of determination.

4.2.3. Are there any significant differences in the participants' conceptions of assessment regarding different variables;

a. Gender

b. Years of Learning English

c. Age

d. Grande Point Average (GPA)

e. Grade levels (2nd, 3rd, 4th grades)

The above questions were asked to examine whether individual differences such as gender, years of learning English, age and grand point average, and grade levels make any statistically significant difference on pre-service English teachers conceptions of assessment. In this part, Multivariate Analysis of Variance test was applied for each dependent variable and for each individual difference, and the statistical results were presented.

4.2.3.1. Gender

The statistical analysis was performed to see whether there was a significant difference between gender difference and assessment conception. At first, descriptive statistics were conducted to make sure that the data had more cases in each cell than the number of dependent variables. It was seen that there was no violation of assumption 1 which means having no violations of normality and equality. Then, Box's Test of Equality of Covariance and Levene's Test of Equality of Error Variance were performed to check whether the data violates the assumption of homogeneity of variance-covariance matrices, and the assumption of and equality of variance or not. Box's M results $F = (10, 45382.064) = .720$, $p <$

.05 indicated that the data had no violation of the assumption of homogeneity of variance-covariance matrices.

Table 4.7: Levene's Test of Equality of Error Variances

Purposes	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
<i>IMP</i>	.040	1	197	.841
<i>SCACC</i>	1.423	1	197	.234
<i>STACC</i>	.304	1	197	.582
<i>IRR</i>	.062	1	197	.803

As shown in the table, Levene’s test results demonstrated that none of the *p* values are less than .05 which indicated that the data also met the assumption of equality of variance for each variable. After the assumptions are met, descriptive statistics were used to check mean differences of conception of assessment based on gender.

Table 4.8: Descriptive statistics of dependent variables for male and female participants

Gender	<i>IMP</i>			<i>SCACC</i>		<i>STACC</i>		<i>IRR</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Male</i>	53	4.11	.69	3.88	1.01	4.08	.75	3.68	.54
<i>Female</i>	146	4.24	.71	3.71	.91	4.00	.74	3.54	.55

The descriptive values were computed to reveal mean differences of pre-service English teachers’ conceptions of assessment regarding gender differences. As shown in the table, participants mean values are slightly different for each dependent variables, therefore a multivariate tests of significance were conducted further to see whether the mean differences were statistically significant.

Table 4.9: Wilks’ Λ for differences in conception between male (n=53) and female (n= 146) participants

	<i>Wilks’ Λ</i>	<i>F(4, 184)</i>	<i>p</i>	<i>Partial eta²</i>
Gender	.976	1.18	.31	.024

p= .05

A one way between groups multivariate analysis of variance was performed to investigate gender differences in conceptions of assessment. Four dependent variables were used: improvement, school accountability, student accountability and irrelevance. Preliminary assumption testing was conducted to check for

normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There were no statistically significant differences between males and females on the combined dependent variables, $F(4, 184) = 1.18, p = .319$; Wilks Lambda = .97; partial eta squared = .02.

4.2.3.2. Years of Learning English

This question was asked in order to see whether there was a significant difference between years of learning English and assessment conceptions. At first, descriptive statistics were conducted to make sure that the data had more cases in each cell than the number of dependent variables. It was seen that there was no violation of assumption 1, which means having no violations of normality and equality. Then, Box's Test of Equality of Covariance and Levene's Test of Equality of Error Variance were performed to check whether the data violate the assumption of homogeneity of variance-covariance matrices, and the assumption of and equality of variance or not. Box's M results calculated as $F = (40, 54777.594) = 1.051, p < .05$ indicated that the data had no violation of the assumption of homogeneity of variance-covariance matrices.

Table 4.10: Levene's Test of Equality of Error Variances

Purposes	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
<i>IMP</i>	1.214	4	192	.306
<i>SCACC</i>	1.494	4	192	.206
<i>STACC</i>	.266	4	192	.899
<i>IRR</i>	1.555	4	192	.188

As shown in the table, Levene's test results demonstrated that none of the *p* values are less than .05 which indicated that the data also met the assumption of equality of variance for each variable. After the assumptions were met, descriptive statistics were used to check mean differences of pre- service English teachers' conceptions of assessment regarding their years of learning English.

Table 4.11: Descriptive statistics of dependent variables for participant's years of English education

Education	<i>N</i>	<i>IMP</i>		<i>SCACC</i>		<i>STACC</i>		<i>IRR</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Less than 10 years</i>	26	4.07	.56	3.85	.90	4.06	.71	3.53	.60
<i>10 years</i>	49	4.19	.65	3.86	.80	4.14	.69	3.73	.48
<i>11 years</i>	42	4.25	.76	3.81	1.03	4.00	.87	3.49	.54

12 years	44	4.04	.70	3.66	.94	3.95	.76	3.53	.62
13 years or more	33	4.11	.82	3.52	1.06	3.92	.70	3.60	.51

The descriptive values were computed to reveal mean differences of pre-service English teachers' conceptions of assessment regarding years of learning English. As shown in the table, participants mean values are slightly different for each dependent variables, therefore a multivariate test of significance was conducted to further explore whether the mean differences were statistically significant or not.

Table 4.12: Wilks' Λ for differences in conceptions between education years; Less than 10 years (n= 26), 10 years (= 49), 11 years (n=42), 12 years (n=44), 13 years or more (n=33) of participants

	<i>Wilks' Λ</i>	<i>F(16, 578)</i>	<i>p</i>	<i>Partial eta²</i>
Education	.930	.86	.611	.018

p= .05

A one way between groups multivariate analysis of variance was performed to investigate English learning time differences in conceptions of assessment. Four dependent variables were used: improvement, school accountability, student accountability and irrelevance. The independent variable was years of learning English. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There were no statistically significant differences among participant's years of learning English on the combined dependent variables, $F(16, 578) = .86$, $p = .611$; Wilks Lambda = .93; partial eta squared = .01.

4.2.3.3. Age

The statistical analysis was performed in order to see whether there was a significant difference between participants' age difference on their assessment conceptions. At first, descriptive statistics were conducted to make sure that the data had more cases in each cell than the number of dependent variables. It was seen that there was no violation of assumption 1 which means having no violations of normality and equality. Then, Box's Test of Equality of Covariance and Levene's Test of Equality of Error Variance were performed to check whether the data violate the assumption of homogeneity of variance-covariance matrices, and the assumption of and equality of variance or not. Box's M results calculated as $F =$

(10, 184465.528) = .660, $p < .05$ indicated that the data had no violation of the assumption of homogeneity of variance-covariance matrices.

Table 4.13: Levene's Test of Equality of Error Variances

Purposes	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
<i>IMP</i>	1.633	1	197	.203
<i>SCACC</i>	.398	1	197	.529
<i>STACC</i>	.054	1	197	.816
<i>IRR</i>	1.185	1	197	.278

As shown in the Table 14, Levene's test results demonstrated that none of the p values are less than .05 which indicated that the data also met the assumption of equality of variance for each variable. After the assumptions were met, descriptive statistics were used to check mean differences of conception of assessment based on age.

