IJATE
International Journal of Assessment Tools in Education

January 2017

International Journal of Assessment Tools in Education

International Journal of Assessment Tools in Education

Content of the Issue

Vol. (4) No. (1)
ISSN: 2148-7456
www.ijate.net

http://www.ijate.net
e-ISSN: 2148-7456 © IJATE
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

INTERNATIONAL JOURNAL OF ASSESSMENT TOOLS IN EDUCATION

Prof. Dr. İzzet KARA

Editor

Publisher : İzzet KARA
Frequency : 2 issues per year
Online ISSN : 2148-7456
Website : http://www.ijate.net/index.php/ijate
Adress : Pamukkale University,
Eduction Faculty,
Department of Mathematics and Science Education
20070, Denizli, Turkey

Phone: +90 258 296 1036
e-Mail: ijate.editor@gmail.com

IJATE is indexed in:

- Tübitak Ulakbim Dergipark
- DOAJ,
- Index Copernicus International,
- Google Scholar,
- Türk Egitim İndeksi,
- Open Access Journals,
- Akademik Dizin,
- Academic Keys,
- CiteFactor (ASJ),
- SIS (Scientific Index Service) Database,
- SCIPIO (Scientific Publishing & Information Online),
- MIAR 2015 (Information Matrix for Analysis of the Journals),
- I2OR Indexing Services,
- JournalTOCs
- Türkiye Akademik Arşivi (ULAKBIM)

- *International Journal of Assessment Tools in Education* (IJATE) is a peer-reviewed on-line journal.
- Author(s) are responsible from the copyrights of the figures, pictures and visuals, and the content of the manuscripts, accuracy of the references and quotations and proposed ideas.

Dear Readers,

Our Journal has been published two issues (January and July) in each year since 2014. Volume 4, Number 1 issue is now available on the website of the journal. We are grateful and thank first the authors and reviewers, then advisory and editorial board, editor assistants and journal secretariat while bringing **International Journal of Assessment Tools in Education (IJATE)** in its present status. Our journal is making an effort to develop its quality in each issue with your contributions.

Sincerely,

Prof. Dr. İzzet KARA

IJATE Editor-in-Chief

IJATE Editor-in-Chief Phone: +90 (258) 296 10 36, E-mail: ijate.editor@gmail.com

Vol 4, No 1 (2017)

Table of Contents

International Journal of Assessment Tools in Education: Vol. 4, Issue 1, (2017) pp. 1-18

http://www.ijate.net
e-ISSN: 2148-7456 © IJATE
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

"Original Research Article"

Exploring Teachers' Assessment Practices and Skills

Bouchaib Benzehaf [1]

Chouaib Doukkali University, El Jadida, Morocco

Abstract

The need for increased use of test results to improve educational outcomes is urgent; yet, there is little understanding in the research literature of practitioners' knowledge and skills in interpreting and using educational data (test results) to enhance classroom instruction and student learning. This study aims to survey 40 high school teachers who work in El Jadida region, distributed between males and females, of different years of experience, with the purpose of learning about their assessment practices, and identifying the barriers that prevent thoughtful applications of formative assessment in classrooms. A questionnaire and an interview were used as a data collection technique. The findings point to use of a varied number of assessment strategies ranging from homework assignments to in-class written tests but mainly for summative purposes. In light of the results, the study ends with implications for teachers and policy makers.

Article Info

Received
14 July 2016

Revised:
08 August 2016

Accepted
20 September 2016

Keywords:
Formative assessment, assessment literacy, descriptive feedback

1. INTRODUCTION

Within the context of reforming the Moroccan educational system, several attempts have been made to boost educational standards, the latest of which is the implementation of the standards-based approach to foreign language teaching. This approach places several obligations on practitioners ranging from identifying students' needs and meeting them through monitoring students' learning to differentiating instruction. However, concrete corresponding changes in assessment practices have been lacking. Undoubtedly, teachers who have been focused merely on the assignment of grades in assessment have neglected an important component of the teaching/learning process. An important function of classroom assessment is promoting students' learning, and raising their motivation levels. This type of assessment is termed "formative assessment".

As its name delineates, formative assessment means that we assess students as part of forming their competences and skills and helping them continue to develop these competences (Brown, 2003). It is used to support and inform learning, build self-confidence, and capacity for success (Stiggins, 2001). It is assessment for learning rather than of learning, and it is becoming increasingly the focus of research (Black & William, 1998; Crooks, 1988; Sadler, 1989; Stiggins,

[1]Corresponding Author Phone:: +21262784923 Email: bbenz841@gmail.com

2148-7456 /© 2017 DOI:10.21449/ijate.254581

2001). On the other hand, summative assessment refers to the assessment of learning, summarizing the development of learners at a particular time. It performs the function of measuring and quantifying the competence or the skill that the student has attained, (Stiggins, 2001).

2. LITERATURE REVIEW

Fleming and Chambers (1983) conducted a survey of teachers' written classroom assessments and came to the conclusion that teachers' test items were of low quality according to principles of good item writing. Specifically, what characterized these items were ambiguity as well as an inclination to rely primarily on recall rather than on higher-order thinking skills. Over all, teachers were found to be deficient in devising quality tests, a finding that testifies to teachers' low assessment literacy levels.

In a similar vein, Marso and Pigge (cited in Wise, 1993) conducted a study which consisted of a direct analysis of teachers' self-constructed tests. They found out that these classroom tests revealed frequent violations of common question writing guidelines. The study consisted also of ratings by classroom teachers, principals and supervisors of classroom teachers that identified needs for a variety of testing competencies. Teachers expressed a high need for competencies including the use of test results for instructional purposes. Over all, they reported they needed training in the following: grading and scoring activities, identifying pupil strengths and weaknesses, and training in test validity-related competencies including matching questions with objectives, writing questions that trigger the use of higher order thinking skills, and measuring true progress of pupils.

In the same line, Plake (1993) conducted a study on teachers' capacity to develop assessment methods appropriate for instructional decisions, scoring, interpreting and communicating results to students, and exploiting results for further learning and right decisions concerning instruction. Results pointed to a weakness on the part of teachers particularly in communicating results to students. The majority (85%) also reported an interest in developing skills in assessment. Similarly, Mertler (2004) compared assessment literacy levels of pre-service teachers and in-service teachers. Overall, results were quite parallel to those reached by Plake (1993) with some quite insignificant differences. Respondents performed quite well in administering, scoring and interpreting test scores. Yet in this study, the lowest scores were on developing valid grading procedures in addition to communicating test results.

Black and Wiliam (1998) reviewed more than 250 articles related to formative assessment. They stated that the studies "show conclusively that formative assessment does improve learning," and that the gains in student achievement are "amongst the largest ever reported" (Black & Wiliam, 1998, p. 61). However, the study pointed to a difficulty on the part of teachers to effectively incorporate formative assessment into their teaching practices. To repair the damage, Black and Wiliam (1998) have suggested that a number of practices may lead to more successful implementation of formative assessment. It is noteworthy that these suggestions are shared by other researchers. First, it is suggested that clear learning targets and criteria upon which performance will be judged are made clear to students (Black & Wiliam, 1998; Schunk, 2003). Second, teachers are advised to administer effective feedback on student performance (Black &Wiliam, 1998; Crooks, 1998). Third, students should be involved in the process of formative assessment so that they can develop meta-cognitive skills (Black &Wiliam, 1998;

Sadler, 1989; Schunk, 2003). Last but not least, results should be timely available to students so as not to miss out on their instructional role and transformative potential (Popham, 2004).

Vaden-Goad (2009) conducted an experimental study in which he compared formative and summative assessment. He found out that the amount of information and motivation levels increased as a corollary of changing the function of assessment from summative to formative. However, continuous assessment in itself is not sufficient to serve the purpose of scaffolding learning because there are conditions that need to be present so that assessment can be formative; notable among these conditions is timely feedback to students. Consequently, teachers need to receive training in administering formative assessment, an important component of which is giving effective, appropriate and timely feedback. Another study bearing on elementary teachers' knowledge and self-efficacy for measurement concepts reported that practicing teachers were relatively skilled at classifying assessment types and interpreting student scores (French & Gotch, 2011). Conversely, results of items assessing teachers' ability to act on standardized scores by using them to make appropriate instructional decisions indicated a weakness in this skill.

Yamtima and Wongwanichb (2014) investigated the levels of classroom assessment literacy of primary school teachers. Nineteen primary school teachers at Wat Phai Rong Wua School completed a classroom assessment literacy questionnaire and 8 teachers participated in a focus group discussion. The findings showed that most of the participants had scores for classroom assessment literacy at the poor level. In light of this finding, the researchers suggested a developmental approach for improving the classroom assessment literacy of primary school teachers which emphasized cooperative learning and teamwork.

Babaii and Damankesh (2015) investigated the effect of high school final examinations on students' test-taking and test-preparation strategy use. The findings revealed that the examinations influenced the students into employing strategies which exerted a negative influence on their learning as they directed them toward a measurement-driven approach to learning rather than to an approach focused on improvement of learning.

The literature on feedback also establishes that while feedback is of paramount importance if given in the form of comments, it fails to deliver on this potential if paired with marks or grades because students tend to overlook comments and content themselves with marks or grades (Butler, 1988). The comment, descriptive though it may be intended by the teacher, will be interpreted as an explanation of the grade. Hence, descriptive comments will only be read as descriptive if they are not accompanied by a grade. In turn, Sadler (1989) stated that empirical evidence demonstrated that feedback can yield positive effects only if intentionally oriented towards improvement of learning. This finding was corroborated by subsequent studies conducted on feedback.

Black and Wiliam (1998) also concluded that when feedback was of high quality, it improved students' work, thus contributing to an increase in standards. In his turn, Hattie (1999) conducted an extensive synthesis of a wide range of educational research and concluded that feedback was the most powerful factor that could enhance achievement. Along the same lines, Higgins et al. (2002) argue that students, despite exhibiting an interest in grades, also demonstrate an intrinsic motivation to learn from feedback. All in all, giving quality feedback serves a scaffolding function which is essential for stretching one's "comfort zone".

In Morocco, Melouk (2001) conducted an exploratory study on classroom assessment in high school. A survey of national exams in the nineties and exams in the latest decade showed that little change has been produced in the way exams are designed. In a way consistent with what has been found out in other parts of the world, Melouk (2001) concluded that the majority of teachers have received a very limited training in assessment. The majority of his respondents (teachers) expressed their interest for training in item production and item management as well as some general background in testing. The study also researched whether learners were trained in new test types and exam papers or not, the result of which was that training was generally insufficient. In addition to that, Melouk explored the content of official exams; more precisely, he investigated teachers' views about whether baccalaureate exams are skills-based or knowledge-based. On this point, a substantial number of teachers thought that exams are knowledge-based, a finding that reveals that exams do not foster creative and critical thinking.

Khtou (2011) investigated students' and teachers' attitudes to assessment in Fez and Rabat faculties of arts, and at a time when the current system was not yet fully implemented (around 2004-5). Utilizing questionnaires and interviews, Khtou probed faculty and students' attitudes to assessment, both terminal and continuous. Concerning teachers, 60% of questionnaire respondents stated that the system of assessment that was prevalent at the time of the study was unfair; and asked for a system that would provide students with feedback on their work to help them learn and improve their performance. In similar terms**,** many students (63%) reported that the system of assessment was unfair, and, in turn, expressed their wish for the inclusion of feedback. These students were dissatisfied with the fact that not only was feedback withheld, but so were the scoring criteria as well. Therefore, there was no room for improvement, and the likelihood for the same mistakes to continue to appear in new situations was strong.

Bouzidi (2009) investigated the type of feedback given to students in higher education. The researcher investigated 2000 marginal and end comments on student essays at Ibn Zohr University. He analyzed these comments in terms of their linguistic features and their intended pragmatic effect on the students. Then, he had a second look at the revised drafts to measure the impact the comments had on the students' revision and to assess the extent of improvement that occurred as a result of the comments. He concluded that the comments were mainly form-based rather than content-based, the most-focused-upon aspects being spelling, punctuation and neatness/appearance of paper while more important aspects like thesis statement, related ideas and development of ideas, for instance, did not obtain equal attention by the teachers. Consequently, the impact of such comments was restricted to some structural changes while the overall essay quality did not improve.

By and large, three conclusions can be drawn from the literature review: first, the value of formative assessment is paramount in driving learning forward (Babaii & Damankesh, 2015; Black & Wiliams, 1998; Vaden-Goad, 2009); second, quality feedback improves students' work; and third, teachers' assessment literacy levels are low in the absence of training (Fleming & Chambers, 1983; Marso & Pigge, 1993; Plake 1993; Yamtima & Wongwanichb, 2014).

With respect to research in Morocco, classroom assessment is still under-researched as attests Melouk's conclusion regarding "scarcity of field research in this area in Morocco" (Melouk, 2001, p.51). Additionally, the few studies surveyed in Morocco do not focus on high school continuous assessment, but on high school terminal assessment (Melouk, 2001) and university assessment (Bouzidi, 2009; Khtou, 2011). Accordingly, research addressing teachers'

assessment practices in high school is highly desirable. This provides the rationale for the present study.

3. METHOD

This study falls within exploratory research type. Research utilizing an exploratory design mainly explores an existing phenomenon; numbers, though, may also be used to characterize individuals or groups. The design, therefore, is both qualitative and quantitative as the study employs a questionnaire and an interview for triangulation purposes.

3.1. Participants

The context of the present study is secondary schools in the town of El Jadida. High school teachers of English constitute the sample of the study. Overall, there are 58 teachers of English in high school in the town of El Jadida, 37 males and 21 females. The sample chosen for this study consists of 40 teachers. It was difficult to include all the teachers in the town as some teachers refused to take part while three female teachers did not return the questionnaire.

The study utilized non-probability sampling, where exclusion or inclusion from the sample is deliberately done by the researcher. Effort was made to collect data from educational practitioners of different backgrounds (age, experience, and education). In this regard, the researcher had opted for equality in number between men and women practitioners, but because male practitioners outnumber female practitioners, this equality was not possible. The final number was 24 male teachers and 16 female teachers distributed as follows:

Table 1. Background of respondents

Items		Count	Percent
Gender	Male	24	60
	Female	16	40
Diploma	Bachelor's degree	31	77.5
	Master's degree	9	22.5
Experience	1-5 years	9	22.5
	6-10	12	30
	11-15	13	32.5
	16-20	3	7.5
	21+	3	7.5

3.2. Materials

3.2.1. Questionnaire

A questionnaire was selected as a data collection technique because, unlike other data collection techniques, it has several advantages. First of all, a questionnaire is cost effective in terms of money and time; it can be administered to a large number of people in one place, thus providing a high proportion of usable responses. Besides, a questionnaire permits anonymity, which would cause respondents to feel at ease and express themselves freely. It also provides respondents with ample time to deeply think about their answers as they are usually not required to fill out the questionnaire on the spot. Moreover, a questionnaire is objective as the researcher's influence is reduced in a questionnaire than in other data-collection instruments.

The questionnaire consists of two sections. Section one includes 14 questions regarding teachers' assessment practices. The participants are, for instance, asked about the frequency of testing their students and the type of assessment and questions that they give to the students. The questionnaire is in the appendix.

3.2.2. Teacher interview schedule

By way of enriching and crosschecking data obtained in the questionnaire, a semi-structured teacher interview which used open-ended questions was designed. It was selected in this study because it is useful in that it provides clues into the reality of teachers' practices, thus filling any gaps that might have arisen from utilizing the questionnaire. Unlike the structured interview which uses questions followed by choices from which the interviewee selects the answer, the semi-structured interview does not provide answers, thus allowing for free individual responses.

The questions in the interview are phrased in such a way as to allow for free answers. There are no choices from which the interviewees can select their responses. The interview aims to uncover the strategies, types of questions that teachers use in their assessment of students, and the purposes for which assessment is carried out, the frequency with which they assess students, the kind of feedback they provide, the turn-around of tests, and the barriers, if any, that hinder the implementation of formative assessment. All in all, the items in the interview are aimed at eliciting answers that will be compared with answers to questions in the questionnaire. The number of interviewed teachers is five. The interview schedule is in the appendix.

4. RESULTS

The respondents and the interviewees were probed about their classroom centered assessment practices. As to frequency of assessment, 72.5% of the questionnaire respondents indicated that they assessed their students once a month; whereas 27.5% of the teachers stated they did so twice a semester:

Table 2. Frequency of assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	once a month	29	72.5	72.5	72.5
	other	11	27.5	27.5	100.0
	Total	40	100.0	100.0	

This finding was supported by findings from the interview. All interviewees claimed that they assessed their students twice a semester; two teachers, though, pointed out using quizzes in addition to tests. All interviewees attributed the frequency with which they tested students to administrative reasons. One teacher, for instance, said that: "*The administration requires two marks, so we administer two tests*" thereby delinking assessment from instruction. To a follow-up question, the interviewees all made it clear that they did not assess at the beginning of the year. In other words, they assessed at the end of a unit of study.

Concerning the type of assessment which teachers conducted in their classes, the answers came as follows:

Table 3. Assessment strategies

Response	Frequency	Percentage
Written tests	40	100
Oral tests	3	7.5
Homework assignments	38	95
Self-assessment	2	5
Peer assessment	2	5
Other	0	0

As can be seen from the chart above, all participants used written tests. Also, most of the respondents (95%) claimed that they assigned homework tasks to students. However, only two percent of the respondents claimed they made use of peer and self-assessment:

Similarly, the interviewees identified a variety of assessment strategies ranging from written tests to homework assignments. However, the overall umbrella for this variety of assessment tools is written tests as teachers talked about dictations, cloze tests and essays. One teacher said: "*I introduce variety into my assessment practices to test a variety of dimensions of intelligence*".

The respondents also indicated that they used all types of questions: multiple-choice questions, short-answer questions, matching questions, essay questions, true/false questions, and fill-in-the-blanks. Two of the respondents added other types; namely, cloze tests, scrambled sentences and dictations. Likewise and in line with the acclaimed variety of assessment strategies, all interviewees reported using a variety of question types ranging from w/h questions, to fill-in-the-blanks and multiple-choice questions. One of them said: "*We need to diversify our questions so as to allow for simple as well as difficult questions to be included*". Another teacher said: "*Variety of questions is important because it enables weak students to answer some of the questions*". Obviously, this teacher was speaking about variety in terms of simple and difficult questions.

In addition, the respondents also reported targeting both levels of difficulty in their tests, deep understanding of concepts on the one hand, and surface knowledge and recall of facts on the other hand. That is, they conducted tests that used a combination of items that disclosed students' thinking processes and deep understanding as well as items that targeted recall and knowledge of facts. However, three of the interviewees admitted that they emphasized recall more than understanding and higher-order questions while the remaining two teachers said they used a variety of questions with some assessing recall and others assessing deep understanding.

Concerning informing students of the objective of tests, only 5% indicated that they informed students of the objective of testing while 87.5 % administered tests and quizzes without any explicit delineation of why the assessment was being conducted:

Table 4. Percent of teachers informing students of objective of testing

Response	Frequency	Percent
Yes	5	12.5
No	35	87.5
Mean frequency & percent	20	50

Similarly, four interviewees clarified that they only informed their students of the date of the assessment without delineating the objective behind testing. One of them, for instance, said: "*No, I don't tell them that, only the day when they sit for the assessment*". The fifth interviewer, however, claimed he informed his students of the objective of assessment: "I tell them what they will be assessed on and why".