Table 4.14: Descriptive statistics of dependent variables for age differences of the participants

Age	<i>N</i>	<i>IMP</i>		<i>SCACC</i>		<i>STACC</i>		<i>IRR</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>20 years or less</i>	97	4.20	.66	3.75	.91	4.04	.74	3.63	.52
<i>21 years or more</i>	102	4.09	.74	3.75	.97	4.00	.75	3.53	.58

Descriptive statistics were computed to reveal mean differences of pre-service English teachers' conceptions of assessment regarding age differences. As shown in the table, participants mean values are slightly different for each dependent variable; therefore, a multivariate test of significance was further conducted to see whether the mean differences were statistically significant.

Table 4.15: Wilks' Λ for differences in conceptions between different ages'; 20 years or less (n=97) and 21 years or more (n= 102) groups

	<i>Wilks' Λ</i>	<i>F (4, 194)</i>	<i>p</i>	<i>Partial eta²</i>
Age	.977	1.15	.33	.023

$p = .05$

A one way between groups multivariate analysis of variance was performed to investigate age differences in conceptions of assessment. Four dependent variables were used: improvement, school accountability, student accountability and irrelevance. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-

covariance matrices, and multicollinearity, with no serious violations noted. There was no statistically significant difference between different age groups on the combined dependent variables, $F(4, 194) = 1.15, p = .331$; Wilks Lambda = .97; partial eta squared = .02.

4.2.3.4. Grade Point Average(GPA)

The statistical analysis was performed in order to see whether there was a significant difference between participants' grade point average (GPA) difference on their assessment conceptions. At first, descriptive statistics were conducted to make sure that the data had more cases in each cell than the number of dependent variables. It was seen that there was no violation of assumption 1 which means having no violations of normality and equality. Then, Box's Test of Equality of Covariance and Levene's Test of Equality of Error Variance were performed to check whether the data violate the assumption of homogeneity of variance-covariance matrices, and the assumption of and equality of variance or not. Box's M results calculated as $F = (10, 35311.504) = .643, p < .05$ indicated that the data had no violation of the assumption of homogeneity of variance-covariance matrices.

Table 4.16: Levene's Test of Equality of Error Variances

Purposes	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
<i>IMP</i>	.555	1	197	.457
<i>SCACC</i>	.100	1	197	.752
<i>STACC</i>	1.068	1	197	.262
<i>IRR</i>	.001	1	197	.981

As shown in the table, Levene's test results demonstrated that none of the p values are less than .05 which indicated that the data also met the assumption of equality of variance for each variable. After the assumptions were met, descriptive statistics were used to check mean differences of conception of assessment based on GPA values.

Table 4.17: Descriptive statistics of dependent variables for grade point average (GPA) scores of the participants

GPA	<i>IMP</i>			<i>SCACC</i>		<i>STACC</i>		<i>IRR</i>	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
2.00-3.00	48	4.16	.75	3.81	.94	3.98	.83	3.69	.52

3.01-4.00 151 4.14 .69 3.74 .94 4.03 .72 3.54 .56

The descriptive values were computed to reveal mean differences of pre-service English teachers' conceptions of assessment regarding their grade point average differences. As shown in the table, participants mean values are slightly different for each dependent variables; therefore, a multivariate test of significance was further conducted to see whether the mean differences were statistically significant.

Table 4.18: Wilks' Λ for differences in conceptions between high (n=48) and medium (n= 151) achievers

	<i>Wilks' Λ</i>	<i>F (4,194)</i>	<i>p</i>	<i>Partial eta²</i>
GPA	.978	1.1077	.36	.022

p= .05

A one way between groups multivariate analysis of variance was performed to investigate grade point average differences and conceptions of assessment. Four dependent variables were used: improvement, school accountability, student accountability and irrelevance. The independent variable was grade point average values. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There were no statistically significant differences between high achievers and medium achievers on the combined dependent variables, $F (4, 194) = 1.1077, p = .369$; Wilks Lambda = .97; partial eta squared = .02.

4.2.3.5. Grade Level

The statistical analysis was performed in order to see whether there was a significant difference between participants' grade levels (sophomore, junior and senior) on their assessment conceptions. At first, descriptive statistics were conducted to make sure that the data had more cases in each cell than the number of dependent variables. It was seen that there was no violation of assumption 1 which means having no violations of normality and equality. Then, Box's Test of Equality of Covariance and Levene's Test of Equality of Error Variance were performed to check whether the data violates the assumption of homogeneity of variance-covariance matrices, and the assumption of and equality of variance or not. Box's M results, $F = (10, 35311.504) = .643, p < .05$ indicated

that the data had no violation of the assumption of homogeneity of variance-covariance matrices.

Table 4.19: Levene's Test of Equality of Error Variances

Purposes	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
<i>IMP</i>	4.686	2	196	.010
<i>SCACC</i>	.319	2	196	.727
<i>STACC</i>	.005	2	196	.995
<i>IRR</i>	1.171	2	196	.312

As shown in the table, Levene's test results demonstrated that none of the *p* values are less than .05 but *p* value of improvement is less than .05. In such as case, Tabachnick and Fidell (2007) offers to set a more conservative level for determining the significance of that specific variable, namely alpha level of .25 or .02 instead of conventional .05 level. In the above test result, improvement value is .01 which indicated that the data also met the assumption of equality of variance for each variable. After the assumptions were met, descriptive statistics were used to check mean differences of conception of assessment based on grade levels.

Table 4.20: Descriptive statistics of dependent variables for grade levels of the participants

Grades	<i>N</i>	<i>IMP</i>		<i>SCACC</i>		<i>STACC</i>		<i>IRR</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Sophomore</i>	86	4.21	.61	3.81	.91	4.06	.77	3.61	.51
<i>Junior</i>	73	4.14	.70	3.75	.93	4.87	.72	3.45	.58
<i>Senior</i>	40	4.01	.86	3.65	1.02	4.20	.71	3.73	.54

The descriptive values were computed to reveal mean differences of pre-service English teachers' conceptions of assessment regarding grade differences. As shown in the table, participants mean values are slightly different for each dependent variables, therefore a multivariate test of significance was further conducted to see whether the mean differences were statistically significant.

Table 4.21: Wilks' Λ for differences in conceptions among 2nd, 3rd and 4th grade students

	<i>Wilks' Λ</i>	<i>F (8, 388)</i>	<i>p</i>	<i>Partial eta²</i>
Grades	.906	2.45	.01	.04

p = .05

A one way between groups multivariate analysis of variance was performed to investigate grade level differences in conceptions of assessment. Four dependent variables were used: improvement, school accountability, student accountability and irrelevance. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There was a statistically significant difference between sophomores, juniors and seniors on the combined dependent variables, $F(8, 386) = 2.45$, $p = .014$; Wilks Lambda = .90; partial eta squared = .04.

Table 4.22: MANOVA for differences in conceptions of assessment based on grade levels

	<i>M</i> ₂	<i>SD</i> ₂	<i>M</i> ₃	<i>SD</i> ₃	<i>M</i> ₄	<i>SD</i> ₄	<i>F</i> (2,196)	<i>p</i>	<i>Part eta</i> ²
<i>IMP</i>	4.21	.61	4.14	.70	4.01	.86	1.04	.35	.011
<i>SCACC</i>	3.81	.91	3.75	.93	3.65	1.02	.410	.66	.004
<i>STACC</i>	4.06	.77	3.87	.72	4.20	.71	2.76	.06	.027
<i>IRR</i>	3.61	.51	3.45	.58	3.73	.54	3.70	.02	.36

When the results for the dependent variables were considered separately, none of the dependent variables reached a statistical significance using a Bonferroni adjusted level alpha level of .012. However, an inspection of the mean scores indicated that sophomores reported slightly higher levels of improvement ($M = 4.21$, $SD = .61$) and school accountability ($M = 3.81$, $SD = .91$), whereas senior students indicated slightly higher levels of student accountability ($M = 4.06$, $SD = .77$) and irrelevance ($M = 3.61$, $SD = .51$).