As for test duration, 60% of the respondents said that they used tests that lasted between one hour and two hours, and 35% claimed they used 30-60 minute tests while only 5% of the respondents claimed that they used short quizzes the duration of which ranged between 15 and 30 minutes. The following diagram provides a good illustration of this point:

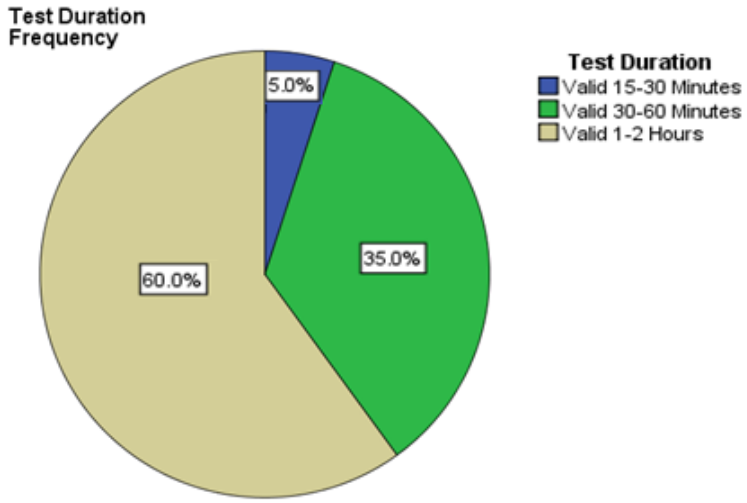


Figure 1. Test duration frequency

Three of the interviewees claimed that the duration of their tests was one hour while the other two teachers said that they also made use of short quizzes the duration of which ranged between quarter of an hour and half an hour. One of them commented that: "*only a long test makes me assess students' true achievement*" while one of the interviewees who reported using quizzes said: "*Quizzes allow for quick check-up of recently learnt material*". However, the teachers did not report a frequent use of quizzes; one even reported using quizzes as a strategy of calming down a noisy classroom, thus indicating using assessment as a punishment instrument rather than as an instructional means.

Respondents were further inquired about availability of test results. Fifty percent of the teachers indicated that results were available in not less than two weeks' time, and 45% returned tests in two weeks' time:

Table 5. Availabilty of results

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than a week	2	5,0	5,0	5,0
	Two weeks	18	45,0	45,0	50,0
	More than two weeks	20	50,0	50,0	100,0
	Total	40	100,0	100,0	

In a like manner, four interviewees reported that they did not return tests in less than two weeks while the fifth teacher said: "*As long as it takes me to correct them*" and indicated that correction was a hard task to do particularly that "*we teach an average of five or six classes of more than 40 students each*". Additionally, all five interviewees indicated that the type of feedback they wrote on papers was no more than a comment on the grade. The comment given to a student with a grade of ten, for instance, was "average", while a student with less than ten would receive less than average or weak. Similarly, a student with more than ten would get "good" or "very good" depending on how high their mark was.

Additionally, 55% of the teachers indicated that they did not hold correction sessions with their students while 45% did. This finding was also supported by findings from the interview with four interviewees claiming that they contented themselves with informing students about their marks. An interviewee said: "*students don't care about correction, why should I make it?*" This statement was echoed by another teacher who commented: "*Even if you correct, the majority of students will not follow with you because all they are interested in is grades*". The remaining teacher who claimed he held correction sessions pointed out that he contented himself with giving students the correct answers.

Table 6. Test correction sessions

Response	Frequency	Percent
Yes	18	45
No	22	55

Respondents were further required to indicate if they disclosed to students scoring rubrics before the test was conducted. 42.5% of the teachers made it clear that they did not disclose to students scoring rubrics while 57.5% said they disclosed scoring rubrics, a finding also confirmed by interview results with all interviewees claiming that they did not disclose their scoring rubrics except on the day when the results were available. "*I show the students the marks for every item when I give them back their papers*," said one interviewee.

Table 7. Percent of teachers disclosing scoring rubrics

Response	Frequency	Percent
Yes	23	57.5
No	17	42.5

Eighty percent of the teachers indicated they did not write comments on students' papers, and the rest (20%) said they did so. By contrast, four interviewees indicated that they wrote comments on test papers. The following table illustrates the point for the questionnaire respondents:

Table 8. Percent of teachers writing comments

Response	Frequency	Percent
Yes	8	20
No	32	80

The respondents who claimed that they offered their students feedback were further required to give description of the type of feedback that they provided to students. As can be seen from the chart below, all 8 teachers who said they gave feedback termed it "descriptive feedback" while the interviewees clarified the point by saying that the type of descriptive feedback they gave was restricted to essay writing when assigned as homework. Otherwise, comments were kept within the bounds of stating whether the work was good or not.

Table 9. Type of feedback given by teachers

Type of feedback	Frequency	Percent
Descriptive	8	100
Rewarding/Punishing	0	0

Respondents were also asked to indicate the purposes for which they administered assessment tasks. They were required to choose from the following: 1) identifying students' strengths and weaknesses, 2) predicting student performance on final Baccalaureate exam, 3) assigning grades for administrative reasons, 4) tracking students' progress toward proficiency in English, or 5) other purposes which teachers were required to name. Respondents could select multiple purposes for assessing students. The following table describes the distribution of the responses. As can be seen from the table below, the vast majority of the tests were administered in order to assign grades for administrative reasons (100%), followed by predicting students' performance on final Baccalaureate exam. Although the purpose of "identifying students' strengths and weaknesses" received 62%, it came third in the ranking.

Table 10. Purposes for assessment

Purposes for assessing	Frequency	Percent
Identifying students' strengths and weaknesses	25	62
Predicting students' performance on final Bac exam	36	90
Assigning grades for administrative reasons	40	100
Tracking students' progress toward proficiency in English	27	67.5
Other	0	0

The interviewees, in turn, indicated more than one purpose for assessment. One of them claimed: "*I assess to make students see themselves in the mirror so that they know where they are from the learning objec*tives" (sic). Another one said: "*I assess to give marks to students*". Still, a third claimed that: "*I assess because without assessment and tests, students will not learn*".

With regard to the barriers that hindered the provision of feedback, returning results quickly, or conducting remedial work, 85% of the respondents indicated that they did face barriers that hindered them from providing feedback, returning results quickly, or conducting remedial work as the following table demonstrates:

Table 11. Percent of teachers indicating the existence of barriers

Response	Frequency	Percent
Yes	34	85
No	6	15

When asked to mention these barriers, all the respondents who answered positively to the previous question indicated that the obstacles were restricted to large classes, pressure to finish the text-book, and weak level of students. Similarly, all interviewees pointed to their concern with finishing the syllabus as well as over-crowdedness of classrooms as barriers to an implementation of a formative framework of assessment. However, other interesting barriers which were mentioned were related to assessment knowledge and skills. One teacher said: "*I do not understand what you mean by formative assessment*". A second interviewee reported that "*teachers have not been trained in conducting such a type of assessment*", while a third raised the issue of incentives and motivation for teachers: "*How do you expect from a teacher who has financial constraints to do his job well? Without motivation, there is nothing*," he commented. A further barrier that was mentioned was the absence of motivation in students.

5. DISCUSSION

The results seem to point to a general formative use of assessment. Respondents indicated that they used different assessment strategies ranging from written tests to homework assignments. The questions included in tests and quizzes have also been found to be varied (multiple choice, true/false, gap filling,…). Additionally, results point to the fact that assessment is conducted for some of the purposes that are formative like tracking student progress toward proficiency in English, identifying student strengths and weaknesses, and predicting student performance on the final Baccalaureate exam. Nevertheless, few respondents indicated that they used results to modify teaching method or instruction, which is an important aspect of formative use. Likewise, few respondents indicated that they used alternative modes of assessment which are at the heart of formative assessment, Vis, projects, portfolios, peer and self-assessment. Feedback, in turn, was found to be lacking in students test and quiz papers according to the respondents. Obviously, feedback is another key component of formative assessment.

The overall tendency seems to be towards the inclusion of some aspects of formative assessment. However, there are serious limitations in teachers' assessment practices. Most respondents, for instance, indicated that they administered no more than two tests a semester. Moreover, no interviewee indicated conducting diagnostic assessment, whose importance cannot be overemphasized in determining the students' learning stages, at the beginning of the year. Such practices run counter to characteristics of formative assessment which require that assessment be frequent, continuous, diagnostic and at the service of instruction. Brookhart (2011) and Popham (2011), for instance, argue that formative assessment is more effective as an assessment instrument when conducted frequently. The teachers, therefore, are called upon to make their assessment as frequent as possible so as to increase student motivation and performance. In this context, the literature points to a strong correlation between formative classroom assessment and student motivation and achievement on both classroom and large-scale assessments (Rea-Dickins & Gardner, 2000; Torrance & Pryor, 2002).

Likewise, although teachers indicated that they used a variety of assessment strategies and a variety of question types, only two respondents indicated that they conducted peer and self-assessment, which are key forms of formative assessment. Similarly, no one indicated the use of projects or performance-based assessments which are another critical aspect of alternative assessment. These types of assessment which are absent from teachers' practices are also at the heart of the standards-based approach that the Moroccan educational system officially adopts, and which is in line with formative assessment. The advantage of these modes of assessment is

that they aim at simulating real-world contexts, focusing on processes as well as products, and drawing upon higher-order thinking and problem-solving skills (Lynch, 2001).

Similarly, respondents said that their assessment targeted both surface knowledge and recall of facts on the one hand, and deep understanding on the other hand while the interviewees made it clear that recall of facts dominated over critical thinking in their assessment practices. This is in line with Melouk's finding (2001) that exams do not foster creative and critical thinking as they are knowledge-based. A critical aspect of formative assessment is to make the second objective (critical thinking) more prevalent. The literature on formative assessment indicates that formative use of assessment tends to disclose students' thinking processes with a view to deepening and sharpening them. Assessment tools, therefore, have to be designed in such a way as to target and nurture a culture of critical thinking.

The literature also indicates that although teachers were familiar with various types of assessment practices (e.g., cloze tests, performance assessments, etc.), they have been found to lack a clear framework for implementing assessments that would reflect and support student learning (Torrance & Pryor, 2002). At the heart of this framework is the disclosure of scoring rubrics and the objective behind testing. Among the findings of the present study is that few teachers disclose to students scoring rubrics (42.5%), and more than that number do not hold correction sessions with their students. In the absence of such a transparent system, students' final grades are likely to appear to them to be arbitrary unlike in the presence of it, not only will students regard the practice as fair and democratic but will also be effectively included in the decision-making process and hence will have a share of the responsibility. According to the literature, assessment can be formative only if learners are involved in the process (Wiliam & Black, 1998). Likewise, formative use of assessment is made possible when teachers are familiar with quality criteria and scoring rubrics which should be shared with the students (Black &Wiliam, 1998; Sadler, 1989; Schunk, 1996; Stiggins, 2007). According to the findings of the present study, tests are created and administered without any explicit delineation of why the assessment is being conducted. Students do not know why they should sit for a test except that it is time for a test as required administratively. Obviously, an important component of formative assessment is that tests serve learning purposes which must be clear to the learners.

In a similar vein, a test which lasts for one hour or more raises the question of turn-around. Most teachers administer tests that are no less than one hour. Besides, they made it clear from their answers that it is difficult to return tests in due time for instructional objectives, given the large number of classes they teach. This is in line with the respondents' responses which are to the effect that teachers return tests in no less than two weeks. The literature on formative assessment (e.g., Popham, 2004) suggests that formative use of assessment results is more likely to occur when results are available in a timely fashion. That is the case because a big time lag between a test and availability of results is likely to lead to students missing out on chances for learning.

Written feedback also tends to be absent from student test papers. Very few respondents and interviewees indicated writing comments on students' test papers which they described as descriptive. In this context, Vaden-Goad (2009) concluded in his study that continuous assessment cannot scaffold learning in the absence of some conditions like the provision of feedback**.** Consequently, teachers are advised to administer effective feedback on student performance (Black & Wiliam, 1998; Crooks, 1988) which in the context of test papers should be written so that it can act as scaffolding towards more developed learning stages.

As to purposes of assessment, in spite of the fact that teachers pointed to some formative purposes, the main purpose of assessment remains assigning grades. All respondents highlighted such a purpose for assessment while the formative purposes did not obtain such a consensus. This is again consistent with what has been found in the literature; namely, that teachers are not quite adept at conducting assessment for formative instructional purposes. This finding raises questions on teachers' assessment practices. Melouk (2001), for instance, states that "the way evaluation is carried out today has stripped it of its pedagogical dimensions" (Melouk, 2001, p. 51). Obviously the pedagogical dimension is for assessment to be put at the service of learning; otherwise, it is more summative than it is formative. Even more dangerous than this is the claim of one interviewee that he uses quizzes as a way of calming down a noisy classroom, a practice which amounts to using assessment as a form of punishment.

The respondents and interviewees were also aware of the fact their assessment practices were far from being totally formative and indicated the existence of some obstacles which hindered the implementation of formative assessment. These obstacles, according to them, were restricted to large classes, pressure to finish the syllabus, and students' weak language proficiency level. However, the proponents of formative assessment argue that adopting a formative theory of assessment is likely to yield solutions to these problems. Concerning large classes, formative assessment offers a solution to this problem by suggesting that students be given scoring rubrics to self and peer correct. As to the second obstacle, teachers fear sacrificing coverage of the textbook; but in the rush to cover the syllabus, students are actually learning less and losing much on opportunities for reinforcement. They are denied time to reflect on and interact meaningfully with new information which affects the amount of learning they assimilate. As to weak language proficiency level of students, formative assessment is the best opportunity to help struggling learners and give them a second chance. These students need scaffolding which is a pillar of formative assessment (Bruner, 1978).

6. CONCLUSION & IMPLICATIONS

This study took preliminary steps to understand practitioners' classroom centered assessment practices and knowledge of assessment issues in a contemporary standards-based environment and within a formative framework of assessment. Formative use of assessment results is an important attribute of effective instruction. Such a use is a critical component of teaching, and when done in an appropriate manner, boosts the quality of instruction students receive. Therefore, implementing such a formative model is likely to result in improved instruction and student learning. Teachers who have limited assessment literacy skills move through the teaching and learning process blindly and are more likely to do harm than good to the students. Accordingly, teachers do need the proper training in assessment issues that will allow them to perform their careers in the best way. Sound assessment practices are not a skill that one typically acquires without support in the form of solid training at training centers and subsequent professional development sessions.

Real change requires teachers to give up old teacher-centered approaches with which they feel comfortable. Teachers are called upon to learn, reflect and experiment with new teaching and assessing practices which are more learner-centered. They should make their assessment strategies as varied as possible to capture different dimensions of intelligence. They should also surrender some of their assessment responsibilities to students. This can be done by promoting practices of peer and self-assessment. In so doing, teachers would nurture in their students

practices of self-reliance, thereby encouraging them to become life-long and self-regulated learners. By sharing the responsibility for assessment, students will also develop into responsible citizens.

On their part, educational policy makers and trainers need to make significant and sustained investments in teacher professional development to support effective teaching and assessment practices. Professional development should be targeted clearly to areas of need which have been identified by the teachers: grade giving, differentiation of instruction according to assessment results, design of tests, provision of feedback and overall formative assessment practices. Hence, it is the duty of educational policy makers to better prepare teachers for the teaching tasks awaiting them, and to raise their awareness as to the way in which the different components (curriculum, instruction, and assessment) interact and feed off each other.

7. REFERENCES

Babaii, E. & Damankesh, M. (2015). On students' test-taking and test-preparation strategies. *Studies in Educational Evaluation, 45*, 62-69.

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan 80* (2): 139–48.

Black, P. & Wiliam, D. (2005).*Changing teaching through formative assessment: Research and practice*. London: King's College.

Bouzid, H. (2009). Teacher commentary and its impact on student revision. In H. Bennoudi and Y. Tamer, (Eds). *The Educational Reform.* Paper presented at Agadir conference, Agadir (pp. 67-82).

Brookhart, S. (2011).Educational assessment knowledge and skill for teachers. *Educational Measurement: Issues and Practice, (30)*1, 3-12.

Brown, H. D. (2003). *Language assessment: principles and classroom practices*. Longman: Longman.com.

Bruner J S (1978). The role of dialogue in language acquisition. In A, Jarvella R & Levelt W J M (eds.) *The child's conception of language*. New York: Springer-Verlag.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*, 1-14.

Crooks, T. J. (1998). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438-481.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows to the classroom. InW. E. Hathaway (Ed.), *Testing in the schools* (pp. 29-47). San Francisco, CA:Jossey-Bass.

Gotch, C. & French, B. (2011).Development of and validity evidence for the teacher educational measurement literacy scale. Paper presented at the annual meeting of the national council on measurement in education conference. New Orleans, LA.

Hattie, J. (1999). Influences on student learning, *Inaugural professorial lecture*. New Zealand: University of Auckland.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89* (2), 140–145.

Higgins, R., Hartley, P., & Skelton, A. (2002). The Conscientious Consumer: Reconsidering the role of assessment feedback in student learning. *Studies inHigher Education*, *27*(1), 53-64.

Khtou, H. (2011). *Assessment in higher education: Students and teachers' perceptions*. Published doctoral dissertation, Saarbrucken: Lambert academic publishing.

Lynch, B. (2001). The ethical potential of alternative language assessment. In C. Elder (Ed). *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 228-239). Cambridge: Cambridge University Press.

Melouk, M. (2001, March 25-29). The state of EFL evaluation in Morocco: The testers and teachers' opinions. In A. Zaki & M. Naji, (Eds). *The Teaching and Assessment of English for Global Purposes.* Paper presented at MATE proceedings, Essaouira (pp. 41-51). Rabat: MATE.

Mertler, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education 33*(1), 49-64.

Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

Plake, B. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Midwestern Educational Researcher 6*(1), 21-27.

Popham, W. J. (2004). Curriculum, instruction, and assessment: Amiable allies or phony friends? *Teachers College Record 106*(3), 417– 428.

Popham, J. (2011). *Classroom assessment: What teachers need to know.* Boston: Pearson.

Rea-Dickins, P. & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17*, 215-243.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science 18*, 119–44.

Schunk, D. (2003). Self-efficacy for reading and writing: Influence of modeling, goal-setting, and self-evaluation. *Reading & Writing Quarterly*, *19*, 159-172.

Stiggins, R. (2009). Essential formative assessment competencies for teachers and school leaders. In. H.L. Andrade & G. J Cizek (Eds.), *Handbook of formative assessment.* New York: Routledge.

Stiggins, R. (2001). *Student Involved Classroom Assessment*. Columbus: Merrill Publishing.

Torrance, H. & Pryor, J. (2002). *Investigating formative assessment: Teaching, learning and* assessment *in the classroom*. England: Open University Press.

Vaden-Goad, R. E. (2009). Leveraging summative assessment for formative purpose. *College Teaching*, *57*(3), 153-155.

Wise, Steven L. (1993). Teachers' testing knowledge, skills, and practices. In N. M. Ronald & L. P. Fred (Eds.), *Teacher Training in Measurement and Assessment Skills (*129-185). Lincoln, NE: Buros-Nebraska Symposium on Measurement & Testing.

Yamtima, V. & Wongwanichb, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences, 116*, 2998-3004.