5. DISCUSSION

5.1. Introduction

This chapter aims to present a summary of the study and to discuss findings. All the research questions will be discussed separately in detail with reference to results section. Additionally, findings of the statistical analyses including descriptive, correlation and multivariate test of variance will be presented and interpreted.

5.2. Summary of the Study

The present study mainly aimed to investigate pre-service English teachers' conceptions of assessment. 204 pre-service English teachers participated into the study voluntarily, and an inventory named Teachers' conceptions of assessment inventory--Abridged (TCoA-III A-Version 3-Abridged) was used to collect data. The inventory was in a 6-point Likert scale format which is ranging from strongly disagree to strongly agree including 27 items. The inventory had also four levels of conceptions of assessment. These are improvement, school accountability, student accountability, and irrelevance.

The data was analyzed by using Statistical Package for the Social Sciences (SPSS 23). After missing values were detected, normality was checked and outliers were deleted, then the data was subjected to descriptive analysis in order to find out participant's agreements for the levels of conceptions of assessment. Descriptive statistics indicated that improvement conceptions have the highest value, and participants moderately agree that assessment should be used for improvement. On the contrary, conception of irrelevance has the lowest mean value among all the levels, and participants moderately disagree to see assessment as irrelevant to teaching and learning processes.

The next step was to further investigate the data to reveal the relationships between different levels of conceptions of assessment. Pearson product-moment correlation coefficient was used to reveal the relations between levels. The correlation results have shown that improvement conception was strongly correlated with both school accountability and student accountability. Similarly,

school accountability and student's accountability conceptions were also strongly correlated. However, it was found out that improvement and irrelevance conceptions were negatively correlated by having small degree of relations.

Following that, the data was subjected to multivariate test of variance in order to unearth the effects of participants' individual differences (gender, years of learning English, age, grade point average and grade levels) on their conceptions of assessment. MANOVA results demonstrated that even though there were no statistically significant differences regarding gender, years of learning English, age and grade point average, participants' grade levels made a statistically significant difference on their conceptions of assessment.

5.3. Discussion of Findings in Terms of Research Questions

5.3.1. Discussion of research question 1

The question ***“What are the pre-service teachers’ conceptions of assessment?”*** tried to reveal participants' purposes of using or understanding assessment. Four levels of conceptions; improvement, school accountability, student accountability, and irrelevance were taken into consideration. Descriptive statistics revealed that conception of assessment held the highest mean value among all the levels ($M = 4.24$, $SD = .70$), and pre-service English teachers moderately agreed that assessment should be used to improve teaching and learning. Brown (2002) stated that the aim of this conception is to “inform the improvement of students' own learning and improve the quality of teaching (p. 27). In this perspective, current study results were also seen to be in line with other studies in the literature. For example, Yüce (2015) in her study on pre-service teachers' conceptions of assessment and assessment practices revealed similar results by reporting that participants moderately agreed with conception of improvement as well. This could be because of the the fact that participants would prefer to use and benefit from assessment as a vehicle for personal improvement in their teaching and learning process. This view was consolidated by Brown and Hirschfeld's (2008) study on students' conceptions of assessment. They suggested that when students believe that assessment is organized to account their individual learning, their results tend to be increased positively. The other two conceptions of the present study; school accountability ($M = 4.02$, $SD = .75$) and student accountability ($M = 3.75$, $SD = .94$) followed improvement conceptions

successively. Furthermore, the participants almost moderately agreed with both conceptions entailing that assessment should be used for accountability. These outcomes were also consolidated by Vardar's study (2010). By investigating sixth, seven, eight grade teachers' conceptions of assessment, she reported that participants moderately agreed that assessment should be benefitted for accountability of students ($M = 3.50$, $SD = .62$). It can be concluded that accountability roles of assessment were valued by participants because of competitive nature of Turkish education system. Not only classroom based assessments (formative and summative) but also high stakes examination (LYS, YDS, ALES, etc.) holds an important role in education system for passing into another grade, getting promotion, entering university, holding a job and so on. Similarly, schools are put into ranks and categories according to their results in high-stake examinations. As a result, participants are inclined to consider accountability as an important purpose of assessment. Accordingly, conception of irrelevance held the lowest mean value of all the levels in the current study ($M = 3.58$, $SD = .55$), and participants moderately disagreed that assessment is useless for education. Seeing assessment as irrelevant could be because of either its adverse effect on teacher autonomy or the view of assessment as "equal to teaching" (Brown, 2002). Because assessment has been a backbone of the Turkish education system for years with so many cultural dimensions in the society, the view of assessment as useless, irrelevant and the like could have provoked the participants to think adversely on the issue.

5.3.2. Discussion of Research Question 2

The purpose of this question "***How do levels of conceptions of assessment relate to each other?***" was to investigate the relationships between different conceptions levels (strong, moderate, small) as well as the direction of the relations (positive, negative or none). Pearson product-moment correlation coefficient results indicated that there were strong, positive correlation between improvement and school accountability ($r = .69$), and improvement and student accountability ($r = .55$). These findings are in line with Yüce's (2015) findings. In her study regarding conceptions of assessment, she found out that there were positive and significant correlations between improvement, school and student accountabilities. Additionally, Brown and Hirschfeld (2010) stated that students

who regard assessment as a tool for personal accountability of their learnings will success more. Similarly, Vardar (2010) also presented that all three conceptions were moderately correlated besides irrelevance which held non-significant correlations with other levels. These indicated that relationships among improvement, school accountability and student accountability were strong and participants agreed that these levels affect each other positively. Similar findings in these studies could be explained by Turkish education system's realities and cultural norms. As explained earlier, Turkish education system is very competitive in its nature. Therefore, parents would like to see not only their students but also their schools accountable. Besides, students' school grades, the ranks of students and their schools in high-stake national examination play key roles on the determination of success and failure, and this leads to the conception that assessment should boost teaching and learning process as well as make this process and outcomes accountable. On the other hand, irrelevance conception was found to be sharing small or non-significant relations with other levels of assessment. Correlation results indicated that improvement and irrelevance conceptions were negatively correlated with a small degree of relationship ($r = -.14$). Similarly, school accountability and irrelevance conceptions were also negatively correlated ($r = -.09$) and held non-significant relationship with each other. These results also correspond to Vardar's (2010) study which also indicated that irrelevance conception shared non-significant relationships with other levels of conceptions of assessment. Brown (2004), in his study on teacher's conceptions of assessment, also suggested that irrelevance conception was also negatively correlated with improvement conceptions. He explained this correlation as "If teachers think assessment is about Improvement then it is unlikely they will consider assessment as Irrelevant ($r = -.69$) (p.313). Therefore, when assessment is accounted for irrelevance, it might be thought that the aim of improving teaching and learning is severed (Brown, 2004).