Appendix

Questionnaire

This questionnaire serves a research function. It aims to explore teachers' assessment literacy levels and their ability to use assessment scores to guide instruction and to make appropriate classroom decisions. Your help is highly requested and appreciated. The information you will provide will be treated in strict confidentiality. Thank you very much for your cooperation.

Section One

1. How frequently do you assess your students? Tick what applies to you:

□ Once every month □ Once every two weeks

□ Once every week □ Other, specify

2. What kind of assessment do you conduct with your students? Tick what applies to you:

□ Written tests □ Self-assessment

□ Oral tests □ Peer assessment

□ Homework assignments □Other, specify: _____

Can you explain why: _____

_____

3. In testing, what kind of questions do you use? Please tick what applies to you; you may use numbers from 1 (most important) to 5 (least important) if more than one choice applies

□ Multiple-choice questions □ Short-answer questions.

□ Matching questions □ Essay questions

□ True/false questions □ Fill-in-the-blanks □ Other, specify:_____

Can you explain why: _____

_____

4. In your assessment/testing of students, do you target deep understanding of concepts or surface knowledge and recall of facts or both?

□ Deep understanding of concepts □ Surface knowledge and recall of facts □ Both

5. Do you inform students why they are taking the assessment?

□ Yes □ No

6. Approximately what is the duration of the test? Tick what applies to you:

□ Fifteen minutes □ Between 30 to 60 minutes

□ Between 15 to 30 minutes □ Between 1 to 2 hours

□ Other, specify: _____

7. Approximately how many days does it take to return test results after completing the test?

□ Results are available in less than a week

□ Results are available in two weeks

□ Results are available in more than two weeks

8. Do you correct with students their errors when you give them back their test papers?

□ Yes □ No

9. Do you disclose to students scoring rubrics?

□ Yes □ No

10. Do you write feedback on students' papers?

□ Yes □ No

11. If yes, what type of feedback is it?

□ Rewarding/punishing □ Descriptive (describes students' errors and shows how they can improve)

12. What are the main purpose(s) of administering assessment? Tick what applies to you:

□ Identifying student strengths and weaknesses

□ Identifying students in need of remedial work

□ Predicting student performance on the final Bac exam

□ Assigning grades for administrative reasons

□ Tracking students' progress toward proficiency in English

□ Other (specify): _____

13. Are there any barriers that prevent you from providing feedback, returning results quickly, or conducting remedial work? □ Yes □ No

14. If yes, please mention these obstacles:

_____

Please add any comments you wish:_____

Section Two

1. Gender:

□ Male □ Female

2. Teaching experience:_____ years

3. Highest academic degree: _____

4. Do you participate in some professional development?

□ Yes □ No

If yes, please describe: _____.

Additional comments

Please use the space below for any comments that you may wish to make about this questionnaire or the topic under investigation.

THANK YOU VERY MUCH FOR YOUR HELP

_____

_____

_____

_____

_____

Teacher interview schedule

The aim of this interview is to get an idea about high school English teachers' assessment practices and skills. Please feel free in your answers. The information you will provide will be treated in strict confidentiality.

1. How frequently do you assess your students? Why?
2. Do you assess at the beginning of the year?
3. What kind of assessment do you conduct with your students?
4. What kind of questions do you use?
5. Do you target recall or higher-order thinking in your questions?
6. Approximately what is the duration of the test?
7. Approximately how many days does it take to return test results after completing the test?
8. Do you hold correction sessions with your students?
9. Do you disclose to students scoring rubrics in the test?
10. Do you write comments on students' test papers when you correct them? If yes, of what type?
11. For what purpose(s) do you administer assessment?
12. What barriers, if any, prevent you from conducting formative assessment?

International Journal of Assessment Tools in Education: Vol. 4, Issue 1, (2017) pp. 19-36

http://www.ijate.net
e-ISSN: 2148-7456 © IJATE
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

"Research Article"

The development of a self-efficacy scale for mathematical modeling competencies

Ilhan Koyuncu[1*], Cem Oktay Guzeller[2] , Didem Akyuz[3]

[1]Hacettepe University, Department of Educational Sciences, Measurement and Evaluation in Education, Turkey, Ankara.

[2]Hacettepe University, Department of Educational Sciences, Measurement and Evaluation in Education, Turkey, Ankara.

[3]Middle East Technical University. Department of Elementary Education, Turkey, Ankara.

Abstract

Mathematical modeling has come into prominence during the last few decades in many countries' mathematics teaching curricula. It combines real life situations with mathematical context. Although evaluating students' mathematical modeling performances with a unique Likert type instrument is questionable, having an instrument about their self-efficacy beliefs in mathematical modeling may help to comment about their ideas related to their competencies in mathematical modeling. The purpose of this study is to develop a reliable and valid measurement scale to determine mathematical modeling self-efficacy of mathematics teacher candidates. For this purpose, the draft and final form of the scale were applied to a total of 562 pre-service elementary mathematics teachers from various public universities in Turkey. The findings of study revealed that the scale is unidimensional according to the results of exploratory factor analysis. The unidimensionality of the scale was validated by confirmatory factor analysis. The reliability of mathematical modeling self-efficacy scale was very high (.97). Finally, it was found that this scale is an appropriate measurement tool to evaluate students' self-efficacy beliefs on their mathematical modeling competencies. Some suggestions related to the scale and for further studies were given at the end.

Article Info

Received
01 July 2016

Revised:
29 September 2016

Accepted
07 October 2016

Keywords:
Factor analysis, mathematical modeling, modeling competencies, pre-service teachers, self-efficacy

1. INTRODUCTION

Mathematical modeling is important part of a real life. People often use statements related to mathematics or geometry, and do calculations for different purposes in their daily life. Mathematical modeling can be defined as a part of real life situation that is expressed mathematically. After the expressions, evaluations are done based on the mathematical model, and it is interpreted again in real life context. During this 'mathematization' process, some

*Corresponding Author Phone:+90 312 297 8550. Fax:+90 312 299 2027
E-mail: ilhankync@gmail.com cguzeller@gmail.com didem.akyuz@yahoo.com

2148-7456 /© 2017 DOI: 10.21449/ijate.256552

physical models are built from real life situations, and transformed to mathematical models. From this point of view, models are concepts, which are already in human mind for making sense of complex structures and systems, and their demonstrations (Lesh & Doerr, 2003). The term 'mathematical model', whereas, is related to explaining the structural characteristics and working principles of real life situation (Lehrer & Schauble, 2007; Lesh & Doerr, 2003). For example, assume that it is planned to design a car parking area. The aim is to locate parking areas for each car such that there is minimum empty place and maximum number of cars located in the area. A drawing or physical manipulative that demonstrate the real life situation is a simple model. However, a mathematical model is formulas or some other mathematical demonstrations that could be used to find the better parking method. When a simple or real model and mathematical model concepts are used within a process, they are considered as parts of mathematical modeling. Blum (1993) identifies mathematical modeling as a process that consists of the following stages;

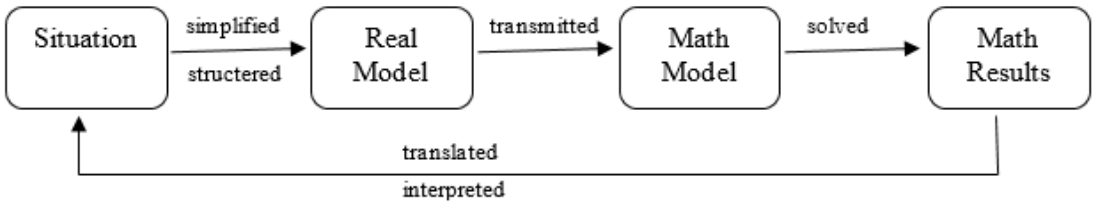


Figure 1. Blum's stages of modeling process

According to the Figure 1, the real life situation is simplified and structured as a real model that can be interpreted and transmitted into a mathematical model. The process continues with solving the problem to get mathematical results and finally, these results are interpreted and translated to the real life situation. In short, mathematical modeling is a cycle of operating on real life situations (Blum, 1993). Similarly, according to Brown (2002), mathematical modeling is formulating a real world problem, solving it by integrating the real life situation and mathematical manipulation, and checking the result using other real life situations. Lingefjrad (2004) identified that mathematical modeling is the process that includes observing an authentic situation, estimating relationships, applying mathematical analysis, obtaining mathematical results, and interpreting again the model.

Especially, during last few decades, mathematical modeling gained much significance in mathematics education (Blum, 2015). A few fundamental reasons for this situation are the fact that using modeling in mathematics education gives opportunity to understand mathematics more meaningfully, learn mathematics by relating with real life, and eliminate inadequacy of available problems (Erbaş et al., 2014). Haines and Crouch (2001) consider that developing mathematical modeling skills is very crucial and, for this purpose, they suggest that many real life problems should be solved in the classroom, and mathematical modeling courses should be added to the curriculum apart from mathematics courses. Maaß (2006) urged that modeling competencies should be paid attention in the class. It is advised that mathematical modeling needs to be included in mathematics courses at every stage of education beginning from early years of education before high school and college levels (Lehrer & Schauble, 2003).

In general manner, mathematical modeling is the ability to make transitions between real world and mathematical world (Crouch & Haines, 2004). Although there are different

approaches based on different theoretical frameworks, there is no consensus on mathematical modeling approaches in the literature (Kaiser & Sriraman, 2006). Though Lesh and Doerr (2003) argue it as a new paradigm beyond the constructivism, Haines and Crouch (2007) regarded modeling as the transition between mathematics and real life. Kaiser and Sriraman (2006) classify modeling approaches that constitute bases for international studies as realistic and applied modeling, contextual modeling, educational modeling, socio-critical modeling, epistemological or theoretical modeling, and cognitive modeling.

Another classification is made based on the aim of using modeling in mathematics education. *Modeling as the purpose of teaching mathematics* and *modeling as a means to teach mathematics* are two general approaches depending on the aim of use (Galbraith, 2012; Gravemeijer, 2002; Niss, Blum, & Galbraith, 2007). Although in the first one the main objective is to develop models and use this models to improve students' mathematical modeling abilities; in the second one the aim is to use mathematical modeling to teach mathematical models and contexts (Erbaş et al., 2014). According to Haines and Crouch (2007), mathematical modeling needs to be regarded as interdisciplinary rather than considering it solely in the mathematical context. Therefore, it is suggested that mathematical skills and competencies that could be used in other disciplines need to be identified and supported in different ways (Erbaş et al., 2014). In the second approach, modeling is used as a teaching tool and it is called the *emergent modeling* approach (Gravemeijer, 2002). It is a result of Modeling and Modeling Perspectives (MMP) in mathematics education (Lesh & Doerr, 2003) and Realistic Mathematics Education (RME) (Freudental, 1991) approaches. MMP is a theoretical approach based on constructivism and socio-cultural theories. It focuses on teaching, learning and problem solving in mathematics. In the context of MMP, a 'model' is product obtained at the end of a process; 'modeling' is a process that constituting the physical, symbolic, or abstract model of a situation (Erbaş et al., 2014). The theoretical basis of another modeling approach offered by RME is the same as MMP (Freudental, 1991; Gravemijer, 2002). In this approach, 'modeling' is not just transferring authentic problem situations into mathematical language; it is also the process to reveal new relationships by organizing facts included in these authentic situations (Gravemeijer & Stephan, 2002).

In addition to this information related to modeling and some main approaches, it seems to be an important matter to identify the term 'mathematical competency' before discussing about what modeling competencies are. Niss (2004, p.120) defines it as: "Mathematical competence then means the ability to understand, judge, do, and use mathematics in a variety of intra- and extra-mathematical contexts and situations in which mathematics plays or could play a role." Based on this definition, it is important to know about the context of mathematical modeling in detail to understand modeling competencies. Maaß (2006) urges that there is close relationship between modeling competencies and the process of modeling. Blum and Kaiser (1997) evaluate modeling competencies as the objectives that needed to be accomplished during modeling process. Therefore, they consider that the students initially need to *understand the real problem* and *set up a model based on reality*. Then, they *set up a mathematical model from the real model* and *solve mathematical questions within this mathematical model*. Finally, the students should i*nterpret mathematical results in a real situation* and *validate the solution* (Blum & Kaiser 1997).

Ikeda and Stephens (1998), and Profke (2000) have similar understanding of modeling competencies with Blum and Kaiser (1997). However, Profke (2000) gives more coverage to

general skills such as being curious instead of competencies like interpreting and verifying than Blum and Kaiser (1997). Although Niss (2004) has similar ideas with Blum and Kaiser about modeling competencies, he differentiates between active modeling and models that have been already prepared. Maaß (2006) develops a holistic point of view and stated that modeling competencies are more competencies than just following steps of modeling process. These competencies are sub-competencies to carry out the single steps of the modeling process, metacognitive competencies, competencies for structuring real world situations, competencies for urging modeling process, and competencies for seeing different solutions of the real world problem and developing positive attitude (Maaß, 2006).

On the other hand, the researchers such as Ross Crouch, John Davis, Andrew Fitzharris, Chris Haines, John Izard, Ken Houston, and Neville Neill define mathematical modeling skills at a micro level and stated them as follows (Lingefjrad, 2004): Identifying and simplifying the given information, making the aim explicit, formulating the problem, identifying variables, parameters, and constants, formulating mathematical expressions, selecting a mathematical model, using graphical representations, and comparing with real life situation and controlling the process. It could be seen that these competencies specified by Lingefjrad (2004) are all included in the sub-competencies suggested by Blum and Kaiser (1997). In the present study, mathematical modeling competencies are regarded at micro level. The main reason for not using the framework suggested by Maaß (2006) is the fact that each sub-competency is not specified well and it is difficult to discriminate between them implicitly. Therefore, Blum and Kaiser's (1997) framework that includes all competencies specified by Lingefjrad (2004) was used to develop item clauses. However, modeling is time consuming to apply in the classroom, does not fit the curriculum, makes mathematics lessons more demanding and less predictable, and assessing modeling is challenging. Peer-to-peer assessment, take-home exams and surveys are some tools for measuring and evaluating students' modeling competencies (Lingefjard & Holmquist, 2004). However, due to the complexity of measuring modeling skills with a unique assessment tool, using a survey will be a convenient way of commenting on mathematical modeling 'self-efficacy beliefs' of teacher candidates which is in the scope of this study.

Whether being used as a means or purpose, mathematical modeling has been an important part of school mathematics. In addition to the studies on mathematical modeling and modeling competencies, teachers' self-efficacy beliefs about their modeling competencies seem to be an important subject that might affect their effectiveness in the classroom. Bandura (1997) identifies the term 'self-efficacy' as beliefs of a person about his/her capacity to do and organize intended course activities to attain given objectives. People with high level of self-efficacy effort much to success and they are more patient in problematic situations (Bandura, 1997). It is found that when the learners are at equal levels of ability, the possibility of finishing a given task for learners who believe to do the task is higher than the ones who do not believe (Schunk & Pajares, 2005).

Another important point is the fact that self-efficacy is not an observed skill or competency, it is internal beliefs of a person related to what to do with this skill (Synder & Lopez, 2002, p. 278). In the context of this study, modeling self-efficacy is related to beliefs of students concerning what to do with their mathematical modeling competencies. In other words, it refers to the beliefs of the students about what they can do with their capacity in mathematical modeling. Bandura (1997) states that four main sources of self-efficacy are mastery experiences, the vicarious experiences provided by social models, social persuasion, and physiological factors. According to Bandura (1997), mastery experiences are the most important and effective

sources of self-efficacy beliefs. For example, if the students with higher performance on mathematical modeling get higher scores from a modeling course, they will develop positive beliefs in their capacity of this subject. However, although they have higher performance, and they get low scores, then, their beliefs on their ability will decrease and it will directly affect their performance. This means that students' personal experiences influence their self-efficacy (Bandura, 1997).

The vicarious experiences, also called 'modeling', are related to take others as models. When people do not have any judgments about their capacities or have limited experience on a subject, vicarious experiences are very effective on their performances (Bandura, 1997). Social persuasion, another source of self-efficacy, is related to encouragement of parents, teachers, or friends on accomplishing a task or a mission. Physiological factors, the last source of self-efficacy, affect significantly one's belief in their capacity. People with high level of anxiety or stress are in tendency to develop lower self-efficacy when compared to ones with low level of negative emotional and physiological feelings. It is urged that people who are able to control their anxiety or stress have high self-efficacy beliefs (Bandura, 1997).

In education, self-efficacy studies generally focus on the relationship of self-efficacy with academic performance, motivational tools, the fields of profession, the choice of profession, teachers' practices in the classroom, and students' products on given tasks (Pajares, 1997). In mathematics education, self-efficacy is found as one of the most important factors that affect students' mathematics performance (Dede, 2008; Pajares & Graham, 1999). Similarly, students with low level of mathematics performance have low level of self-efficacy (Lee, 2009). This situation justifies the claim of Bandura (1997) related to mastery experiences source of self-efficacy, which is the fact that students' personal experiences influence their self-efficacy.

As justified by some researchers (e.g. Bandura, 1997; Dede, 2008; Lee, 2009; Pajares & Graham, 1999) there is a close relationship between students' performances and their self-efficacies. From mathematical modeling perspective, it seems to be important to assess students' beliefs about their capabilities in mathematical modeling as these competencies have important implications for their mathematical modeling performances. Therefore, the purpose of this study is to develop and verify Mathematical Modeling Self-Efficacy Scale regarding mathematical modeling competencies. Blum and Kaiser's (1997) framework that includes all competencies specified by Lingefjrad (2004) was used to develop item clauses. The validity of the scale is established by structural equation models. Content and construct related validity evidences are obtained by means of these models and the opinions of scholars, teachers, and students. The internal consistency of the scale was interpreted by evaluating Cronbach's and McDonald's reliability coefficients. During the verification process, it is aimed to specify the following questions:

1. What is the validity of Self-Efficacy Scale in measuring students' mathematical modeling competencies?
2. What is the consistency level of Self-Efficacy Scale in measuring students' mathematical modeling competencies?

2. METHOD

In the present study, a descriptive research design was used to develop a scale to measure the level of students' mathematical modeling self-efficacy. The indices for mathematical modeling competencies were matched by appropriate expressions and the students were expected to select the degree to agree or disagree with given situation. Therefore, it was aimed to describe students' beliefs and ideas about their mathematical competencies. From these perspectives, the present study represents the characteristics of descriptive studies (Frankel & Wallen, 2011).

2.1. Participants

Participants of the present study were selected from four public universities in Turkey. Each university was selected from different regions including Eastern, Mediterranean, Black Sea and Central Anatolia. The participants were elementary mathematics teacher candidates from all class levels. The reasons for selecting pre-service teachers are that because mathematical modeling is included in almost all levels of education, they need to be aware of their modeling competencies to assist the students effectively, and they meet mathematical modeling activities directly or indirectly during their university education. The study was carried out in three application steps. In the first application, the data were collected from 72 (Female=45, Male=27) pre-service teachers to analyze whether the items work or not in terms of sentence structures and item parameters. The second application was carried out with 180 (Female=127, Male=53) pre-service teachers to explore the structure of the scale by performing an exploratory factor analysis. Finally, the third application was carried out with 310 (Female=230, Male=80) pre-service teachers to confirm hypothetical structure of the scale by performing a confirmatory factor analysis.