5.3.3. Discussion of Research Question 3

5.3.3.1. Gender

The question "***Are there any significant difference in the participants' conceptions of assessment regarding their gender difference?***" aimed to unveil any possible effect of gender difference on pre-service English teachers'

conceptions of purposes of assessment regarding four levels of conceptions. Multivariate test of variance results indicated that there was no statistically significant difference between males and females regarding their conceptions of assessment (Wilks' Lambda = .97, $p = .31$). Similar results were yielded by Zaimoğlu's (2015) study in which she found out statistically no difference between males and females as well (Pillai Trace = .20, $p = .17$). Descriptive analyses indicated slightly different values for males on females for their conceptions of assessment, but their agreement levels for each conception was the same according to descriptive results. It was seen that both males and females were inclined to see assessment as a tool for improvement of teaching and learning with a moderately agreement level. In this respect, Zaimoğlu (2015) concluded that "whatever teachers' gender is, they give importance to the function of assessment, which improves teaching and students' learning" (p.55). Similarly, student accountability conception held a moderately agreement level by both male and female participants as well. Brown et al. (2011) found out strong correlation of accountability with improvement conception in Chinese context. They asserted that this was because of the policy and tradition, which drive assessment to improve quality of teaching and student learning. This could be echoed to current research as well. Regardless of the gender difference, pre-service teachers preferred to see assessment as a vehicle of accountability and improvement due to Turkish traditions and educational policies as explained earlier. Irrelevance conception held the lowest values for both males and females that they disagreed with irrelevant view of assessment. As a result, gender has made very limited difference on pre-service teachers' conceptions of assessment, rather it was concluded that participants perceived purpose of assessment as a tool to improve and account their learning and quality of teaching regardless of gender difference.

5.3.3.2. Years of learning English

The aim of the question "***Is there any significant differences in the participants' conceptions of assessment based on their years of learning English?***" was to investigate how the changes in participants' English language learning years (less than 10 years, 10 years, 11 years, 12 years, 13 years and more) could influence their view on the purpose of assessment. Analysis of the data was carried out by using multivariate test of variance which indicated that

differences in the years of learning English made no statistically significant difference regarding conceptions of assessment. In a similar study, Zaimoğlu (2015) investigated the effect of teaching experience over participants' conception levels. She found out no statistically significance difference as well (Pillai Trace = .23, $p = .86$). Descriptive analyses yielded similar results as seen in the gender case, though some slight mean differences were presented among different learning experience groups. Participants moderately agreed with improvement and student accountability conceptions whereas irrelevance conception held the lowest mean value with "moderately disagree" agreement level. Improvement conception held the highest agreement level among all the levels and 10 years' experienced participants mostly agree that assessment should be used to improve quality of teaching and learning ($M = 4.25$, $SD = .76$). 11 years experienced participants, on the other hand, asserted the highest disagreement level for the conception of irrelevance which implies that assessment is useless ($M = 3.49$, $SD = .54$). The non-significant or similar results could be explained by very similar years of experience among participants. English is included in the course of education from 4th grade on in the primary level in Turkish context, and a sophomore student is expected to be having roughly ten years of English learning background. It was seen that slight differences did not cause wide differences on participants' conceptions, rather participants would prefer to follow their conception on the basis of assessment should enhance quality of teaching and learning as well as provide accountability for individual learnings. Therefore, pre-service teachers agreed on improvement and accountability functions of assessment and rejected to see it as irrelevant or useless regardless of their English learning durations.

5.3.3.3. Age

The question ***"Is there any significant difference in the participants' conceptions of assessment regarding their age difference?"*** were formulated to unveil how age factor influenced participants' conceptions of assessment. After the participants were divided into two groups as 20 years or less and 21 years or more (the range was between 18 to 25 years), a multivariate analysis of variance were performed to investigate the difference. Statistical results found no significant difference between different age groups and conceptions of assessment. These results are inline with those of the previous studies in the conception of

assessment literature. Brown (2004) found no statistically significant difference in participants' mean scores for each conception regarding their age difference in his study in which he investigated primary school teachers and managers' conceptions of assessment in New Zealand context. In the current study, descriptive results indicated that both groups' conceptions of assessments are similar even though some slight mean differences were detected. Both group of students indicated that they moderately agreed with improvement and accountability conceptions and disagreed with irrelevance conception as seen in other independent variable values of the study. This could be explained by the close range of age groups, similar grade levels and similar experiences they had gone through. It may be assumed that if other values and conditions such as place, ranks, degree of education etc., small age differences would not lead into significant differences in participant's assessment conceptions. This created the belief that students, regardless of their age differences, conceive assessment as a tool for their personal improvement and accountability of their improvement at the same time. However, referring assessment as irrelevant was disagreed by almost all age groups since assessment practices hold a common ground for any age groups in Turkish educational context.

5.3.3.4. Grade point average (GPA)

The purpose of the question "Is there any significant difference in the participants' conceptions of assessment regarding their grade point average differences?" was to investigate whether achievement levels of the participants make a significant difference on their conceptions of assessment. The data was categorized into two groups, and it was analyzed by using multivariate analysis of variance (MANOVA) method. The statistical analyses indicated that there was no statistically significant difference on participants' conceptions of assessment regarding their GPA values which entails that participants' academic achievements did not make a significant change on their understanding of assessment purposes. Descriptive statistics revealed that both high and medium achieving students moderately agree that assessment provides improvement to teaching and learning processes. It was interesting to see that high achievers agree with the student accountability conception contrary to medium achievers who moderately disagree that assessment accounts students' outcomes even though the mean values were

slightly different. This could be accounted for because of the fact that medium achieving students might conceive assessment as not evaluating their competencies in a crystal clear manner since assessment puts them into medium or moderately achieving group. For the irrelevance conception, both high and medium achieving groups indicated a moderate level disagreement. Regardless of their academic achievement, the participants agreed that irrelevance or useless view of assessment should be rejected. This could be explained by the educational culture which presents assessment-based education for all levels of students. Even if assessment labels people as low, medium or high achievers, all the participants agreed that assessment is inseparable part of education system and it should be taken as a vehicle of improvement and accountability of the products instead of approaching to assessment as irrelevant, useless or bad.

5.3.3.5. Grade levels

The question ***“Is there any significant difference in the participants’ conceptions of assessment regarding their grade levels?”*** was formulated to unearth how different grade levels (second, third and fourth grades) made a difference in the participants’ conceptions of assessment. Multivariate analysis of variance results indicated that there was statistically significant difference between grade levels and participants’ conceptions of assessment. However, when the data was further analyzed for in depth results by using multivariate test and Bonferroni adjustment, none of the dependent variables was reached to statistical significance. To put it simply, grade levels made a significant difference on participants’ conceptions of assessment when taken as a whole, but not considered separately. Moinvaziri (2015) conducted a study to examine university teachers’ conceptions of assessment. She found out that there is a strong correlation between teaching experience and accountability: the more they are experienced, the higher values they presented for accountability conceptions. This could be concluded as experience makes difference in participants’ conceptions of assessment even though conditions of participants (pre-service teachers vs. university teachers) were different. Descriptive statistics indicated that second-grade participants reported slightly higher level of improvement conception, whereas fourth-grade participants asserted that assessment should be used for student accountability. Third-grade participants held the middle ground in

general in their conceptions. These results could be explained by the course they had taken. Pre-service teachers were provided with two different assessment related courses during their undergraduate studies. Measurement and evaluation course is given in the spring semester of second-grade, and measurement and evaluation in a foreign language course is given in the spring semester of fourth-grade. Sophomores' higher levels in the improvement conceptions could be due to the fact that they have not still completed an assessment related course. That is why, they considered assessment as a means of improvement instead of accountability. On the other hand, seniors scored higher in student accountability even though they have completed the same assessment course with juniors. The difference could be explained by the employment exam which senior students have to take after they complete their degrees in order to get a job. The realities of educational policies and applications they have begun to face could lead them to see assessment as an accountability tool for their qualifications.

5.3.4. Discussion of future assessment practices

Assessment is an inseparable side of educational processes for a great deal of time, and it is widely benefited in different educational contexts for accountability purposes regardless of whether it is mandated or not. Therefore, assessment places an important place both for students and teachers and for the other parties such as policy makers, parents etc. During the implementation of assessment tools, teachers' beliefs and practices plays a significant role for the type of assessment tool used, purpose, timing and assessment returns. Brown (2002) stated the importance of teachers' beliefs on assessment as;

all pedagogical acts, including teachers' perceptions of and evaluations of student behaviour and performance (i.e., assessment), are affected by the conceptions teachers have about their own confidence to teach, the act of teaching, the nature of curriculum or subjects, the process and purpose of assessment, and the nature of learning among many educational beliefs. (p. 3).

Similarly, Munoz, Palacio and Escobar put forward that "teachers' assessments of student behavior and performance, among others, are shaped by the theories they have in relation to teaching, assessment, and the nature of learning" (p. 144). This idea is supported by Harlen's (2005) thought of assessment process as how we interpret it. So, teachers' interpretation of assessment needs or results shape the purpose and outcomes of assessments. Asch (1976) argued that teachers' beliefs

over students “closely linked to one’s choice of evaluation techniques (as cited in Brown, 2002, p. 2)

A handful of studies have been investigated the teachers’ conceptions of assessment and preferences of assessment tools they are using or will use in the near future. In her study, Vardar (2010) provided her participants to opt for assessment tools they prefer to use for their classroom assessments out of a checklist including a range of assessment tools including objective and subjective methods. It was revealed that most of the participants opted for objective tools such as multiple choice, fill in the blanks and true false. However, alternative assessment tools such as performance-tasks or portfolios were also ranked very high according to study results. Similarly, Zaimoğlu (2013) revealed that participants mostly opted for objective techniques even though their assessment practices were greatly varied. She concluded that participants preferred measures indicated that they aimed to use assessment as a way of improving students’ learnings and higher order skills.

Statistical analysis of the data already indicated that improvement conception of assessment held the highest agreement level among participants. When it further analyzed item by item, students indicated a moderate level of agreement with statements such as “Assessment provides feedback to students about their performance, Assessment feeds back to students their learning needs, and Assessment is a way to determine how much students have learned from teaching”. This results indicated that the teacher candidate will mostly benefit from formative assessment techniques for assessment purposes in their real classrooms. The results revealed that the participants opted for the purpose of assessment for improvement of teaching and learning and they paid attention to the importance of feedback. Brown (2002) noted down that “improvement conception is associated with the term formative” (p. 28) and formative assessment mostly calls for feedbacks. Therefore, it could be concluded that pre-service English teachers will be benefiting from formative assessment and feedback for the improvement of their quality of teaching and students learning. It could also be deduced that peer assessment and peer feedback can also be benefited in their real applications besides teacher assessment and feedback. Brown (2002) also put forward that improvement conceptions refuses the idea of

testing lower order skills, and it should include the identification of higher order skills as well. Hence, student teachers will likely to provide their students with more in depth tools not just formal testing tools such as multiple- choice in order to evaluate a broad range of abilities of the learners.

Secondly, participants indicated a moderately agreement level for the conception of accountability. Munoz et al. (2012) have withdrawn two aim of the assessment from the relevant literature: pedagogical and administrative aims. Pedagogical goals refer to development and improvement of students and administrative goals refer basically to accountability. In this line, statistical results demonstrated that student teachers agreed that assessment should be used for accountability of schools or students. “Assessment provides information on how well schools are doing, and Assessment is assigning a grade or level to student work” were the highly agreed items for school and school accountabilities. Brown (2002) noted that accountability refers to summative assessment. From this perspective, it could be deduced that student teachers will be benefiting from summative assessment tools which includes traditional (multiple choice, true-false) or performance (portfolio, interview) assessments. Therefore, learner will be subjected to summative assessment techniques at the end of the term or year for the accountability of their own learning outcomes as well as how well school is doing.

Shortly, the study results indicated that student teachers moderately agreed with the purposes of assessment for improvement and accountability. In this direction, it could be deduced that they will benefit and use a mixture of formative and summative assessment together to provide feedback students for their learnings and provide accountability for students and schools overall results or success outcomes.

6. CONCLUSION

6.1. Introduction

This chapter is designed to presents implications of study findings, suggestion for further researches, limitations of the study, and a brief conclusion to study.

6.2. Implications for Practice

The study results indicated that school accountability and students' accountability placed an important role for pre-service teachers' conceptions of assessment. This idea could be supported by competitive nature of the Turkish education system where high stake tests places a key role for students' future progress as well as schools put into ranks from most successful to less. However, improvement conception held the highest mean value and agreement level of all the conception levels, this demonstrated that pre-service teachers are eager to benefit from assessment for improvement of teaching and learning process. Therefore, textbooks, assessment procedures and the like should be organized and revised by accounting for improvement conception together with school and student accountabilities.

It was also seen that irrelevance conception still holds a place in student teachers' conceptions even though it has the lowest mean value of all. Therefore, assessment related course should be varied and emphasized during during undergraduate education process for all teacher candidates. Besides accountability of competence and related works, not only the books but also the lecturers should present assessment more thoroughly so that pre-service teachers internalize it as a key factor for development instead of as a burden on their shoulders both as student teachers and as real teachers.

Additionally, the purpose of the study should be made crystal clear before the education process. Apart from formative and summative use of assessment which are used either for providing feedback or evaluating progress, wash back effect of the assessment should be prioritized since it "positively influences what and how teachers teach and learn" (Browan & Abywicrama, 2010, p. 38). Therefore, washback could enhance improvement conception of assessment at the same time decrease irrelevant view of assessment.

6.3. Suggestions for Further Research

The present study was conducted with 204 pre-service English teachers at Hacettepe University. In further studies, participants from different universities and contexts could be used to compare and deeply analyze participants' conceptions of assessments'.

Teachers conceptions of assessment inventory- Abridged (TCoA-III A- Version3- Abridged) was used to collect data. In a further study, original scale could be applied to test participants' conception levels more thoroughly.

The data was collected quantitative tools and only quantitative analysis were used to investigate the data. In a further research, both qualitative and quantitative tools could be applied to gather data and mixed analysis might be used to reveal more in depth outcomes.

Only pre-service English teachers were used as participants. Apart from teacher candidates, students, parents, managers and other stakeholders should be included in the study in order to investigate their conceptions of assessment for a broader understanding of conceptions of assessment.

Original version of the inventory, which is in English, was used in this study since participants had enough competence in the target language. In a further study with English language teachers or teachers' candidates, both original version of the inventory and adapted version for Turkish should be delivered at the same time in order to eliminate any possible effect of cultural implication(viewpoint) of the language.

In a follow up study, senior pre-service teachers and novice teachers could be analyzed and compared in order to examine the effects of short-time real class experience on participants' conceptions of assessment.

6.4. Limitations of the Study

In this thesis, listed reasons would be seen as the limitations of the study especially with generalizability of the results.

1. The data were collected and analyzed by using quantitative methods. Absence of any qualitative method could be a limitation.