2.2. Scale development process

In general manner, the purpose of the present study was to develop a scale to measure an affective construct in mathematics education research. Ryang (2014) suggest following steps to develop a scale for this purpose: Defining research problem and significance of the study, literature review, theoretical framework, collecting data, sample selection, target population, developing/adapting measurement scale, data analysis, reporting the results, and reliability and validity studies. In addition, Crocker and Algina (1986) proposed a more technical scale development plan than Ryang's one and focused on developing/adapting measurement scale, data analysis and reliability and validity studies steps in Ryang's process. In the present study, the scale development process was designed by considering scale development process suggested by Ryang (2014), and Crocker and Algina (1986).

Maaß (2006) criticized that modeling competencies are generally associated with modeling process, and stated that modeling competencies are more competencies than just following steps of modeling process. However, as stated before, each sub-competency is not specified well in Maaß (2006) framework, and it is difficult to discriminate between them implicitly. Many researchers, such as Ross Crouch, John Davis, Andrew Fitzharris, Chris Haines, John Izard, Ken Houston, and Neville Neill define mathematical modeling skills at a micro level (Lingefjrad, 2004). In the present study, mathematical modeling competencies are also regarded at micro level instead of Maaß holistic point of view. Therefore, Blum and Kaiser's (1997) framework that includes all competencies specified by Lingefjrad (2004) was used to develop item clauses.

For each mathematical modeling index (sub-competency) given in Blum and Kaiser (1997, p.9) five to seven items were developed and an item pool consisting of 32 Likert type items was

prepared (see Appendix). The appropriateness of these items was controlled by two scholars having their PhD degree in mathematical modeling and other two scholars having their PhD degree in the field of measurement and evaluation in education. The criteria for evaluating the items was suitability of the items with the indices of mathematical modeling competencies, appropriateness of item formats, suitability of item levels for pre-service elementary mathematics teachers. In order to increase the quality of the items and to develop items representing the construct ideally, the scholars were also requested to suggest delete or add new items if possible. After revisions of the scholars, some items were deleted, new ones were written, and unclear items were modified. As a result, Mathematical Modeling Self-Efficacy Scale consisting of 32 items was prepared as the first draft (see Appendix).

2.3. Data analysis

The first draft was applied to 72 students to analyze some basic psychometric properties of items including item-total score correlations, item mean, and standard deviations. By doing this, the researcher had also the possibility of observing how students react to the expressions and students' ideas about the item structures. The results of the preliminary analysis revealed that item parameters were appropriate and none of the items were needed to be deleted except modifying some of them to make more understandable.

Verification process of the scale consisted of two applications. Exploratory factor analysis and confirmatory factor analysis were carried out in each application. In order to have evidence for internal reliability of the scale, Cronbach α and McDonald ω coefficients were calculated. The assumptions of factor analysis were checked before doing this analysis. In order to test the appropriateness of sample size, Kaiser-Meyer-Olkin sample suitably test was done and it was found that the sample size was adequate. When the descriptive statistics of the data were examined, there was not any missing values and outliers. As another assumption for factor analysis, there should not be multicollinearity. Since, principle component analysis was done, this assumption will not create problem and there is no need to check (Tabachnick & Fidell, 2013). In order to check univariate and multivariate normality, Chi-square statistics was evaluated and this assumption was not satisfied ($p<0.05$). For this reason, Robust Maximum Likelihood method was used to estimate the parameters. For data analysis, IBM SPSS 18.0, LISREL 8.80, and Microsoft Office Excel 2010 software were used.

3. RESULTS

In this section, the results of item analysis for preliminary application, validity and reliability studies were reported in detail.

3.1. Item Analysis of Preliminary Application

Before verifying appropriateness of the scale structure on a large trial group, it will be beneficial to observe the suitably of the scale in practice on a small group. For this purpose, the first scale template was applied to 72 students and the feasibility of the items was examined. For item analysis two methods are generally preferred: Simple and Henryson methods. Simple method bases on upper and lower 27% of whole group and it is appropriate for a sample of 300 or more participants. Henryson method is usually used for a sample of 60 or higher participants. When the sample is big enough, the results of two methods are similar. In the present study, descriptive statistics were calculated based on Henryson method due to the sample size of the first application (N=72).

In Table 1, item means (μ), standard deviations (*s*) and item-total score correlations (r_{IT}) are given. According to the results, item means differ from 3.13 to 3.85 and standard deviations differ from 0.72 to 1.16. Item-total score correlations differ from 0.30 to 0.66 and all of them are significant ($p<0.05$). Since, it was aimed to develop a scale with high internal consistency, 0.30 and higher correlations are enough for intended purpose. The mean of the items is higher or lower than the mean of all items by standard deviations that are higher than 0.60, which is lower bound.

Table 1. Descriptive Statistics

Item	μ	*s*	*n*	r_{IT}	*Item*	μ	*s*	*n*	r_{IT}
1	3.76	0.853	71	0.43	17	3.3	0.962	71	0.47
2	3.65	0.937	72	0.51	18	3.85	0.98	71	0.39
3	3.71	0.721	72	0.4	19	3.81	0.839	70	0.5
4	3.46	0.992	72	0.43	20	3.25	1.143	71	0.66
5	3.26	0.904	72	0.61	21	3.38	1.156	72	0.47
6	3.18	1.142	72	0.39	22	3.51	1.061	72	0.37
7	3.63	1.131	72	0.36	23	3.33	0.856	72	0.6
8	3.42	1.196	72	0.44	24	3.64	0.844	72	0.58
9	3.68	0.819	72	0.52	25	3.13	1.055	71	0.3
10	3.51	1.061	72	0.41	26	3.38	1.08	72	0.51
11	3.47	1.007	72	0.5	27	3.55	0.983	71	0.61
12	3.73	0.962	70	0.63	28	3.63	0.941	72	0.39
13	3.44	1.06	72	0.41	29	3.4	0.944	72	0.43
14	3.38	0.868	71	0.53	30	3.24	1.12	72	0.54
15	3.81	0.959	72	0.53	31	3.26	0.934	72	0.54
16	3.32	1.005	72	0.6	32	3.44	1.005	72	0.48

3.2. Validity Studies

Validity is a process in which evidences are collected to support inferences done based on test scores (Cronbach, 1984). According to the well accepted classification, validity consists of content, construct and criterion related evidences. Content validity is related to the fact that the items are a sample of subject and behavior domain (Cronbach & Meehl, 1955). In the present study, scholar views were taken as a rational evidence for content validity. Four scholars' suggestions were taken into account during whole scale development process including forming item pool, modifying or deleting items that are not consistent with mathematical modeling construct. Criterion-based evidence is related to the fact that the test measures what it intended to measure (Cureton, 1951). In order to provide evidence for criterion-based validity, the correlation between developed scale and an already existed scale that measures the same construct is examined. Since there could not be found any scale that measures mathematical modeling self-efficacy, criterion-based validity evidence could not be obtained for the present scale.

Construct validity is related to the construct that test measures instead of criterion scores. Cronbach and Meehl (1955) stated that nomological networks that indicates how constructs will be measured and shows the relationships between each other are essential for construct validity. Campbell and Fiske (1959) made nomological networks more concrete and suggest multi-method multi-trait matrix to show the relationships between variables. They also suggest to analyze convergent and divergent validity evidences together when any concrete criteria do not

exist. In the present study, factor analysis was used as an empirical method to provide evidences for the construct validity of the scale. For this purpose, the structure of the scale was explored with exploratory factor analysis after item analysis of preliminary application. The obtained structure of the scale was hypothetically tested with confirmatory factor analysis. After verifying the structure of the scale, convergent validity coefficient was calculated. Due to the unidimensional structure of the scale, divergent validity coefficient could not be calculated.

3.2.1. Exploratory factor analysis.

In order to examine construct validity of the scale, first of all, an exploratory factor analysis was conducted with the data collected from 180 pre-service elementary mathematics teachers. The appropriateness of the data for the analysis was investigated by examining the results of the Keiser-Meier-Olkin (KMO) and Bartlett sphericity tests. According to Tabachnick and Fidell (2013), KMO value should be greater than 0.60 and Bartlett test result need to be significant for an exploratory factor analysis to be conducted. The scale's values for the KMO test was 0.88 and Bartlet test results were significant (χ^2=2044.23, p=0.000). Therefore, it can be said that the data were appropriate for the analysis.

According to Büyüköztürk (2013), the factors that have eigenvalues equal or greater than 1 are assumed to be significant factors. Accordingly, there are 7 significant factors that have eigenvalues equal or greater than 1. Additionally, Büyüköztürk (2013) suggests that if explained variance for a scale that designed as unidimensional is greater than 0.30, it can be accepted enough for ensuring the unidimensionality of the scale. In the present study, the first significant factor has a factor loading value of 0.342. The unidimensional structure of the scale could also be observed from scree plot. The curve of the plot decline dramatically after the first significant factor. This is also an indication for unidimensional structure of the scale. Since the scale was unidimensional as expected, there was no need for rotation.

In the present study, it was aimed that the items which have factor loadings in the first factor are expected to have factor loadings at least 0.50. For this reason, the items (21, 18, 22, 7, 4, 6, 8 and 13) having factor loadings lower than 0.50 were removed from the scale. When these items were deleted one by one, the structure of the scale varied and the items (9, 1, 19, 15, 26, 32 and 2) also had factor loadings lower than 0.50 and were removed from the scale. When all items that were removed from the scale examined they had high relationships with each other and they were lower relationship with the aim of the scale compared to other items according to scholar's views. After reducing dimensions, explanatory factor analysis was repeated with remaining items. Scree plot for dimension reduction analysis is given in Figure 2.

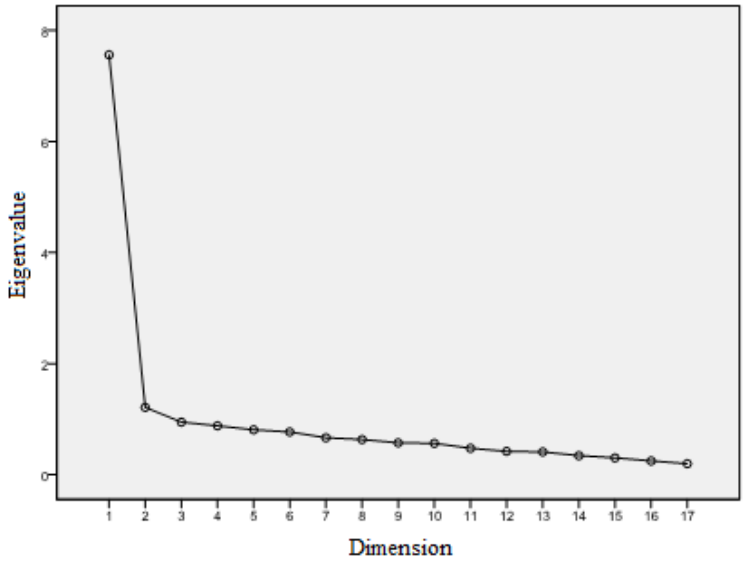


Figure 2. Scree plot

According to the scree plot given in Figure 2, the curve decreases dramatically after the first factor. The second and the other factors have the values very close to each other and the decrease between any two factors is not remarkable. Although the scree plot indicates a unidimensional structure for the scale, it is important to examine component matrix and explained variance proportions.

For the second exploratory factor analysis of remaining 17 items, the scale's values for the KMO test was 0.91 and Bartlett test results were statistically significant (χ^2=1058.85, p=0.000). Therefore, it can be said that the data were appropriate for the analysis. There were two significant factors that have eigenvalues equal or greater than 1. Explained variance for the first factor was 0.445 and hence it indicates a unidimensional scale as observed in the scree plot. All remaining items had factor loadings equal or greater than 0.56 for the first factor (Table 2).

Table 2. Factor Loadings (Λ) and Total Explained Variance

Item	*λ*	*Item*	*λ*	*Item*	*λ*	*Item*	*λ*
5	.776	30	.702	28	.638	25	.616
23	.755	16	.690	11	.638	3	.574
24	.733	27	.682	17	.637	12	.562
31	.729	14	.664	29	.620	10	.558
20	.708						
Eigenvalue = 7.561							
Total variance explained (%) = 44.476							

According to the exploratory factor analysis results, it was concluded that 17 items explained sufficiently mathematical modeling self-efficacy. In order to verify proposed scale structure by this analysis, a confirmatory factor analysis was performed.

3.2.2. Confirmatory factor analysis.

A confirmatory factor analysis was carried out to validate that the scale with 17 items is proper to measure mathematical modeling self-efficacy of pre-service teachers. LISREL 8.80 software was used to perform the analysis and obtain evidences for construct validity of the scale. In order to calculate model parameters, maximum likelihood technique was used (Jöreskog & Sörbom, 2004). Univariate and multivariate normality were checked and it was found that these assumptions were not satisfied as the prerequisite of the analysis. Therefore, a robust method for maximum likelihood technique was performed. Tabachnick and Fidell (2013) suggest carrying out the analysis with a sample of approximately 300 participants. Therefore, the scale with 17 items was applied to 310 pre-service elementary mathematics teachers from different public universities around Turkey.

KMO and Bartlett test results were examined before doing the analysis. According to Tabachnick and Fidell (2013), KMO value should be greater than 0.60 and Bartlett test result need to be significant for a confirmatory factor analysis to be conducted. The scale's value for the KMO test was 0.88 and Bartlett test result was significant (χ^2=1904.52, p=0.000). Therefore, it can be said that the data were appropriate for the analysis.

As shown in Table 3, χ^2 and χ^2/df statistics, the normed fit index (NFI), the non-normed fit index (NNFI; also known as Tucker-Lewis index), the relative fit index (RFI), the comparative fit index (CFI), the incremental fit index (IFI), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the root mean square residual (RMR), and the root mean square error of approximation (RMSEA) were used to interpret the fit of the model to the data (Kline, 2011). Among the modifications given for the decrease in χ^2 values in LISREL output, one of them, which is between items 9 and 10, was done.

Table 3. χ^2 Statistics, Error, and Fit Indices

χ^2	χ^2/df	ρ	RMSEA	NFI	NNFI	RFI	CFI	GFI	AGFI	SRMR	IFI
303.38*	2.55	.000	0.071	0.95	0.96	0.94	0.97	0.87	0.83	0.058	0.97

Notes. ρ<0.01

As shown in Table 3, χ^2 statistics is significant (ρ<0.001) and χ^2/df statistics is 2.55. Although it is advised that χ^2/df value need to be lower than 5 (Anderson & Gerbing, 1984), Kline (2011) stresses that using this value to evaluate the fit of data to the model has not any logical and statistical base. For this reason, interpreting other approximation and fit indices given in Table 3 will be much appropriate. The model RMSEA and SRMR values for the present scale were 0.071 and 0.057, respectively. The acceptable maximum cutoff value for RMSEA is 0.06 and for SRMR it is 0.08 (Hu & Bentler, 1999). However, Steiger (2007) proposes a maximum cutoff value of 0.07 for RMSEA. Hence the model has acceptable fit to the data for RMSEA and SRMR. Inversely, the acceptable value for GFI and AGFI indices is to be greater than 0.80 (Cole, 1987; Marsh, Balla & McDonald, 1988). Since the model GFI and AGFI values for the present scale was 0.87 and 0.83, respectively, the model has again acceptable fit to the data. The model values greater than 0.90 represents good fit, and greater than 0.95 indicates perfect fit of the model to the data (Hair, Anderson, Tatham & Black, 1998). Relative fit indices, RFI, IFI, and CFI have values 0.94, 0.97, and 0.97, respectively. Therefore, RFI indicates well; IFI and CFI have perfect fit of the model to the data. Normed and non-normed fit indices are also interpreted

similar to relative fit indices (Hu & Bentler, 1999). Since the model NFI and NNFI values for the present scale were 0.95 and 0.96, respectively, the model has perfect fit to the data.

The item values of standardized factor loadings (λ), unstandardized factor loadings (λ'), t values, standardized error variances (σ_e), unstandardized error variances (σ_e'), and determination coefficients (R^2) were calculated for the theoretical model and given in Table 4.

Table 4. Factor Loadings, t Values, Error Variances, and Determination Coefficients

Item	λ	λ'	t	σ_e	σ_e'	R^2
3	0.64	1	10.55	0.60	0.6	0.40
25	0.57	0.87	11.44	0.67	0.64	0.33
10	0.56	0.9	11.13	0.69	0.74	0.31
17	0.57	0.94	11.03	0.68	0.75	0.32
16	0.61	1.01	11.39	0.63	0.7	0.37
5	0.63	0.93	12.30	0.60	0.53	0.40
11	0.56	0.93	10.75	0.68	0.75	0.32
12	0.58	0.9	10.41	0.66	0.65	0.34
23	0.59	0.9	11.77	0.65	0.63	0.35
20	0.64	0.98	12.27	0.59	0.57	0.41
31	0.61	0.88	11.70	0.63	0.54	0.37
30	0.66	0.91	10.55	0.57	0.45	0.43
24	0.59	0.87	11.71	0.65	0.58	0.35
14	0.64	0.96	10.15	0.58	0.53	0.42
27	0.52	0.78	10.69	0.73	0.67	0.27
29	0.66	1.02	10.52	0.57	0.55	0.43
28	0.58	0.93	10.74	0.67	0.7	0.33

Kline (2011) suggests that the absolute values of standardized factor loadings are expected to be greater than 0.10. In addition, it is also stressed that the values lower than 0.10 indicate small effect; values between 0.30 and 0.50 represent medium effect; and values greater than 0.50 show large effect. Standardized factor loadings for the present scale vary between 0.56 and 0.66 and hence all of them indicate large effect. In addition, *t* values greater than the critical value 1.96 show that all items fit to the unidimensional model. The standardized error variances (σ_e) for the items of the present scale vary between 0.57 and 0.73. These values show that error variances are little higher than medium level. Correspondingly, explained variances vary between 0.37 and 0.43 and they are little lower than medium level. When all findings obtained from the confirmatory factor analysis were interpreted together, it was found that all 17 items fit to the theoretical model.

Convergent validity. The present scale consists of congeneric items. These items do not have equal factor loadings when compared to parallel, tau-equivalent, and essentially tau-equivalent items. Therefore, the reliability and validity coefficients for congeneric items were evaluated differently. McDonald (1985) suggests to use ω coefficient for such items. The value of this coefficient for this scale was calculated as 0.97. Campbell and Fiske (1959) proposed convergent validity to establish construct validity. Convergent validity could be evaluated by using

reliability coefficient. $\sqrt{\omega}$ is equal to the correlation between observed and true scores in classical test theory. The value 0.99 indicates that the construct validated by confirmatory factor analysis has a very high convergent validity and this value constitutes a strong evidence for construct validity of the scale. Since the scale was unidimensional, discriminant validity which shows the discrepancy between two dissimilar constructs could not be evaluated.

3.3. Reliability of the Scale

Reliability coefficient was defined differently in the literature. Gulliksen (1950) identified that it is equal to the correlation between observed scores obtained from parallel test forms. Cureton (1958) stated that the ratio of true score variance to the observed score variance corresponds to the reliability coefficient. Lord and Novick (1968) defined it as the square of correlation between true and observed scores. In order to calculate reliability coefficient corresponds to internal consistency of a scale, different reliability coefficients are used according to the equality of item means, standard deviations, error variances, and factor loadings (Yurdugül, 2006). Since items factor loadings of the present scale were not equal, ω coefficient (McDonald, 1985) was used to calculate the reliability of the scale. Kline (2011) suggested that a reliability coefficient greater than 0.90 is reliable at perfect level. As it was calculated in the equation 4, ω internal consistency coefficient of the scale is 0.97. This value indicates that the reliability of the present scale is very high.