2. Of all the participants were from the same setting and absence of participants from different setting could be a limitation to generalization.
3. Participants' possible future assessment applications are withdrawn from their answers to survey items. An interview with students would be more effective to make inference.

6.5. Conclusion

The main purpose of the study was to investigate pre-service English teachers' conceptions of assessment. After the data was analyzed statistically, it was seen that participants agreed with the conception that assessment should be used for improvement of teaching and learning. They remarked that irrelevant view of assessment had little place on their understanding of assessment purposes. Then, improvement, school accountability and student accountability conceptions correlated significantly and it was revealed that there was a strong positive correlation among them whereas improvement and irrelevance conceptions were negatively correlated. Finally, it was seen that each individual difference had a slight mean difference for different conceptions; however, grade level is the only variable making statistically significant difference on pre-service English teachers' conceptions of assessment.

REFERENCES

- Aksu, H. H. (2008). A study on the determination of secondary school mathematic teachers' views on alternative assessment. *Humanity & Social Sciences Journal*, 3(2), 89-96.
- Airasian, P. W., & Madaus, G. F. (1972). Criterion-referenced testing in the classroom. *Measurement in Education*, 3(4), 1-7.
- Azis, A. (2012). Teachers' conceptions and use of assessment in student learning. *Indonesian Journal of Applied Linguistics*. 2(1), 41-51.
- Azis, A. (2015). Conceptions and practices of assessment: A case of teachers representing improvement conception. *TEFLIN Journal*, 26(2), 129-154.
- Baird, J. A. (2011). Does the learning happen inside the black box?. *Assessment in Education: Principles, Policy & Practice*, 18(4), 343-345.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment?. *Assessment in Education: Principles, Policy & Practice*, 5(1), 103-110. doi: 10.1080/0969595980050106
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *PhiDeltaKappan*, 80(2), 139-148.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning*. (p. 206-230). London, England: Sage.
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. London: McGraw- Hill International.
- Bond, L, A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2). Retrieved from <http://PAREonline.net/getvn.asp?v=5&n=2>
- Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, 13, 259-278.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in higher education*, 15(1), 101-111.
- Brown, G.T.L. (2002). Teacher conceptions of assessment. (Unpublished doctoral dissertation), University of Auckland, New Zealand.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implication for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318. doi: 10.1080/0969594042000304609

- Brown, G. T. L. (2006). *Teachers' conceptions of assessment inventory--Abridged* (TCoA-III-A-Version 3-Abridged). Unpublished test. Auckland, NZ: University of Auckland.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students conceptions of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational & Developmental Psychology*, 7, 63-74.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3-17. doi: 10.1080/09695940701876003
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347-363. doi: 10.1080/09695940903319737
- Brown, G. T. L., Lake, R. & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27, 210-220. doi: 10.1016/j.tate.2010.08.003
- Brown G. T. L., & Michaelides M.P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *Eur J Psychol Educ*, 26, 319–337. doi: 10.1007/s10212-010-0052-3
- Brown, G. T. L., Hui, S.K.F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*. 50, 307-320. doi: 10.1016/j.ijer.2011.10.00.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4), 653-675
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253-271.
- Chirimbu, S. (2013). Using alternative assessment methods in foreign language teaching. Case study: Alternative assessment of business English for university students. *Buletinul Stiintific al Universitatii Politehnica din Timisoara, Seria Limbi Moderne*, 12, 91-99.
- Corcoran, C. A., Dershimer, E. L., & Tichenor, M.S. (2004). A teacher's guide to alternative assessment: Taking the first steps. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 77(5), 213-218. doi: 10.3200/ TCHS.77.5.213-218
- Danielson, C. (2008). Assessment for learning: For teachers as well as students. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 191-213). New York: Taylor & Francis.

- Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. San Francisco, CA: Jossey-Bass.
- Fenton, R., (1996). Performance assessment system development. *Alaska Educational Research Journal*, 2(1), 13-22.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (1993). *How to design and evaluate research in education* (7th Ed.). New York, NY: McGraw-Hill.
- Frank, M. & Barzilai, A. (2004). Integrating alternative assessment in a project-based learning course for pre-service science and technology teachers. *Assessment & Evaluation in Higher Education*, 29(1), 41-61. doi: 10.1080/0260293042000160401
- Gonzales, G. C. (2003). Assessment in English language classrooms: Constructing and empowering meaningful learning. *Revista Pensamiento Actual*, 4(5), 84-92.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51.
- Griffiths, T., Gore, J., & Ladwig, J. (2006). Teachers' fundamental beliefs, commitment to reform and the quality of pedagogy. In: *Paper prepared for presentation at Australian association for research in education annual conference, Adelaide, November 26-30*.
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Boston, MA: Allyn & Bacon.
- Gulikers, J. T., Bastiaens, T. J., Kirschner, P. A., & Kester, L. (2006). Relations between student perceptions of assessment authenticity, study approaches and learning outcome. *Studies in Educational Evaluation*, 32(4), 381-400.
- Hamayan, E.V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212-226.
- Hancock, C. R. (1994). Alternative assessment and second-language study. *ERIC Digest*. Retrieved from http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED_376695.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16(2), 207-223. doi: 10.1080/09585170500136093
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381. doi: 10.1080/09695940903319745
- Heaton, J. B. (1988). *Writing English language tests*. New York, NY: Longman.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hornby, W. (2003). Assessing using grade-related criteria: A single currency for

- universities?. *Assessment and Evaluation in Higher Education*, 28: 435- 454.
- Jandra, A. (2011). The importance of assessment in ELT. Retrieved from <http://prezi.com/gjwma82zl-hm/the-importance-of-assessment-in-elt/>
- Kim, S. I., Lee, M. J., Chung, Y., & Bong, M. (2010). Comparison of brain activation during norm-referenced versus criterion-referenced feedback: The role of perceived competence and performance-approach goals. *Contemporary Educational Psychology*, 35(2), 141-152.
- Knight, P. (2001). *A briefing on key concepts: Formative and summative, criterion and norm-referenced assessment*. York, UK: Learning and Teaching Support Network.
- Knight, P. T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education*, 27(3), 275-286. doi: 10.1080/03075070220000662
- Maclellan, E. (2004). Authenticity in assessment tasks: A heuristic exploration of academics' perceptions. *Higher Education Research & Development*, 23(1), 19-33.
- Moiinvaziri, M. (2015). University teachers' conception of assessment: A structural equation modeling approach. *Journal of Language, Linguistics and Literature*, 1(3), 75-85.
- Moss, C. M., & Brookhart, S. M. (2010). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Alexandria, VA: ASCD.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers' Professional Development*, 14(1), 143-158.
- Pallant, J. (2010). *SPSS survival manual*. London, McGraw-Hill Education.
- Palmer, S. (2004). Authenticity in assessment: reflecting undergraduate study and professional practice. *European Journal of Engineering Education*, 29(2), 193-202.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2), 91-105.
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18, 238- 250.
- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42(4), 203-220.
- Sadler, D.R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84. doi: 10.1080/0969595980050104
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.

- Stanley, J. C. (1964). *Measurement in today's schools*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69(5), 363-368.
- Şahin, Ç., & Karaman, P. (2013). Sınıf öğretmeni adaylarının ölçme ve değerlendirmeye ilişkin inançları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28(2), 394-407.
- Thompson, A.G. (1992). Teachers beliefs and conceptions: A synthesis of the research. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127- 146). New York: Macmillan.
- Tyler, R. W., & Wolf, R. M. (1974). *Crucial issues in testing*. Berkeley, CA: McCutchan Pub. Corp.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, 31(5), 505-521. doi: 10.1080/02602930600679506
- Vardar, E. (2010). *Sixth, Seventh and Eighth Grade Teachers' Conceptions of Assessment*. (Unpublished Master's Thesis), Middle East Technical University Graduate School of Social Sciences, Ankara.
- Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, MA: Cambridge University Press.
- Volante, L., Beckett, D., Reid, J., & Drake, S. (2010). Teachers' views on conducting formative assessment within contemporary classrooms. *Annual Meeting of the American Educational Research Association*, 1–23.
- Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wynne, H., & Mary J. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379. doi: 10.1080/0969594970040304
- Yıldırım, A. (2004). Student assessment in high school social studies courses in Turkey: Teachers' and students' perceptions. *International Review of Education*, 50(2), 157-175.
- Yüce, Z. (2015). *Pre-service English language teachers' conceptions of assessment and assessment practices*. (Unpublished Master's Thesis), Çaç University Institute of Social Sciences, Ankara.
- Zaimoğlu, S. (2013). *Teachers' and students' conceptions of assessment in a university EFL preparatory school context*. (Unpublished Master's Thesis), Çaç University Institute of Social Sciences, Mersin.

APPENDICES

APPENDIX A. ETİK KOMİSYONU ONAY BİLDİRİMİ



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Sayı : 35853172/ 433-895

28 Mart 2016

EĞİTİM BİLİMLERİ ENSTİTÜ MÜDÜRLÜĞÜNE

İlgi: 17.03.2016 tarih ve 704 sayılı yazınız.

Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Bilim Dalı tezli yüksek lisans programı öğrencilerinden **Ramazan YETKİN**'in Yrd. Doç. Dr. **Hüseyin ÖZ** danışmanlığında yürüttüğü "**Türkiye'deki İngilizce Öğretmen Adaylarının Değerlendirme Algısı ve Gelecekteki Değerlendirme Uygulamaları (Pre-Service English Teachers Conception of Assesment and Their Future Assesment Practies in a Turkish Context)**" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **22 Mart 2016** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

APPENDIX B. ORJİNALLİK RAPORU



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLER ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANA BİLİM DALI BAŞKANLIĞI'NA

Tarih: 16/6/2017

Tez Başlığı : TÜRKİYE' DEKİ İNGİLİZCE ÖĞRETMEN ADAYLARININ DEĞERLENDİRME ALGISI VE GELECEKTEKİ DEĞERLENDİRME UYGULAMALARI

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak Turnitin adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir.

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Endeksi	Gönderim Numarası
16/6/2017	87	132235	15/6 /2017	%15	825440980

Uygulanan filtreler:

- 1- Kaynakça hariç
- 2- Alıntılar dâhil
- 3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Adı Soyadı: Ramazan Yetkin
Öğrenci No: N14228826
Anabilim Dalı: Yabancı Diller Eğitimi
Programı: İngiliz Dili Eğitimi
Statüsü: Y.Lisans Doktora Bütünleşik Dr.

Tarih ve İmza
16/6/2017

DANIŞMAN ONAYI

UYGUNDUR.

(Unvan, Ad Soyad-İmza)

Yrd. Doç. Dr. Hüseyin Öz



HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
THESIS ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
TO THE DEPARTMENT OF FOREIGN LANGUAGE EDUCATION

Date: 16/6/2017

Thesis Title : PRE-SERVICE ENGLISH TEACHERS' CONCEPTIONS OF ASSESSMENT AND THEIR
FUTURE ASSESSMENT PRACTICES IN A TURKISH CONTEXT

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

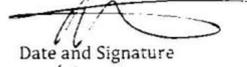
Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
16/6 /2017	87	132235	15/6 / 2017	15%	825440980

Filtering options applied:

1. Bibliography excluded
2. Quotes excluded
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.


Date and Signature
16/6/2017

Name Surname: Ramazan Yetkin

Student No: N14228826

Department: Foreign Language Education

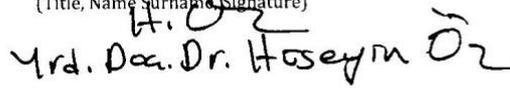
Program: English Language Teaching

Status: Masters Ph.D. Integrated Ph.D.

ADVISOR APPROVAL

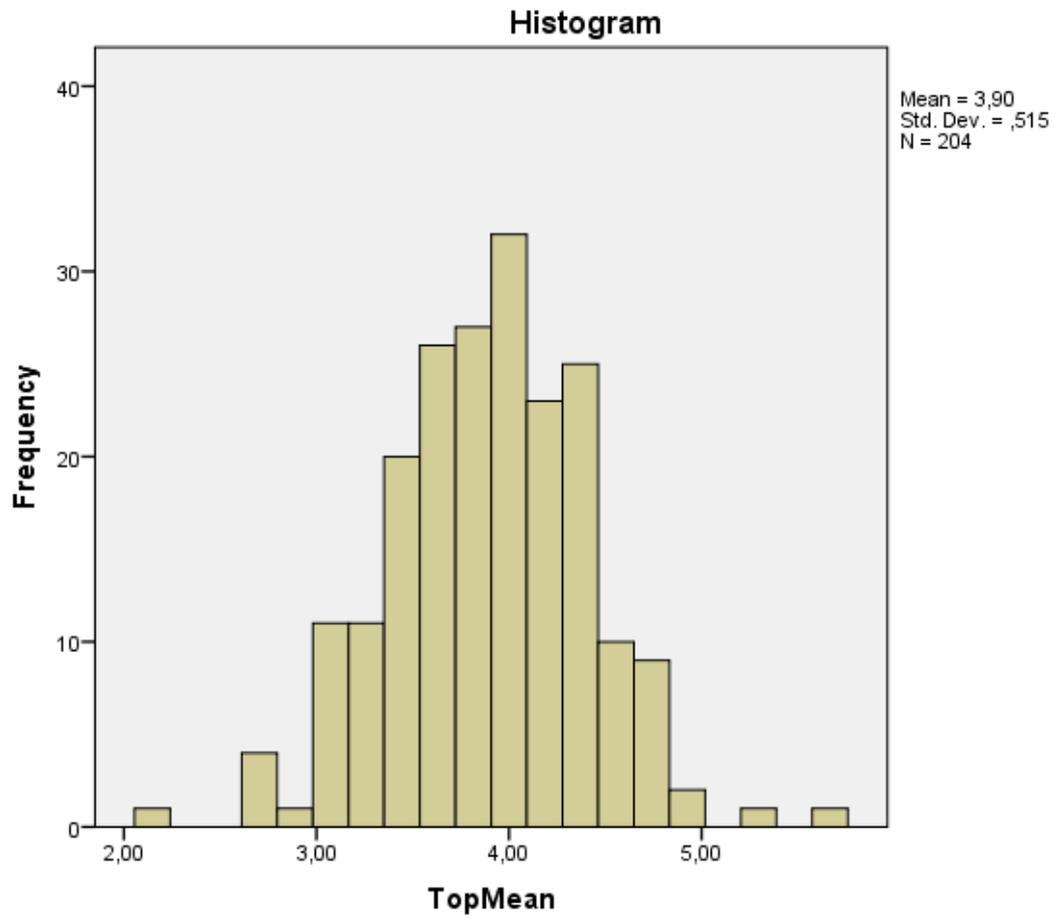
APPROVED

(Title, Name Surname, Signature)