In addition to ω coefficient, Cronbach's α coefficient was also calculated as standardized factor loadings of the items are close to each other. McDonald's ω coefficient is equal or higher than Cronbach's α coefficient in all measurements (Bacon, Sauer & Young, 1995). For the present scale, Cronbach's α reliability coefficient was calculated as 0.91. This value is the lower bound for the reliability of the scale. Although α is lower than ω coefficient, it also indicates a perfect level reliability for the present scale.

Since McDonald 's ω and Cronbach's α coefficient could have values between 0.00 and 1.00, if the reliability value found is subtracted from 1.00, the new value found indicates total observed score variance arising from random errors (Kline, 2011). When McDonald's ω and Cronbach's α coefficient for the presented scale are subtracted from 1.00, the random error variance is 3% and 9%, respectively. It means that maximum total observed score variance arising from random errors is 0.09. These findings show that the present scale has a very low total observed score variance arising from random errors.

4. CONCLUSIONS AND SUGGESTIONS

The aim of the present study is to develop a self-efficacy scale to measure pre-service elementary mathematics teachers' belief on their competencies in mathematical modeling. The scale is unidimensional and it is constructed according to Blum and Kaiser's (1997) mathematical modeling competencies framework that includes all competencies specified by Lingefjrad (2004). The final form of the scale consists of 17 items and they are in the form of Likert format which is scored 1 to 5 points. When the items with negative meaning reversed, the scale scores vary between 17 to 85 points. Higher scale score means higher level self-efficacy of mathematical modeling competencies. Indices for items, and the items included and not included in the final form are given in Appendix. The Turkish version of the final form will be provided to the researchers that are interested in self-efficacy of prospective teachers related to mathematical modeling competencies.

The validity studies revealed that the scale is verified in terms of its content and construct to be measured. The evidences are obtained by taking the ideas of scholars, teachers and students for the content of the scale and it is concluded that the scale measures what it intends to measure. Moreover, exploratory and confirmatory evidences obtained by factor analysis provided strong evidences for the construct validity of the scale. In addition, the reliability analysis revealed high level of internal consistency according to both Cronbach's and McDonald's reliability coefficients. This finding also constitutes an evidence for construct validity of the scale. When all findings are interpreted together, an appropriate tool is developed to measure pre-service teachers' self-efficacy beliefs on their mathematical modeling competencies.

Since assessing mathematical modeling performances is more complicated than expected (Blum, 1993; Lingefjard & Holmquist, 2004), this scale is considered to be a convenient tool that could be used in the field of mathematical modeling. Scholars and teachers can utilize this scale to make interpretations about students' self-efficacies which is one of the most important indicators for performance of the students as justified by some researchers (e.g., Bandura, 1997; Dede, 2008; Lee, 2009; Pajares & Graham, 1999).

In other research studies, it can also be used to investigate on the relationship between modeling performances and students' self-efficacies. Moreover, the scale could be used in mathematical modeling studies for diverse purposes such as examining the effects of other mathematical constructs, their relationship with different demographic variables, etc. In addition, evidences related to criterion-related validity, test-retest, split-half and equivalent form reliability can be collected to enhance the scale. Finally, this scale can be adapted to the high school level for different regions and countries.

5. REFERENCES

Anderson, J.C., & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155–173.

Bacon, D.R., Sauer, P.L. & Young M. (1995). Composite reliability in structural equations modeling. *Educational and Psychological Measurement, 55*(3), 394-406.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Blum, W. (1993). *Mathematical modeling in mathematics education and instruction*. Germany: Ellis Horwood Limited.

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do? In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education: Intellectual and attitudinal challenges* (pp. 73-96). New York, NY: Springer.

Blum, W., & Kaiser, G. (1997). *Vergleichende empirische Untersuchungen zu mathematischen anwendungsfähigkeiten von englischen und deutschen Lernenden* [Comparative empirical studies at mathematical application skills of English and German learners]. Unpublished manuscript, German Research Foundation, Bonn, Germany.

Brown, R. (2002). Mathematical modeling in the international baccalaureate, teacher beliefs and technology usage. *Teaching Mathematics and Its Applications, 21*(2), 67-74.

Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabi* [The data analysis manual for social sciences]. Ankara: Pegem Akademi.

Campbell, D.T., & Fiske D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.

Cole, D.A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*, 1019–1031.

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

Cronbach, L.J. (1984). *Essentials of psychological testing*. New York: Harper.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Crouch, R. & Haines C. (2004). Mathematical modeling: Transitions between the real world and the mathematical model. *International Journal of Mathematics Education Science Technology*, *35*(2), 197-206.

Cureton, E. E. (1951). Validity. In E.F. Lindquist (Ed.), *Educational measurement* (1st ed., pp. 621-694).Washington, DC: American Council on Education.

Cureton, E.E. (1958). The definition and estimation of test reliability. *Educational and Psychological Measurement, 18*, 715-738.

Dede, Y. (2008). Matematik öğretmenlerinin öğretimlerine yönelik öz-yeterlik inançları [Mathematics teachers' self-efficacy beliefs toward their teaching]. *Türk Eğitim Bilimleri Dergisi*, *6*(4), 741-757.

Erbaş, A.K., Çetinkaya, B., Alacacı, C., Kertil, M., Çakıroğlu, E., & Baş, S. (2014). Matematik eğitiminde matematiksel modelleme: Temel Kavramlar ve farklı yaklaşımlar [Mathematical modeling in mathematics education: Basic concepts and approaches]. *Kuram ve Uygulamada Eğitim Bilimleri, 14*(4), 1607-1627.

Fraenkel, J.R. & Wallen, N.E. (2011). *How to design and evaluate research in education* (6th Ed.). New York: McGraw-Hill, Inc.

Freudental, H. (1991). *Revisiting mathematics education*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Galbraith, P. (2012). Models of modelling: Genres, purposes or perspectives. *Journal of Mathematical Modeling and Application, 1*(5), 3-16.

Gravemeijer, K. (2002). Preamble: From models to modeling. In K. Gravemeijer, R. Lehrer, B. Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 7-22). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Gravemeijer, K., & Stephan, M. (2002). Emergent models as an instructional design heuristic. In K. Gravemeijer, R. Lehrer, B. Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 145-169). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Haines, C., & Crouch, R. (2001). Recognizing constructs within mathematical modelling. *Teaching Mathematics and its Applications*, *20*(3), 129-138.

Haines, C., & Crouch, R. (2007). Mathematical modeling and applications: Ability and competence frameworks. In W. Blum, P. L. Galbraith, H. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 417-424). New York, NY: Springer.

Hair, J.F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. New Jersey: Pearson Education.

Hu, L.T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Ikeda, T., & Stephens, M. (1998). The influence of problem format on students' approaches to mathematical modelling. In P. Galbraith, W. Blum, G. Booker, & I. Huntley (Eds.), *Mathematical modelling, teaching and assessment in a technology-rich world* (pp.223-232). Chichester: Horwood Publishing.

Jöreskog, K.G., & Sörbom, D. (2004). *Lisrel 8.7: Structural equation modeling with the Simplis command language*. Lincolnwood, IL: Scientific Software International.

Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM – The International Journal on Mathematics Education, 38*(3), 302-310.

Kline, R.B. (2011). *Principles and practice of structural equation modeling: Methodology in the social sciences.* New York: The Guilford Press.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and individual differences, 19*(3), 355-365.

Lehrer, R., & Schauble, L. (2003). Origins and evaluation of model-based reasoning in mathematics and science. In R. Lesh, & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 59-70). Mahwah, NJ: Lawrence Erlbaum.

Lehrer, R., & Schauble, L. (2007). A developmental approach for supporting the epistemology of modeling. In W. Blum, P. L. Galbraith, H-W. Henn, & M. Niss (Eds.), *Modeling and applications in mathematics education* (pp. 153-160). New York, NY: Springer.

Lesh, R., Cramer, K., Doerr, H. M., Post, T., & Zawojewski, J. S. (2003). Model development sequences. In R. Lesh, & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 3-33). Mahwah, NJ: Lawrence Erlbaum.

Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning, and problem solving. In R. Lesh, & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 3-33). Mahwah, NJ: Lawrence Erlbaum.

Lesh, R., & Zawojewski, J. S. (2007). Problem solving and modeling. In F. Lester (Ed.), *The handbook of research on mathematics teaching and learning* (2nd ed., pp. 763-804). Reston, VA: National Council of Teachers of Mathematics; Charlotte, NC: Information Age Publishing.

Lingefjard, T. (2004). *Assessing engineering student's modeling skills*. Retrieved from http://www.cdio.org/files/ assess_model_skls.pdf

Lingefjard T., & Holmquist, M. (2004). To assess students' attitudes, skills and competencies in mathematical modeling. *Teaching Mathematics and Its Applications, 24*, 2-3.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley,.

Maaß, K. (2006). What are modelling competencies? *The International Journal on Mathematics Education, 38*(2), 113-142.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.

McDonald, R. (1985). *Factor analysis and related methods.* Hillsdale, N J: Erlbaum.

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Niss, M. (2004). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagtsis & Papastavridis (Eds), *Proceedings of 3rd Mediterranean Conference on Mathematical Education* (pp. 115-124). Greece, Athens: The Hellenic Mathematical Society.

Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In W. Blum, P. Galbraith, H. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 3-32). New York: Springer.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement.* (Vol.10, p 1-49). Greenwich, CT: JAI Press.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*,124-139.

Profke, L. (2000). Modellbildung für alle Schüler. In Hischer, Horst (Ed.), *Modellbildung, Computer und Mathematikunterricht* (pp.24-38), Hildesheim: Franzbecker.

Ryang, D. (2014). How to develop a scale measuring an affective construct in mathematics education research. *J. Korean Soc. Math. Educ., 18*(1), 75–87.

Schunk, D.H., & Pajares F. (2005). Competence beliefs in academic functioning. In A. J. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). New York: Guilford Press.

Snyder C R & Lopez S. (2002). *Handbook of positive psychology*. Oxford University Press, US.

Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences 42* (5), 893-98.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics*. Needham Heights, MA: Allyn and Bacon.

Talim ve Terbiye Kurulu Başkanlığı (2013). *Ortaöğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı* [Secondary (Grades levels 9, 10, 11, and 12) Mathematics Curriculum]. Ankara: T.C. Milli Eğitim Bakanlığı.

Yurdugül, H. (2006). Paralel, eşdeğer ve konjenerik ölçmelerde güvenirlik katsayılarının karşılaştırılması [The comparison of reliability coefficients in parallel, tau-equivalent, and congeneric measurements]. *A.Ü. Eğitim Bilimleri Fakültesi Dergisi, 39*(1). 15-37.

Appendix

Indices for mathematical modeling self-efficacy scale

Indices	*#*	*Item*	*Inclusion*
Competencies to understand the real problem and to set up a model based on reality	1	I could understand real life problem situation by simplifying.	NI
	2	I could make assumptions to understand and interpret real life problems.	NI
	3	I could identify real life situations differently.	I
	4	I have difficulty in planning to solve a real life problem.	NI
	5	I could benefit from relations between variables to make estimations from given situation.	I
	6	I have difficulty in setting up a figure, drawing or model to describe real life situation.	NI
Competencies to set up a mathematical model from the real model	7	I have difficulty in establishing relationships between mathematical models (formula or graphics) and mathematical materials (unit cubes, geometrical strips, etc.).	NI
	8	I could not decide on relevant information to set up a mathematical model.	NI
	9	I could see mathematical relationships in real life situation.	NI
	10	I could reflect on a mathematical model in depth.	I
	11	I could use different materials to set up a mathematical model.	I
	12	I could choose appropriate mathematical notations (graphic, function, etc.) to set up a mathematical model.	I
Competencies to solve mathematical questions within this mathematical model	13	I have difficulty in understanding mathematical and cognitive processes in developing mathematical formulas or notations.	NI
	14	I could compare mathematical models developed for different problem situations.	I
	15	I could decide on how to use mathematics in different problem situations.	I
	16	I could design mathematical models for different mathematical subjects.	I
	17	I could use a formula developed for solving a math problem in developing formulas for similar problems.	I
	18	I could demonstrate a function on a graphical model.	NI
Competencies to interpret mathematical results in a real situation	19	I could interpret mathematical results in social and daily life.	NI
	20	I could apply the solution for a mathematical problem to the real life situations.	
	21	I have difficulty in understanding mathematical formulas or graphics used in other disciplines (physics, chemistry, etc.).	NI
	22	I have difficulty in interpreting mathematical formulas or graphics applied to real life situations.	NI
	23	I could generalize mathematical solutions into different real life situations.	I
	24	I could demonstrate the logic behind a mathematical formula in real life situations.	I
	25	I could develop formulas or graphics that enable to take actions for the future based on a given dataset.	I
Competencies to validate the solution	26	I could validate the model that I developed by mathematical modeling.	NI
	27	I feel confident to demonstrate the accuracy of a mathematical model.	I
	28	I could critically check the solution that I obtained by mathematical modeling.	I
	29	I could review the modeling process after developing a solution for a mathematical problem situation.	I
	30	I could develop alternative solutions during mathematical modeling process.	I
	31	I could develop creative solutions by checking possible mistakes done during modeling process.	I
	32	I could develop problems that could be solved by mathematical formulas or graphics.	NI

Notes.
NI: Items not included in the final form
I: Items included in the final form

International Journal of Assessment Tools in Education: Vol. 4, Issue 1, (2017) pp. 37-53

http://www.ijate.ne
e-ISSN: 2148-7456 © IJATI
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

"Research Article"

Developmental Mathematics Students: Who are They and What is Their Mathematics Self-Efficacy?

Ryan Baxter[1], Alan Bates[1], Adel Tawfig Al-Bataineh[1,*]

[1]Illinois State University, School of Teaching and Learning, College of Education, Campus Box 5300, Normal, Il 61790-5300

Abstract

The purpose of this quantitative study was to determine differences in developmental mathematics students' self-efficacy within the demographic data from the survey. Data from a sample of 240 Intermediate Algebra students at a single four-year university using the Mathematics Self-Efficacy Results indicate that males possess higher levels of mathematics self-efficacy and confidence with their mathematical abilities than females. Students who completed a lower developmental mathematics course prior to Intermediate Algebra possess lower levels of mathematics self-efficacy. The results of this study suggest developmental mathematics instructors should refine their teaching methodologies by incorporating strategies to increase their student's self-efficacy.

Article Info

Received
08 September 2016

Revised:
06 October 2016

Accepted
26 October 2016

Keywords:
Developmental mathematics, Higher education, Self-Efficacy, Traditional and non-traditional students, Gender, Ethnicity

1. INTRODUCTION

Most teaching professionals spend their entire careers refining their instructional methods in the pursuit of teaching excellence. This refining process continually challenges teachers to adapt teaching methodologies in order to improve student performance and engagement. However, this refining process becomes even more critical when the teaching professional teaches underprepared college students in mathematics (Smittle, 2003). Traditional and non-traditional students enroll in community colleges and universities every year lacking the foundation and skills required for college level mathematics. Students who lack the foundational skills in mathematics place into developmental mathematics courses in order to become prepared to succeed in their mathematics course (s) required for graduation. In the 1990 study by the Conference Board of the Mathematical

*Corresponding Author E-mail: atalba@ilstu.edu

2148-7456 /© 2017 DOI: 10.21449/ijate.264026

Sciences (CBMS) (Albers, Loftsgaarden, Rung, & Watkins, 1992), it was reported that 56% of all students studying mathematics at two-year colleges were studying at the remedial level. Nearly a decade later, the Fall 2000 CBMS survey reported 60% of all students enrolled in two-year colleges annually take remedial math courses (Lutzer, Maxwell, & Rodi, 2002). Instructors for developmental mathematics courses at these higher education institutions serve as a gateway for success throughout a student's collegiate career. These instructors face the challenge of building, or rebuilding, the necessary foundation of mathematical skills and attitudes necessary to succeed in credit generating mathematics courses required for graduation.

Instructors face this challenge of building mathematical skills and attitudes upon an insecure mathematical foundation many students bring into colleges and universities. Unfortunately, the average United States student's ability in math, according to the National Assessment of Educational Progress (NAEP), SAT, and ACT scores, is not keeping pace with society's demands (Scherer, 2002). According to Scherer (2002), the average SAT math score has only increased three points since 1967 and only forty percent of these students earned a score of 22 or higher on the ACT math portion (the equivalent of predicting a C or better in a first-year college-level course). With such a high percentage of college students not prepared for college mathematics, something must be done differently to prepare students and help develop mathematical skills at the college level. More specifically, developmental mathematics students have generally been unsuccessful with traditional instructional methods and materials. Effective developmental mathematics teachers must be able to present mathematics in different ways, requiring teachers to have in-depth knowledge of the concepts and skills they are teaching as well as higher level content knowledge in the field (Smittle, 2003).

Along with providing different teaching strategies, another key component of a developmental mathematics course is to raise the self-efficacy of the developmental mathematics student. Much research has focused on mathematics anxiety and achievement yet little research exists on the factors affecting the self-efficacy of the developmental mathematics student. Understanding the self-efficacy of developmental mathematics students could lead to intervention strategies or teaching strategies aimed to promote a positive sense of mathematical ability which influences mathematics achievement (Pajares, 2002). Developmental mathematics instructors seeking to refine their teaching methodologies should strive to improve mathematical ability while simultaneously improving student's self-efficacy. Increasing these two components in the developmental mathematics classroom begins the building, or rebuilding, process of a solid mathematical foundation for underprepared mathematics students.

1.1 Statement of the Problem

According to the National Center for Education Statistics (NCES), in 1999-2000, 32% of all freshmen in four-year colleges and universities along with 41% of community college freshmen required developmental education, which includes developmental mathematics (NCES, 2001, as cited in Smittle, 2003). Other studies indicate about 40% of traditional undergraduates take at least one such course (Woodham, 1998, as cited in Attewell, Lavin, Domina, & Levey, 2006). Developmental mathematics educators have been attempting to improve struggling learner's ability to learn and succeed in mathematics for decades. Students' ability to succeed in college level mathematics courses are required for graduation and seems to be a "determinate of not only choice of a college major but also serves as a determinant in the acquisition of a college degree" (Hall & Ponton, 2005, p. 26). If students cannot successfully complete their developmental mathematics course (s), and then proceed to successfully complete their required mathematics

courses, they cannot graduate. Students in developmental mathematics courses show promise of succeeding at the college level by displaying strengths in some academic areas but they show weakness and struggle with mathematics (Attewell et al., 2006).

Many struggling learners believe they cannot succeed in school (Pajares, 2003). When developmental mathematics students enter the mathematics classroom, they bring negative past experiences and most believe that they will not do well in this mathematics course. This belief is also referred to as a student's mathematics self-efficacy. Self-efficacy refers to "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). Over three decades of research findings "amply support the contention that students' self-efficacy beliefs powerfully affect their academic performance in various ways" (Mills, Pajares, Herron, 2007, p. 417). Also, low self-efficacy beliefs "impede academic achievement and, in the long run, create self-fulfilling prophecies of failure and learned helplessness that can devastate psychological well-being" (Margolis & McCabe, 2006, p. 219). With almost one-third of new students entering colleges and universities taking developmental courses, what factors impact the self-efficacy of developmental mathematics students? What instructional strategies are instructors utilizing to ensure students increase their self-efficacy while simultaneously becoming competent to complete the required courses?