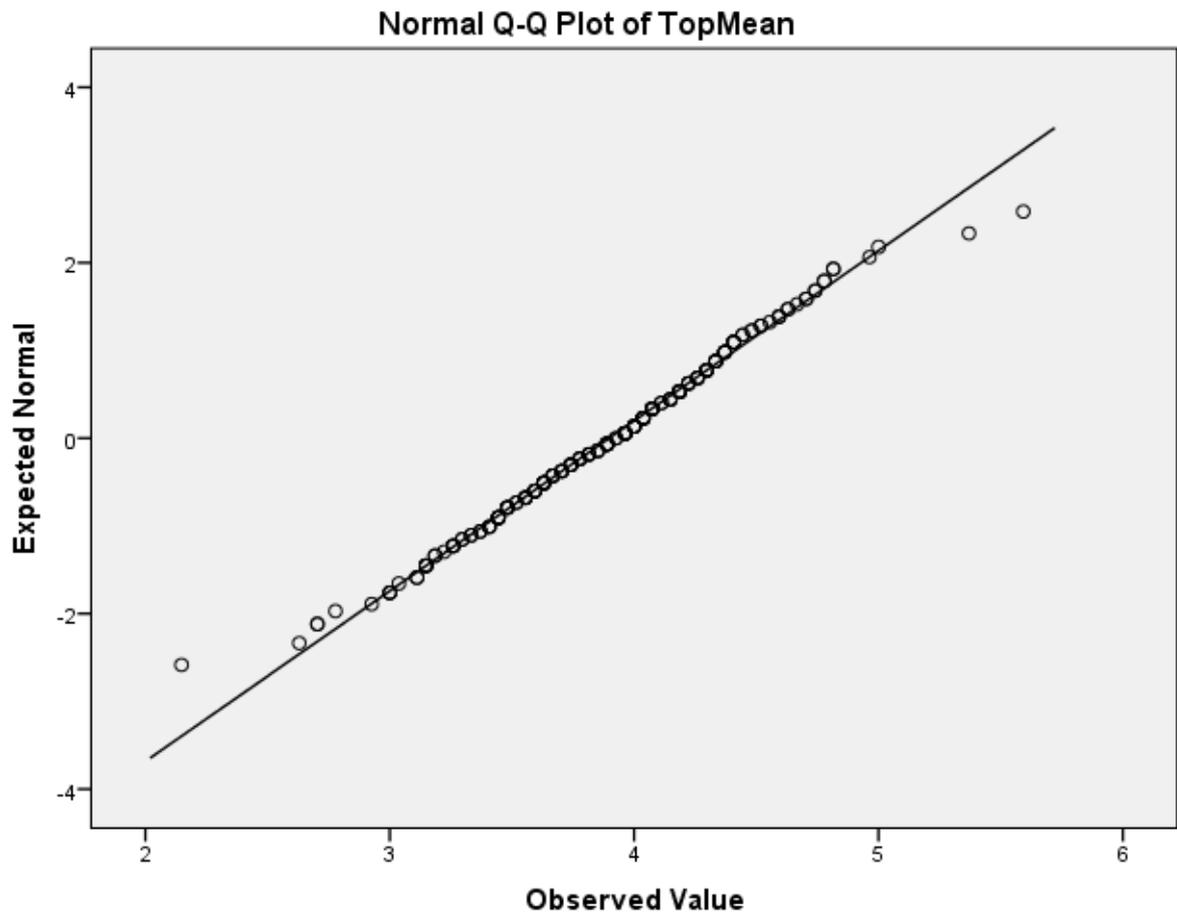

Yrd. Doç. Dr. Huseyin Öz

69

APPENDIX C. TEST OF NORMALITY- HISTOGRAM



APPENDIX D. TEST OF NORMALITY- Q-Q PLOT



APPENDIX E. DATA GATHERING INVENTORY SAMPLE

Teachers' Conceptions of Assessment Inventory (TCoA-III A- Version 3- Abdrigid)

Dear Participant,

The following survey is administered in order to find out your conception of assessment. There is no right or wrong answers in this list of statements. Please make sure that the answers you give in these questionnaires will remain confidential. Your answers will have a valuable contribution to the study. Thank you very much for your participation. Hacettepe University - ELT Department

Your gender: Female Male **Your age:** _____ years old.

Your grade: _____

What is your current Grade-Point Average (GPA = Academic Average)? _____

What are your years of English Education? _____

Part A:

This instrument is composed of 27 statements concerning how you conceive the assessment. Please indicate the degree to which each statement applies to you by marking whether you feel the statement is:

1 = Strongly Disagree 2 = Mostly Disagree 3 = Slightly Agree
4 = Moderately Agree 5 = Mostly Agree 6 = Strongly Agree

	ITEMS – CONCEPTION OF ASSESSMENT	SD	MD	SA	MA	MA	SA
1.	Assessment provides information on how well schools are doing.	①	②	③	④	⑤	⑥
2.	Assessment places students into categories.	①	②	③	④	⑤	⑥
3.	Assessment is a way to determine how much students have learned from teaching.	①	②	③	④	⑤	⑥
4.	Assessment provides feedback to students about their performance.	①	②	③	④	⑤	⑥
5.	Assessment is integrated with teaching practice.	①	②	③	④	⑤	⑥

6.	Assessment results are trustworthy.	①	②	③	④	⑤	⑥
7.	Assessment forces teachers to teach in a way against their beliefs.	①	②	③	④	⑤	⑥
8.	Teachers conduct assessments but make little use of the results.	①	②	③	④	⑤	⑥
9.	Assessment results should be treated cautiously because of measurement error.	①	②	③	④	⑤	⑥
10.	Assessment is an accurate indicator of a school's quality.	①	②	③	④	⑤	⑥
11.	Assessment is assigning a grade or level to student work.	①	②	③	④	⑤	⑥
12.	Assessment establishes what students have learned.	①	②	③	④	⑤	⑥
13.	Assessment feeds back to students their learning needs.	①	②	③	④	⑤	⑥
14.	Assessment information modifies ongoing teaching of students.	①	②	③	④	⑤	⑥
15.	Assessment results are consistent.	①	②	③	④	⑤	⑥
16.	Assessment is unfair to students.	①	②	③	④	⑤	⑥
17.	Assessment results are filed & ignored.	①	②	③	④	⑤	⑥
18.	Teachers should take into account the error and imprecision in all assessment.	①	②	③	④	⑤	⑥
19.	Assessment is a good way to evaluate a school.	①	②	③	④	⑤	⑥
20.	Assessment determines if students meet qualifications standards.	①	②	③	④	⑤	⑥
21.	Assessment measures students 'higher order thinking skills.	①	②	③	④	⑤	⑥
22.	Assessment helps students improve their learning.	①	②	③	④	⑤	⑥
23.	Assessment allows different students to get different instruction.	①	②	③	④	⑤	⑥
24.	Assessment results can be depended on.	①	②	③	④	⑤	⑥
25.	Assessment interferes with teaching.	①	②	③	④	⑤	⑥
26.	Assessment has little impact on teaching.	①	②	③	④	⑤	⑥
27.	Assessment is an imprecise process.	①	②	③	④	⑤	⑥