The purpose of this quantitative study is to determine differences in developmental mathematics students' self-efficacy, within the demographic data from the survey, based upon the Mathematics Self-Efficacy Scale results.

2. REVIEW OF RELATED LITERATURE

Developmental mathematics instructors face this challenge of building mathematical skills and attitudes upon an insecure mathematical foundation many developmental mathematics students bring into colleges and universities. Students in developmental mathematics courses show promise of succeeding at the college level by displaying strengths in some academic areas but they show weakness and struggle with mathematics (Attewell et al., 2006). However, if students cannot successfully complete their developmental mathematics course (s), and then proceed to successfully complete their required mathematics courses, they cannot graduate. Since students' ability to succeed in college level mathematics courses is required for graduation, this requirement seems to be a "determinate of not only choice of a college major but also serves as a determinant in the acquisition of a college degree" (Hall & Ponton, 2005, p. 26). Developmental mathematics students have generally been unsuccessful with traditional instructional methods and materials. Effective developmental mathematics instructors must be able to present mathematics in different ways, requiring teachers to have in-depth knowledge of the concepts and skills they are teaching as well as higher level content knowledge in the field (Smittle, 2003). With such a high percentage of college students not prepared for college mathematics, something must be done differently to develop students' mathematical skills and appropriate attitudes to be successful at the college level.

One such difference in the developmental mathematics classroom is the holistic approach taken to prepare students to succeed. Many struggling learners believe they cannot succeed in school (Pajares, 2003). More specifically, when developmental mathematics students enter the mathematics classroom, they bring negative past experiences, usually some apprehension, and most believe that they will not do well in this mathematics course. This belief is also referred to as a student's self-efficacy toward mathematics. Self-efficacy refers to "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura,

1997, p. 3). Over three decades of research findings "amply support the contention that students' self-efficacy beliefs…powerfully affect their academic performance in various ways" (Mills et al., 2007, p. 417). Also, low self-efficacy beliefs "impede academic achievement and, in the long run, create self-fulfilling prophecies of failure and learned helplessness that can devastate psychological well-being" (Margolis & McCabe, 2006, p. 219). With almost one-third of new students entering colleges and universities taking developmental courses, what approach does the developmental educator take to increase not only mathematical skills but the self-efficacy of the student? Increasing these two components in the developmental mathematics classroom lays the groundwork for building, or rebuilding, a solid mathematical foundation for underprepared mathematics students.

Understanding the factors that impact the self-efficacy of developmental mathematics students is the focus of this study. Understanding the self-efficacy of developmental mathematics students could lead to intervention strategies or teaching strategies aimed to promote a positive sense of mathematical ability which influences mathematics achievement (Pajares, 2002). This literature review discusses the background of developmental education and placement into developmental mathematics courses. The literature then examines self-efficacy related to academic achievement; the sources of self-efficacy; self-efficacy regarding gender and race; and a brief history of assessing mathematics self-efficacy. This literature review primarily focuses on articles describing college students and not articles discussing middle or high school students. Topics not thoroughly discussed include self-efficacy relating to self-regulation and self-efficacy relating to self-concept. Following the literature is a summary of the research findings.

2.1 Research Questions

1. Which factors have an effect on developmental mathematics students' self-efficacy?
2. What is the relationship of developmental mathematics students' MSES scores in regards to gender?
3. What is the relationship of developmental mathematics students' MSES scores with race and gender?
4. What implications are evident by analyzing the developmental mathematics students' self-efficacy within the developmental mathematics classroom?

2.2 Definition of Terms

Affective domain: This refers to an emotional component. It consists of attitudes, or one's tendency to respond in a certain way, along with memories of past failures and successes. Affective variables include math anxiety, self-confidence in learning and doing mathematics, liking or disliking mathematics, interest in mathematics, attributions for success and failure in mathematics, as well as beliefs about oneself as a learner of mathematics, and beliefs about math's usefulness (Bassarear, 1991).

Attitude toward mathematics: This may be defined as "the level of like or dislike felt by an individual toward mathematics" (Quinn, 1997, p. 108).

COMPASS mathematics placement exam: COMPASS is defined by Illinois State University as a Placement Exam determines which math courses students are eligible to take at Illinois State University. Placement exam results are provided to assist in determining initial placement in a math course and are discussed with Academic Advisors. All examinees receive questions in Algebra. Depending on their Algebra score, they are then be routed into either Pre-Algebra

questions OR College Algebra and Trigonometry questions. COMPASS is an untimed exam. The average time needed to complete the exam is less than one hour.

Cognitive Domain: "The logical component that processes thought, that stores and retrieves information, that deals with aptitude for learning math, and that matches learning readiness to teaching strategies" (Martinez & Martinez, 1996, as cited in Shields, 2006).

Developmental Mathematics Student: A student displaying moderate skill deficiencies in mathematics that requires cognitive and affective growth before enrolling in credit courses. Throughout this process of growth, the student is expected to function adequately. Students are placed in developmental mathematics courses following a mathematics placement exam (Shields, 2006).

Math Anxiety: "Feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematics problems in a wide variety of ordinary life and academic situations" (Richardson & Suinn, 1972, p. 551).

Mathematics Self-Efficacy Scale: This self-report instrument is useful for measuring college students' mathematics self-efficacy and consists of two subscales. The Likert-style questionnaire consists of 34 self-reported items where the student rates his or her level of confidence relating to mathematical tasks. The purpose of the Mathematics Tasks subscale is to measure student confidence in the ability to perform everyday mathematical tasks. The purpose of the Mathematics Courses subscale is to assess student confidence in their ability to earn a B or better in college courses that require mathematical skills (Hall & Ponton, 2005).

Self-Efficacy: This concept refers to "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3).

Self-regulation: This concept refers to a "metacognitive process that requires students to explore their own thought processes so as to evaluate the results of their actions and plan alternative pathways to success" (Usher & Pajares, 2009, p. 443). Furthermore, successful self-regulating students organize their work, set proximal and distal goals, seek help when needed, and manage their time well.

Traditional students: This concept refers to students are who are often below the age of 24. They enroll in college immediately after graduation from high school and pursue undergraduate education on a full-time basis. Most traditional students are financially dependent on others and are employed only on a part-time basis. They often do not have family with children and education is their primary responsibility.

Nontraditional students: This concept refers to students are who considered to be adult learners who often have family and work responsibilities. They are often over the age of 24 and return to college to seek out additional education that is necessary for job transitioning in the workforce.

2.3 Limitations of the Study

This study has been limited to adult college students from developmental mathematics courses enrolled in MAT 104: Intermediate Algebra during the Spring 2010 semester. Adult students will be classified as traditional or non-traditional students during the data collection. The data collected may not be representative of the entire population regarding the self-efficacy of developmental mathematics students. Even though Informed Consent forms are given before

research begins, and the course instructor will not be present when students fill out the survey, some students could feel the answers they provide may affect their grade in some way. The sample size is limited due to the following qualifications: participants who volunteered, were 18 years of age or older, and placed in developmental mathematics courses for the Spring 2010 semester. There is an assumption the participants accurately and honestly responded to the survey and demographic questions. For this reason, it is assumed the data is accurate to the best of the students' abilities.

3. METHODOLOGY

3.1 Participants

The sample population consisted of 240 male and female college students who were eighteen or older and enrolled in MAT 104: Intermediate Algebra for the Spring 2010 semester at a Midwestern four-year public university. From the 240 participants, 79 (33%) of the students were male, and 158 (66%) of the students were female (see Table 1). Although a more balanced sample would have been ideal, more women are typical in developmental mathematics classes. Analyzing the university's Fall 2009 student enrollment data showed 11886 (57%) undergraduate students were female compared to 8970 (43%) male students (University Facts, 2009). Although women represented a vast majority of the sample, this would probably be true in most courses at this university and is typically the case in developmental mathematics courses. The sample consisted of many racial backgrounds including 4 (2%) American Indian/Native Alaskan students; 41 (17%) Black/Non-Hispanic students; 6 (3%) Asian/Pacific Islander students; 23 (10%) Hispanic students; 158 (66%) White/Non-Hispanic students; and 8 (3%) students classified themselves as other. The sample had 24 (10%) non-traditional students with 207 (86%) students classified as traditional students. Ninety-six (40%) of the sample completed the Basic Algebra course prior to enrolling in Intermediate Algebra. Almost all students enroll in developmental mathematics courses based upon completion of the COMPASS mathematics placement exam. The rationale for selecting developmental mathematics students in only Intermediate Algebra was due to the researcher's belief students who place into this level of mathematics will demonstrate a low level of self-efficacy toward mathematics.

Table 1. Descriptive Data of the Sample

Category	Description	N	%
Gender	Male	79	32.9
	Female	158	65.8
	Not Indicated	3	1.3
	Total	240	
Race	American Indian/Native Alaskan	4	1.7
	Black/Non-Hispanic	41	17.1
	Asian/Pacific Islander	6	2.5
	Hispanic	23	9.6
	White/Non-Hispanic	158	65.8
	Other	8	3.3
	Total	240	
Minority vs. Majority	Minority	82	34.2
	White/Non-Hispanic	158	65.8
	Total	240	
Credit Hours Earned	0-29	150	62.5
	30-59	48	20.0
	60-89	27	11.3
	90+	9	3.8
	Unsure	1	0.4
	Not Indicated	5	2.1
	Total	240	
Completed Basic Algebra	Yes	96	40.0
	No	142	59.2
	Not Indicated	2	0.8
	Total	240	
Repeated Intermediate Algebra	Yes	37	15.4
	No	203	84.6
	Total	240	
Student Status	Traditional	207	86.3
	Non-Traditional	24	10.0
	Not Indicated	9	3.8
	Total	240	

Due to the quantitative nature of this study, the convenience sampling includes participants from all eleven sections of Intermediate Algebra courses offered in the Spring 2010 semester.

3.2. Instrumentation

Participants were asked to complete the Mathematics Self-Efficacy Scale (MSES). The intended outcome of the MSES was to accurately measure student confidence in the ability to perform every day mathematical tasks. The MSES was originally developed in 1983 by Betz and Hackett and contained 75 items. However, after a revision in 1993, the survey became more concise and now contains 34 items. The MSES contains a Mathematics Tasks subscale and a Mathematics Courses subscale. The purpose of the Mathematics Tasks subscale measures student confidence in the ability to perform everyday mathematics tasks. The purpose of the Mathematics Courses subscale is to determine student confidence in their ability to earn a B or better in college courses that require mathematical skills (Betz & Hackett, 1993). Betz & Hackett (1983) reported internal consistency using the coefficient alpha to be .96 for the total scale and .92, .96, and .92 for the Tasks, Problems, and Courses subscales, respectively. Lent et al. (1991) reported a coefficient alpha of .92 and a two-week test-retest reliability of .94. Based upon the findings in this current study, our Cronbach's alpha is .95. Therefore, the findings in this research are consistent with previous reports and should be considered reliable and valid data.

The Likert-style questionnaire consists of 34 self-reported items where the student rates his or her level of confidence. Participants rate their level of confidence using categories such as "no confidence at all," "very little confidence," "some confidence," "much confidence," or "complete confidence." Scoring for each question ranges from 0 = no confidence at all, to 9 = complete confidence. To compute the MSES score, the mean of all 34 items is calculated. The range of MSES scores could fall between 0.000 and 9.000. If a student failed to respond to an item, the sum is calculated based upon the items that were completed. However, if more than 3 items were not completed, the survey is not considered valid. Such surveys were not included in the sample for this research study. Table 2 provides the approximate percentile equivalents to aid in interpreting the MSES scores. These percentile equivalents are separated by gender since significant gender differences were found when creating the mathematics self-efficacy scale. For example, if a female receives a MSES score of 6.223, she falls within the 60-70th percentile. This result indicates the female participant exhibits a stronger sense of mathematics confidence than approximately 65% of the female population. However, if a male receives a MSES score of 6.223, he falls within the 40-50th percentile. This result indicates the male participant exhibits a stronger sense of mathematics confidence than approximately 45% of the male population.

This instrument was selected due to its reliability and validity to measure college-level students' mathematics self-efficacy. Furthermore, Betz and Hackett (1993) note that the content validity for the MSES has been demonstrated through research that validates each area measured by the instrument. The MSES has a positive correlation between other mathematics scales such as math anxiety (r = .56), confidence in doing mathematics (r = .66), perceived usefulness of mathematics (r = .47), and the effectance motivation in math (r = .46), thus enhancing the validity of this instrument. Permission was granted to print and distribute the MSES on January 21, 2010 for up to 300 participants (see Appendix B).

Table 2. Approximate Percentile Equivalents for Mathematics Self-Efficacy Scores

	Total Score	
Percentile	**Females**	**Males**
95	7.9	8.5
90	7.5	8.1
80	6.9	7.5
70	6.5	7.1
60	6.1	6.7
50	5.8	6.4
40	5.5	6.1
30	5.1	5.7
20	4.7	5.3
10	4.1	4.7
5	3.7	4.3

4. RESULTS

An independent samples *t*-test was conducted to determine if a significant difference between the mathematics self-efficacy of male ($N = 79$) and female ($N = 158$) students as measured by the MSES exists. The mean MSES score for all male students in Intermediate Algebra was 5.977 ($SD = 1.174$); and the mean MSES score for all female students was 5.243 ($SD = 1.272$). The results of the *t*-test ($t = 4.293$, $p = .000$) suggested that the means are not equal. Therefore, a significant difference exists between the level of mathematics self-efficacy for male and female students in Intermediate Algebra. Male students demonstrate a higher level of mathematics self-efficacy than female students.

A one-way analysis of variance (ANOVA) was performed to determine the main effects of race on the dependent variable for MSES score. Tukey and Scheffe post hoc tests were performed to further analyze the interactions between the individual racial groups (see Table 3). The results indicate no significant differences between and within racial groups $F(5, 234) = 1.290$, $p = .269$. The Tukey post hoc results indicate no significant difference exists between racial groups ($p = .120$). Similarly, the Scheffe post hoc results demonstrate no significant difference exists between racial groups ($p = .274$). Since race could interact with gender to create significant differences based upon race and gender, a multivariate analysis of variance (MANOVA) was performed. Similar to the ANOVA results, the MANOVA results indicate no interaction between race and gender. Race does not have statistical significance and does not appear to be a significant factor in the mathematics self-efficacy of Intermediate Algebra students.

Since many of the racial groups consist of very few participants, the relationship between all minority students and White/Non-Hispanic students was examined. An independent samples *t*-test was conducted to determine if there was a significant difference between the minority group ($N = 82$) and the White/Non-Hispanic group ($N = 158$). The results indicate the mean MSES score of the minority students was 5.332 ($SD = 1.366$); and the White/Non-Hispanic students had a mean MSES score of 5.584 ($SD = 1.244$). The results of the *t*-test ($t = 1.436$, $p = .152$) indicate there is no significant difference in the MSES scores comparing minority students to White/Non-Hispanic students. Combining the MANOVA results, prior ANOVA results, and the results from this *t*-test suggest that race, as its own stand alone variable, does not have a significant effect on MSES

scores. Therefore, the results suggest race does not have statistical significance regarding the mathematics self-efficacy of Intermediate Algebra students.

Table 3. Tukey and Scheffe post hoc tests of MSES Score

	Race	N	MSES Score
Tukey HSD[a]	Other	8	4.6738
	Hispanic	23	5.2980
	Black/Non-Hispanic	41	5.3380
	White/Non-Hispanic	157	5.5759
	Asian/Pacific Islander	6	5.7585
	American Indian/Native Alaskan	2	6.2794
	Sig.		.150
Scheffe[a]	Other	8	4.6738
	Hispanic	23	5.2980
	Black/Non-Hispanic	41	5.3380
	White/Non-Hispanic	157	5.5759
	Asian/Pacific Islander	6	5.7585
	American Indian/Native Alaskan	2	6.2794
	Sig.		.319

An independent samples *t*-test was conducted to determine if there was a significant difference between the mathematics self-efficacy of non-traditional (N = 24) and traditional students (N = 207) as measured by the MSES. Although the non-traditional students do not have enough participants in their group to display statistical significance, the *t*-test was still completed. The mean MSES score for non-traditional students was 5.832 (SD = 1.427); and the mean MSES score for all traditional students was 5.494 (SD = 1.286). The results of the *t*-test (t = -1.206, p = .229) suggest there is no significant difference in the MSES scores when comparing non-traditional and traditional students.

An independent samples *t*-test was conducted to determine if there was a significant difference between the mathematics self-efficacy of students who placed into a lower developmental mathematics course (N = 96) and students who placed directly into Intermediate Algebra (N = 142) as measured by the MSES. The mean MSES score for all students placing in a lower course was 5.290 (SD = 1.320); and the mean MSES score for all students placing directly into Intermediate Algebra was 5.655 (SD = 1.253). The results of the *t*-test (t = -2.154, p = .032) suggest that the means are not equal. Therefore, a significant difference exists between the mathematics self-efficacy of students placing into a lower developmental mathematics course prior to enrolling in Intermediate Algebra. Students placing directly into Intermediate Algebra demonstrate a higher level of mathematics self-efficacy than students who place into Basic Algebra, or lower, developmental mathematics course.

An independent samples *t*-test was conducted to determine if there was a significant difference between the mathematics self-efficacy of students repeating Intermediate Algebra (N = 37) and students enrolled in Intermediate Algebra for the first time (N = 203) as measured by the

MSES. The mean MSES score for all students repeating Intermediate Algebra was 5.649 (*SD* = 1.261); and the mean MSES score for all students enrolling in Intermediate Algebra for the first time was 5.470 (*SD* = 1.296). The results of the *t*-test (*t* = .773, *p* = .440) suggest there is no significant difference in the MSES scores of students enrolling for the first time in Intermediate Algebra in comparison to students repeating the course.

A one-way analysis of variance was performed to determine the main effects of credit hours earned on the dependent variable MSES score (see Table 4). The results indicate no significant differences between and within credit hour groups $F(6, 233) = 1.544$, $p = .165$. These results suggest that the status of a student based upon credit hours does not have statistical significance regarding the mathematics self-efficacy of Intermediate Algebra students.

Table 4. One-Way Analysis of Variance of MSES Score by Credit Hours

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.201	6	2.534	1.544	.165
Within Groups	382.252	233	1.641		
Total	397.453	239			

A two-way analysis of variance was performed to determine the main effects and interactions of credit hours and gender on the dependent variable MSES score (see Table 5). The results indicate a significant main effect for gender ($F = 6.321$, $p = .013$) and no significant interaction between gender and credit hours on MSES score ($F = .280$, $p = .840$). Consistent with our gender *t*-test results, gender displays a significant difference in this ANOVA. However, when gender and credit hours are combined, this interaction does not appear to be a significant factor in the mathematics self-efficacy of Intermediate Algebra students.

Table 5. Two-Way Analysis of Variance of MSES Score by Gender and Credit Hours Earned

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	44.672[a]	10	4.467	2.925	.002
Intercept	997.257	1	997.257	653.016	.000
Gender	9.653	1	9.653	6.321	.013
Credit Hours	13.586	6	2.264	1.483	.185
Gender * Credit Hours	1.281	3	.427	.280	.840
Error	345.137	226	1.527		
Total	7527.450	237			
Corrected Total	389.809	236			

a. R Squared = .115 (Adjusted R Squared = .075)

5. DISCUSSION

Based upon the research findings in this study, students enrolled in Intermediate Algebra possess low levels of mathematics self-efficacy. Students with low levels of self-efficacy often tend to complete only simple academic tasks where they apply the minimal amount of effort necessary and do not persist when the task becomes challenging. In other instances, students will choose not to complete the academic assignment altogether. (Mills et al., 2007; Margolis & McCabe, 2006). The current study found the mean MSES score for all male students in Intermediate Algebra was 5.977 (*SD* = 1.174), which falls somewhere in the 30-40th percentile when interpreting the results based upon the approximate percentile chart provided by Betz and Hackett (1993). This indicates male students in Intermediate Algebra rank in the 30-40th percentile when comparing the mathematics self-efficacy of all male mathematics students. The findings are similar to previous research with Intermediate Algebra students. Hall and Ponton (2005) found male Intermediate Algebra students possessed a mean MSES score of 5.392 (*SD* = 1.301). However, an interesting difference between the studies indicates Hall and Ponton's (2005) male students would fall into the 20-30th percentile based upon the percentile chart. When comparing the two groups of male students, the mean MSES scores could be significantly different. Although the findings from each study could produce significant differences, both research studies consistently demonstrated male Intermediate Algebra students possess low levels of mathematics self-efficacy. Both studies confirm that male college students in Intermediate Algebra display less confidence in their mathematics abilities.

Examining the mean MSES score for all female students in Intermediate Algebra compared to previous research displayed very consistent findings. This research found the mean MSES score for female students in Intermediate Algebra was 5.243 (*SD* = 1.272), which falls into the 30-40th percentile based upon the approximate percentile chart provided by Betz and Hackett (1993). Similar to previous research, Hall and Ponton (2005) found female Intermediate Algebra students possessed a mean MSES score of 5.294 (*SD* = 1.545). The female Intermediate Algebra students in the Hall and Ponton (2005) study fell into the into the 30-40th percentile based upon the percentile chart. Similar to the male findings, both research studies consistently demonstrate female Intermediate Algebra students possess low levels of mathematics self-efficacy.

Although male and female Intermediate Algebra students possess low levels of mathematics self-efficacy, one of the significant findings from this research indicated female students possess lower levels of mathematics self-efficacy than male students. The results from this study support previous research indicating females possess lower levels of mathematics self-efficacy than males (Betz & Hackett, 1983; Lent, Lopez, Brown & Gore, 1996; O'Brien, Martinez-Pons, & Kopala, 1999; Pajares, 2002). On the other hand, the results from this study do not support previous research by Hall and Ponton (2005) who found no significant difference regarding mathematics self-efficacy between males and females. The mean MSES score for the entire sample of this research study was 5.498 (*SD* = 1.290) while Hall and Ponton (2005) found Intermediate Algebra students possessed a mean MSES score of 5.33 (*SD* = 1.447). With very similar mean MSES scores, some possible differences in the studies should be considered. The current research has a larger sample size of Intermediate Algebra students (*N* = 240) compared with Hall and Ponton's (2005) sample size (*N* = 105). Breaking down the sample sizes indicates females (*N* = 158) have almost twice the amount of participants in this sample compared to our males (*N* = 79). Whereas Hall and Ponton (2005) had females (*N* = 63) in a lower ratio compared to males (*N* = 42). Another variable to consider within the sample includes how Hall and Ponton (2005) only analyzed Intermediate Algebra students who were of freshmen status. By not including all participants

enrolled in Intermediate Algebra, like our sample, this could cause some differences in the data. However, the current research had 146/237, or 62%, that are considered freshmen status. Although multiple variables could be analyzed within each study, most research concludes that females possess lower levels of mathematics self-efficacy.

When determining what factors impact the mathematics self-efficacy of developmental mathematics students, one main factor in this study, along with previous research, are the gender differences. Developmental mathematics instructors need to be aware of the "gender dynamic routinely at work in the classroom and strive to involve the minority gender in discussions on content. Teachers need to remember that teacher gender also can influence participation from students and work to include both males and females in questions and answers" (Waycaster, 2001, p. 413). Developmental mathematics instructors should be aware how their gender could influence participation in the classroom. This study found females possess lower levels of self-efficacy than males. Females also consist of the majority gender in the developmental mathematics classroom. Instructors should make an effort to keep a balance of students from each gender responding to questions and providing answers. Also, previous research has indicated "social persuasions and vicarious experiences were critical sources of women's self-efficacy beliefs, and that they recalled those types of incidents to a greater extent than they recalled performance accomplishments" (Zeldin et al., 2007, p. 1039). As developmental mathematics instructors attempt to strengthen self-efficacy toward mathematics in each gender, females tend to improve their self-efficacy through cooperative learning in groups or finding a role model to provide support. However, males tend to rely more on successful experiences from previous attempts and working through the material on an individual level to build their confidence. Instructors could provide this opportunity in class or could possibly assign a group homework assignment. However, having students find their own group of students they are comfortable with would be ideal. Each gender may follow different paths in order to improve confidence and mathematics self-efficacy. Instructors should "convey the message that academic success is a matter of desire, effort, and commitment rather than of gender or established social structure" (Pajares, 2002, p. 123). Developmental mathematics instructors should attempt to incorporate multiple learning opportunities throughout their course in order to enhance the self-efficacy for all students which will ultimately enhance academic achievement.

While gender differences were significantly different, this study did not find any significance between male and female Intermediate Algebra students when comparing racial backgrounds. Based upon previous research, minority students have consistently demonstrated lower self-efficacy than White/Non-Hispanic students (Stevens et al., 2004; O'Brien et al., 1999). Due to the limited number of research studies involving mathematics self-efficacy and race, previous research has only focused on high school students. Comparing high school students with college students causes some concern. Typically only the students who were academically successful in high school transition to the university level. The minority students at the university level were probably the more self-efficacious students from their high schools. Although no significant differences were found between the racial background of Intermediate Algebra students, examining the demographics of the university compared to the Intermediate Algebra sample demonstrates some stark differences. For example, this study has 34% of the sample being labeled as minority students. The university's overall demographics indicate 17% of the student body is minority students (University Facts, 2009). Intermediate Algebra students enroll twice the number of minority students than is typical at this university. Other interesting numbers indicate that this sample had 17% being labeled as Black/Non-Hispanic and 10% labeled as Hispanic. The

university's overall demographics include only 6% of the student body being labeled Black/Non-Hispanic and 4% as Hispanic. Once again, our ANOVA'S indicate race does not significantly impact Intermediate Algebra student's mathematics self-efficacy. This would suggest that some extraneous variables we have not measured are somehow influencing the higher percentage of minority students in developmental courses.

Racial backgrounds did not significantly influence MSES scores and the same could be said about traditional students and non-traditional students. Cassazza (1999) has shown that the fastest growing segment of higher education is the number of non-traditional learners. Hall and Ponton (2005) called for more research involving the mathematics self-efficacy of non-traditional students and traditional students. In response to this call, this study compared the mathematics self-efficacy of non-traditional (N = 24) and traditional students (N = 207) as measured by the MSES. Although Cassazza (1999) determined this was a fast growing segment of the higher education population, the non-traditional students were not well represented in this sample. With only 24 participants in the non-traditional group, this sample size does not have enough participants for statistical purposes. The results also indicated no significant differences between traditional and non-traditional students. The comparison between traditional and non-traditional students should be conducted at the community college where a higher number of students likely would be classified as non-traditional students.

Similar to the traditional and non-traditional findings, when comparing credit hour information and Intermediate Algebra students, no significant differences were found. Previous research by Hall and Ponton (2005) found a significant difference between the mathematics self-efficacy of freshmen Calculus I students compared to freshmen Intermediate Algebra students. The Calculus I students displayed significantly higher levels of mathematics self-efficacy. The current study compares students only enrolled in Intermediate Algebra and shows no significant differences based upon credit hour status. With 62% of students being freshmen, 20% being sophomores, 11% being juniors, and only 4% labeled themselves as seniors, there are no significant differences in mathematics self-efficacy based upon whether you are a freshmen, sophomore, junior, or senior, enrolled in Intermediate Algebra. Since our sample of junior students (N = 27) and senior students (N = 9) are small, the statistical validity of comparing freshmen with seniors is not completely accurate. However, no significant differences occurred between and within gender and credit hours. This interaction does not appear to be a significant factor in the mathematics self-efficacy of Intermediate Algebra students.

Along with credit hour status not being a significant difference, students enrolling in Intermediate Algebra for a second or third attempt did not show any significance when compared to students enrolling for the first time. Based upon Bandura's (1986, 1997) self-efficacy theory, students who have failed to complete Intermediate Algebra on the first attempt would probably possess lower levels of self-efficacy. Mastery experiences are the primary source of self-efficacy information for almost every person. It could be assumed that withdrawing from, or failing, a course would negatively impact student's mathematics self-efficacy. However, the results indicated that students who are repeating Intermediate Algebra (N = 37) are not significantly different than students who are taking Intermediate Algebra for the first time (N = 203) based upon MSES scores. Some possible explanations include the fact that females comprise a strong majority of the students in Intermediate Algebra courses. Previous research indicated "social persuasions and vicarious experiences were critical sources of women's self-efficacy beliefs, and that they recalled those types of incidents to a greater extent than they recalled performance accomplishments" (Zeldin et al., 2007, p. 1039). Females may not necessarily gauge a tremendous

amount of their mathematics self-efficacy from mastery experiences, or failed experiences, in this case. The results clearly show no significance between groups of students repeating Intermediate Algebra and students enrolling in Intermediate Algebra for the first time. This is encouraging to developmental mathematics instructors as they attempt to build repeat Intermediate Algebra student's mathematics self-efficacy. Knowing students enrolled in Intermediate Algebra for a second, or third attempt, do not have lower mathematics self-efficacy is one less barrier impeding student's academic achievement in a mathematics course.

Conversely, students enrolled in Intermediate Algebra after completing a lower developmental mathematics course did have significantly lower levels of mathematics self-efficacy compared to students placing directly into Intermediate Algebra. Previous research by Hall and Ponton (2005) found a significant difference between the mathematics self-efficacy of freshmen Calculus I students compared to freshmen Intermediate Algebra students. The Calculus I students displayed significantly higher levels of mathematics self-efficacy. Calculus I and Intermediate Algebra are separated by usually three or four mathematics courses. By only separating the students with one mathematics course, the results from this current study support Hall and Ponton (2005) findings of significant differences between two mathematics courses. The results from this study also support previous research indicating students who perform at lower academic levels report significantly less self-efficacy than students operating at higher academic levels. Whether comparing gifted and regular students (Zimmerman & Martinez-Pons, 1990) or regular and low achieving students (Multon, Brown, & Lent, 1991), the students in the higher courses displayed significantly higher self-efficacy than students in the lower level courses. The current study is in contrast to previous research by Young and Ley (2002) which found students in developmental mathematics courses had similar levels of mathematics self-efficacy as students placing in regular mathematics courses. Based upon Bandura's (1997) self-efficacy theory, mastery experiences, or previous experiences, with mathematics is the most influential aspect of a person's self-efficacy. Students taking a lower mathematics course than Intermediate Algebra have probably experienced many negative feelings toward mathematics by not having experienced much success in previous mathematics courses. The results from this study seem to support the principles of self-efficacy theory and previously established research.

5.1. CONCLUSION

Developmental mathematics educators have been attempting to improve underprepared student's mathematical abilities for decades. Increasing the developmental mathematics student's self-efficacy should improve their confidence in mathematics while simultaneously improving their mathematical ability. Adjusting instructional methodologies to incorporate more mastery experiences, verbal persuasion, and cooperative learning in the classroom are some strategies for instructors to implement. However, instructors must realize there are no quick fixes to improve low levels of mathematics self-efficacy. Students with low levels of mathematics self-efficacy have faced an uphill battle to comprehend mathematics for many years. Nevertheless, continual attempts, however slow and arduous, to improve developmental mathematics student's self-efficacy should be implemented by developmental mathematics instructors to improve student's chances of successfully completing their college degree. As developmental mathematics educators seek teaching excellence, incorporating the improvement of developmental mathematics students' self-efficacy throughout the process of improving instructional methods should lay the groundwork for building a solid mathematical foundation for underprepared students to succeed.

6. REFERENCES

Albers, D. J., Loftsgaarden, D. O., Rung, D. C., & Watkins, A. E. (1992). *Statistical abstracts of undergraduate programs in the mathematical sciences and computer science in the United States*. 1990-91 CBMS Survey (MAA Notes Number 23). Washington DC: Mathematical Association of America.

Attewell, P., Lavin, D., Domina, T., Levey, T. (2006). New evidence on college remediation. *The Journal of Higher Education, 77*(5), 886-924.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). Self-Efficacy: The exercise of control. New York: W.H. Freeman and Company.

Bassarear, T. J. (1991). An examination of the influence of attitudes and beliefs on achievement in a college developmental mathematics course. *Research & Teaching in Developmental Education, 7*(2), 43-56.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*, 329-345.

Betz, N. E., & Hackett, G. (1993). *Mathematics self-efficacy scale*. Palo Alto, CA: Mind Garden Press.

Hackett, G. (1985). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20*(3), 261-73.

Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20*(3), 261-73.

Hall, J. M., & Ponton, M. K. (2005). Mathematics self-efficacy of college freshmen. *Journal of Developmental Education, 28*(3), 26-33.

Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology, 38*(4), 424-430.

Lent, R.W., Lopez, F.G., Brown, S. D., & Gore, P. A. (1996). Latent structure of the sources of mathematics self-efficacy. *Journal of Vocational Behavior, 49*, 292-308.

Lutzer, A. J., Maxwell, J. W., & Rodi, S. B. (2002). Statistical abstract of undergraduate programs in the mathematical sciences in the United States. Fall 2000 Conference Board of Mathematical Sciences (CBMS 2000). American

Margolis, H., & McCabe, P. P. (2006). Improving self-efficacy and motivation: What to do, what to say. *Intervention in School and Clinic, 41*(4), 218-227.

Mills, N., Pajares, F., Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning, 57*(3), 417-442.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*, 30-38.

O'Brien, V., Martinez-Pons, M., Kopala, M., (1999). Mathematics self-efficacy, ethnic identity, gender, and career interests related to mathematics and science. *Journal of Educational Research, 92*(4), 231-235.

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice*, *41*(2), 116-125.

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, *19*(2), 139.

Pajares, F., & Valiante, G. (2006). Self-efficacy beliefs and motivation in writing. In C. A. Macarther, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (p. 158-170). New York: Guilford Press.

Quinn, R. J. (1997). Effects of mathematical methods courses on the mathematical attitudes and content knowledge of preservice teachers. *The Journal of Educational Research*, *91*(2), 108-113.

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*, 551-554.

Scherer, M. (2002). Love and hate for math and science. *Educational Leadership*, *5*(61), 5.

Shields, D. J. (2006). *Causes of math anxiety: The student perspective.* Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Smittle, P. (2003). Principles for effective teaching in developmental education. *Journal of Developmental Education*,*26*(3), 10-16.

Stevens, T., Olivarez A., Jr., & Hamman, D. (2006). The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students. *Hispanic Journal of Behavior Sciences, 28*, 161-186.

Stevens, T., Olivarez A., Jr., Lan, W. Y., & Tallent-Runnels, M. (2004). Role of mathematics self-efficacy and motivation in mathematics performance across ethnicity. *Journal of Educational Research, 97*(4), 208-221.

University College. Illinois State University. COMPASS math placement exam: What is the COMPASS exam? http://universitycollege.illinoisstate.edu/testing/compass/, retrieved October 6, 2016.

University Facts. (2009, Fall). Planning and Institutional Research Office. Illinois State University, Normal, IL.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, *34*(1), 89-101.

Waycaster, P. (2001). Factors impacting success in community college developmental mathematics courses and subsequent courses. *Community College Journal of Research and Practice*, *25*, 403-416

Young, D. B., & Ley, K. (2002). Self-efficacy of developmental college students. *Journal of College Reading and Learning, 33*(1), 21-31.

Zeldin, A. L., Britner, S. L., Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching, 45*(9), 1036-1058.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal, 37*, 215-246.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82-91.

International Journal of Assessment Tools in Education: Vol. 4, Issue 1, (2017) pp. 54-78

http://www.ijate.net
e-ISSN: 2148-7456 © IJATE
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

"Research Article"

Formative Assessment in Teaching the Macedonian Language (Primary Education in R. Macedonia)

Violeta JANUŠEVA[1,*] Jana JURUKOVSKA[2]

Faculty of Education, St. "Vasko Karangelevski" - Bitola, Republic of Macedonia University "St Kliment Ohridski" - 7000 Bitola, Republic of Macedonia

Abstract

In the Republic of Macedonia, the formative assessment represents a new assessment paradigm that focuses on students as active participants during the teaching process. This paradigm was established in 2007 as part of the Primary Education Project (PEP) by USAID (United States Agency for International Development). This research is focused on the analysis of Macedonian language teaching curriculum from first to ninth grade of primary education in the Republic of Macedonia, on the analysis of the suitability of the tasks for formative assessment in one Macedonian language textbook, as well as on the examination of students' and teachers' viewpoints, concerning various aspects of formative assessment in Bitola, a region in the Republic of Macedonia. From the analysis it could be seen that there is strong support of formative assessment by the authorized structures, and that there are many positive and negative aspects of the use of formative assessment in the Macedonian language teaching process. The research has qualitative and quantitative paradigm. The methods for collecting and processing the data, and for gaining scientific conclusion are analysis, synthesis and comparison. The instrument consists of two questionnaires, one for students and one for teachers.

Article Info

Received
01 July 2016

Revised:
29 September 2016

Accepted
07 October 2016

Keywords:
Formative assessment; Teaching process; Macedonian language; Teachers' and students' viewpoint

1. INTRODUCTION

An important segment of the entire reforms in education is the improvement of the system of assessment regarding students' achievements. The assessment contributes to the improvement of the quality of education and its development to a great extent. The formative assessment in primary education in the Republic of Macedonia was implemented for the first time, in structural and systematic manner, in 2007 as part of the Primary Education Project (PEP) by USAID (United States Agency for International Development). That project comprised of 433 primary and secondary schools, and over 17 523 teachers, school and community representatives, (Almanah, 2011; Almanac, 2011). The project activities contributed significantly to the improvement of

[*]Corresponding Author E-mail: violetajanuseva@gmail.com jana_juruk@yahoo.com

2148-7456 /© 2017 DOI: 10.21449/ijate.266059

teachers' capacities in the Republic of Macedonia by encouraging them to implement qualitative manners for realization of formative assessment and to focus on the needs of the students.

As a continuous process of observation and evaluation of students' activities in the teaching process, the formative assessment provides information for student's achievements. This information serves as a basis for summative assessment, (Brookhart and Nitko, 2014; Poposki, 2005). The formative assessment is done throughout the entire teaching process and has preventive and corrective character. It motivates and directs the future performance of students and teachers. It has an influence on the flow of the teaching process and on the students' achievements. Formative assessment has a direct influence on the processes of teaching and learning, as well as on the entire students' development. It is directed towards modeling of the teaching process so that the established goals and standards could be achieved. During the formative assessment, teachers and students receive feedback for the quality and the effects of their activities. Teachers acquire information about the manner of teaching, and students about the process of learning, (Black and William, 2001; Black, 2007; Gojkov, 2003: 149). The formative assessment does not emphasize the result of the achievements, i.e. the grade, but it refers to the facilitation of the process of assimilating knowledge. As such, it is a manner which leads to achieving higher standards in learning, (Black and William, 1998; Wiggins, 2012).

The objective of the research is to provide information that concerns the use of formative assessment in the Macedonian language teaching process in Bitola, a region in the Republic of Macedonia. This information could be used as a certain indicator of the effects of its use on a national level. The analyses of the literature for formative assessment in the Republic of Macedonia, the Macedonian language teaching curriculum from first to ninth grade for primary education, the four tasks in the Curriculum (one task) and the Textbook (three tasks), as well as the viewpoints of Macedonian language teachers and students display strong institutional support of the concept of formative assessment. Additionally, the analysed materials demonstrate the contributions but also the difficulties in its realization in the Macedonian language teaching process. The analysis further demonstrates the estimation of teachers and students concerning the suitability of the tasks for realization of formative assessment (in the Curriculum and in the Textbook). Moreover, the analysis indicates the extent to which the tasks activate the higher cognitive processes of the students.

2. OVERVIEW OF THE LITERATURE FOR FORMATIVE ASSESSMENT IN THE REPUBLIC OF MACEDONIA

The characteristics and specificities of formative assessment in the Republic of Macedonia were initially analyzed by Poposki. He includes the formative assessment in the group of the summative and diagnostic assessment, according to the time period of its realization. He describes formative assessment as a continuous process that has formative function and preventive character, and he provides detailed analysis of the characteristics of formative assessment (continuity, differentiation, prevention, etc.). Further, he gives special attention to the feedback, emphasizing its importance in teaching and future activities of teachers, as well as in learning and progress of students, (Poposki, 2005: 170–180).

A great number of handbooks for training teachers for implementation of formative assessment in the teaching practice were written, as a result of the project activities from 2007. For example, there is the Handbook: "Vrednuvanje na znaenjata na postiganjata na ucenicite" [Assessment of the knowledge and achievements of students, 2007] written by a group of authors.

This Handbook significantly relies on the book "Učilišna dokimologija" [School dokimology, 2005] by Poposki. Further, in the Handbook "Unapreduvanje na ocenuvanjeto vo učilištata" [Enhancement of School Based Assessment, 2007] the authors emphasize the difference between the formative and summative assessment, the importance of Bloom's Taxonomy, and they provide essay questions, tests, and other kinds of questions intended for realization of formative assessment, (Schiel and Kitching, 2007). In addition, there is the Handbook "Ocenuvanje so testovi na znaenje" [Assessment with knowledge tests, 2007]. The author of this Handbook, Violeta Petroska-Beška, emphasizes the manner of creation and use of the essay questions, their advantages and disadvantages.

Further, in 2011 an International conference was held: "Ocenuvanje za ucenje vo 21 vek" [Assessment for learning in the 21 century], as a result of the project activities. On the Conference a great number of foreign experts, teachers in primary, secondary and high education, and counselors from the Bureau for Development of Education in Republic of Macedonia, shared their viewpoints regarding the effects of the use of formative assessment in the teaching practice. Their viewpoints, i.e. papers were published in a Proceeding of the Conference. The analysis of the papers in the Proceeding demonstrates that the most common elaborated subjects are: the role of formative assessment in the teaching practice of teachers and the learning process of students; the characteristics of formative assessment; the different segments of formative assessment; the connection between formative assessment and teaching goals, standards and criteria for success; the character of the feedback, (Zbornik, "Ocenuvanje za ucenje vo 21 vek" [Proceeding, Assessment for learning in the 21 century], 2011).

Moreover, the number of research studies in Macedonia, regarding the effects of the implementation of formative assessment in the teaching practice has increased as a result of the project activities. These studies indicate that teachers are getting familiar with the essential characteristics of formative assessment. Formative assessment is recognized by teachers as a developmental assessment that helps students' progress. Teachers also believe that this assessment provides feedback for them and their students. These facts are indisputable indicators that there is a solid basis for implementation of formative assessment. They also demonstrate that the effects of its implementation provide quality of the teaching process, (Author and Pejchinovska, 2011; Talevski et al., 2014). Additionally, research studies, (Author and Pejchinovska, 2011; Pejchinovska and Author, 2014) indicate that almost all teachers plan the formative assessment in the yearly and thematic lesson plans. In these documents they note the techniques for assessment of students' achievements. Teachers also prepare daily lesson plans where they provide more detailed description of the techniques and instruments for assessment of students' achievements.

According to the results of the research studies, (Pejchinovska and Author, 2014; Talevski et al., 2014) it could be concluded that teachers pay attention to the connection between the formative assessment and the teaching goals. This is a significant indicator which demonstrates that teachers operationalize the teaching goals that are given in the teaching plans for all students. They also differentiate and individualize the approach in the realization of the tasks. Thus, in the range of the global teaching goals, teachers set more specific goals according to the needs and the abilities of each student. The individualization and differentiation of tasks as planed and systematic activity, indisputably speaks of the teachers' competencies for implementation of formative assessment. The connection between the formative assessment and the teaching goals is a necessary prerequisite for effective and efficient implementation of formative assessment. Additionally, this is a condition for providing solid bases for promotion of the quality of teaching and learning, i.e. the quality of the teaching process in general.

3. METHODOLOGY OF THE RESEARCH

This research has qualitative and quantitative paradigm. The methods for collecting and processing the data, and for gaining scientific conclusion are analysis, synthesis and comparison. Data collection is done with questionnaires for students and teachers, prepared by the authors of this paper.

An analysis of the Macedonian language teaching curriculum from first to ninth grade for the nine-year primary education in Republic of Macedonia is done, in regard of the institutional support of the concept of formative assessment. This analysis is done in accordance with the directions for realization of formative assessment given by the Ministry of Education and Science and the Bureau for Development of Education. The Teaching Curriculum is retrieved from the official website of the Bureau for Development of Education. Then, there are two separate analyses of questionnaires given to Macedonian language teachers and students. These questionnaires examine their viewpoint regarding the following: the contributions of formative assessment but also the difficulties in the daily Macedonian language teaching process; the extent to which the tasks in the Macedonian language textbooks are suitable for formative assessment in general; and especially, the extent to which one task given in the Teaching curriculum for sixth grade, and three tasks given in the Macedonian Language Textbook for sixth grade are suitable for formative assessment, i.e. for improvement of students' achievements and also for activation of their higher cognitive processes. The Textbook is retrieved from the official website for digital textbooks of the Ministry of Education and Science. Both documents (Curriculum and Textbook) are currently used in the teaching process in sixth grade of the nine-year primary education. The analysis, in fact, interprets the results from the questionnaires and makes a comparison between them. In the questionnaires, among the questions, four tasks have been used. These tasks are analysed regarding the Bloom's Taxonomy, i.e. the extent to which they activate the higher cognitive processes of the students. The four tasks from the Teaching Curriculum and the Textbook refer to the section "Language" for the teaching unit "Grammatical category mood of verbs". The tasks are:

1. Task from the Macedonian Language Teaching Curriculum for Sixth Grade, (Nastavna programa po makedonski jazik VI oddelenie, Macedonian Language Teaching Curriculum for sixth grade, 2008: 7): After an analysis of a text from a textbook, students have the task to underline all verbs in a given paragraph. Then, they should write the underlined verbs in their notebooks and determine the verb form regarding the person, number, aspect and mood.

2. Three tasks from the Macedonian Language Textbook for Sixth Grade, (Velkova, Jovanovska, 2011: 20): 2.1. Apply your knowledge – Write three sentences for each mood of the verbs; 2.2. Recognize the mood of the verbs in the following sentences: Toj vežba redovno (He exercises regularly); Goce, pišuvaj pobrzo (Goce, write faster); Jana bi peela postojano (Jana would sing constantly); Donesete voda, ve molam (Bring water, please); Jas bi jadela sladolen (I would eat ice-cream); Kalina redovno vežba (Calina regularly exercises); 2.3. Transform the verb form in the sentences, from indicative mood into imperative and subjunctive mood. (Example: Baba zboruva., Grandmother talks.; Babo, zboruvaj!, Grandmother, talk!; Baba bi zboruvala., Grandmother would talk.) – 1. Tato odi brzo (Father goes fast); 2. Tanja gleda film. (Tanja watches a movie); 3. Ribarot fati riba (The fisherman caught a fish). The research includes 30 Macedonian language teachers who teach from fifth to ninth grade of the nine-year primary education, and 90 students who currently attended sixth grade, in the following primary schools in the municipality of Bitola and Novaci: "Kliment Ohridski", "Todor Angelevski", "Kiril i Metodij", "Elpida

Karamandi", "Goce Delcev", "Dame Gruev" (Bitola); "Slavko Lumbarkovski" – Bač (Novaci); "Brakja Miladinovci" – Dedebalci (Mogila).

4. RESULTS, ANALYSIS AND DISCUSSION

4.1. Formative Assessment in the Macedonian Language Teaching Curriculum for Primary Education

In the Macedonian language teaching curriculum from first to ninth grade for the nine-year primary education there is a special section referring to assessment of students' achievements. This section provides directions for the implementation of formative assessment.

In the Macedonian language teaching curriculum for first, second and third grade there is a recommendation for implementation of formative assessment during the teaching process. This includes creation of portfolio for the students that contains: collection of indicators such as students' creations, products, statements, etc. for each student individually; current (formative) evaluation lists prepared in advance for each student that are filled in after every particular activity of the student; case study in which the teacher notes the actual state; instruments that refer to every particular teaching section (the teacher writes information about students' achievements) and provide information of all aspects that are stimulated by the Macedonian language curriculum (the intellectual, socio-emotional and psychomotor aspect). It is also stated that the evaluation lists refer to the teaching goals that are completely achieved by the entire class. Further, it is stated that micro-summative assessment is provided at the end of the second trimester, on the basis of the information collected with the formative assessment, (Macedonian Language Teaching Curriculum for First grade, 2007: 22, 23; Second grade, 2007: 15,16; Third grade, 2007: 13, 14).

In the Macedonian language teaching curriculum from fourth to ninth grade there are more information regarding the formative assessment. Firstly, there is a recommendation for implementation of formative assessment during the teaching process. Further, it is said that teachers could realize the formative observation and assessment of students' achievements by using various methods and procedures. Some of them are the following: conversations teacher – student, oral presentations, written exercises, tests, homework and portfolio – collection of indicators for the achievements of each student individually. According to the given directions, the portfolio should be accessible to the parents throughout the entire school year. In that way they could also contribute to the quality of the realization of the Macedonian language teaching process. In addition, it is stated that the portfolio of each student would continue to be used for the following grade with purified and selected information. Further, it is stated that the results of the observation of students' achievements provide creation of planed oral and written feedbacks. These feedbacks (in form of descriptions) are intended for the students, the parents and the teachers. Finally, it is indicated that micro-summative assessment is provided at the end of every trimester on the basis of the information collected with the formative assessment. (Macedonian Language Teaching Curriculum for Fourth grade, 2009: 19; Fifth grade, 2008: 17; Sixth grade, 2008: 18; Seventh grade, 2008: 16, 17; Eighth grade, 2008: 15, 16; Ninth grade, 2009: 17, 18).

The Macedonian language teaching curriculum from sixth to ninth grade indicate that all of the methods for formative assessment could be evaluated by the teacher with a help of evaluation lists. This list should be prepared in advance for each student and it should be filled in after every particular activity of the student. (Macedonian Language Teaching Curriculum for Sixth grade, 2008: 18; Seventh grade, 2008: 16, 17; Eighth grade, 2008: 15, 16; Ninth grade, 2009: 17, 18).

From the analysis of the Macedonian language teaching curriculum it could be concluded that the Ministry of Education and Science and the Bureau for Development of Education provide strong institutional support of the concept of formative assessment. It is evident that there are directions and recommendations for realization of formative assessment. Those directions and recommendations could significantly contribute to the improvement of the quality of learning by the students and the education in general.

4.2 Analysis of the Results of the Questionnaire for Macedonian Language Teachers

The following are the results of the questionnaire for teachers, along with the analysis and the discussion. The result regarding the work experience of the teachers is presented in Figure 1. Namely, 47% of the teachers (14 respondents) have work experience of more than ten years, 30% of them (9 respondents) have work experience of five to ten years, and 23% (7 respondents) have work experience of one to five years. There are not teachers with work experience of less than a year.

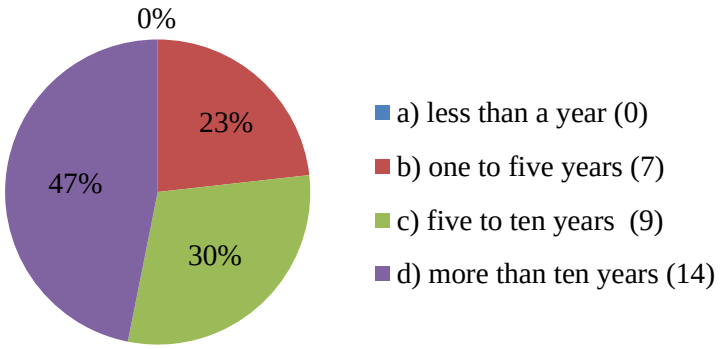


Figure 1. Question no.1. What's the length of your work experience?

The result of the first question is directly connected with the result of the second question. Figure 2 presented the results of the second question which refers to the participation of teachers in projects, seminars or trainings for formative assessment. Namely, 13 respondents (43%) have participated in these forms of professional development only once, the biggest part of the teachers, that is, 16 respondents (54%) have participated many times, and only one respondent has not participated in project, seminar of training of this kind. These results provide an illustration of the real situation regarding the participation of teachers in these forms of professional development. It is a fact that, after the implementation of formative assessment in the Republic of Macedonia, these forms of professional development about formative assessment are held continually. Nowadays, there is not a teacher who has not participated in one of these forms of professional development, organized mostly by the Ministry of Education and Science and the Bureau for Development of Education.

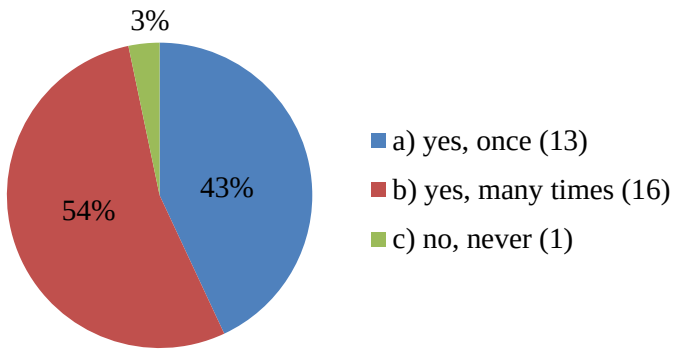


Figure 2. Question no.2. Have you even participated in projects, seminars or trainings for formative assessment?

Further, the result, presented in Figure 3, demonstrates that 83% of the teachers (25 respondents) correctly associate the formative assessment with the entire development of the student's personality, i.e. the cognitive, affective and psychomotor domain, and 5 respondents (17%) associate the formative assessment only with the cognitive domain. This result is a clear indicator of the developed competencies of most teachers regarding the formative assessment. However, this result also indicates the need for their further professional development, so that the knowledge of all teachers could be improved.

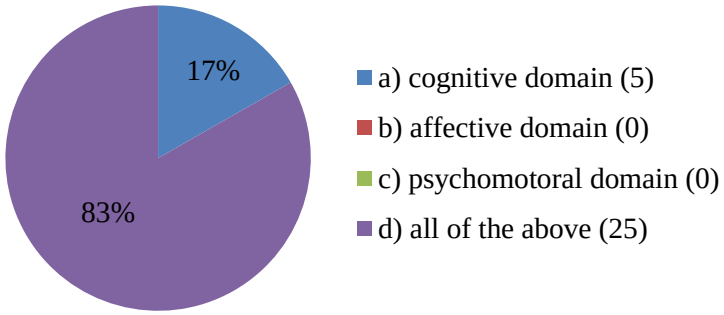


Figure 3. Question no.3. What does the term "formative assessment" refer to?

Moreover, from the result presented in Figure 4, it is evident that 27 respondents (90%) comprise the continuous observation and evaluation of students' activities in the formative assessment, and 20 respondents (67%) additionally comprise the process of providing constructive feedback. Only 12 respondents (40%) state that the effective formative assessment further comprises clearly defined goals in the teaching process. The teachers had the possibility to choose more than one answer, but it is evident that most of them consider that the formative assessment only includes the continuous observation and evaluation of students' activities. A small number of them additionally included the effective feedback and the goals of the teaching process. This information indicates the need for further development of teachers' competencies regarding the characteristics and specificities of the formative assessment. The constructive, constant, well-timed, individualized and, above all, corrective feedback, and the goals of the teaching process are closely related to the formative assessment, (Black and William, 1998; Brookhart, 2008; Wiggins, 2012).

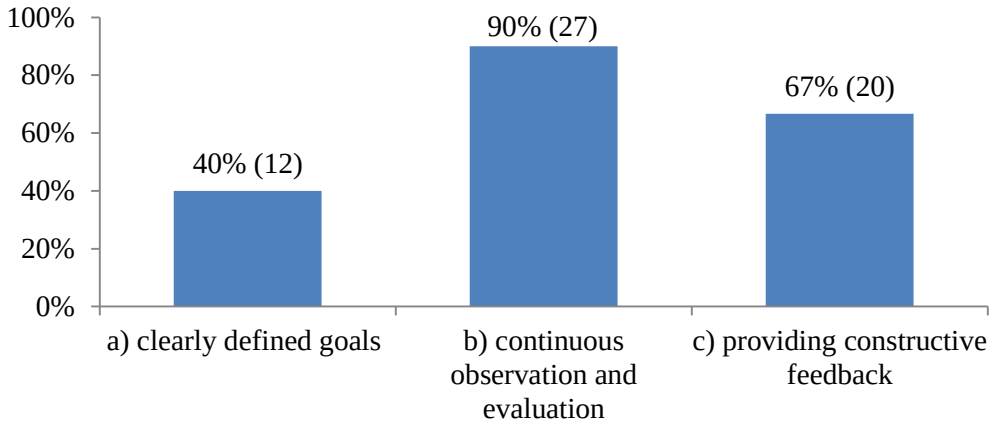


Figure 4. Question no.4. What does the formative assessment comprise?

The following result, presented in Figure 5, refers to the stability of formative assessment and its characteristic in the daily teaching practice of teachers. 87% (26 respondents) state that they constantly use formative assessment in their daily teaching practice, and 13% (4 respondents) use it occasionally. This result indicates the need for further research of the causes that evoke some teachers to employ formative assessment occasionally, and not constantly. However, it is also very important to emphasize that the number of those teachers is significantly smaller, and that there are not teachers who do not use formative assessment at all. Moreover, this information demonstrates that teachers employ formative assessment in their daily teaching practice and they understand the importance of formative assessment in the teaching process. A great number of studies connect the improvement of students' achievements with the use of formative assessment, (Black, 2007; Gojkov, 2003).

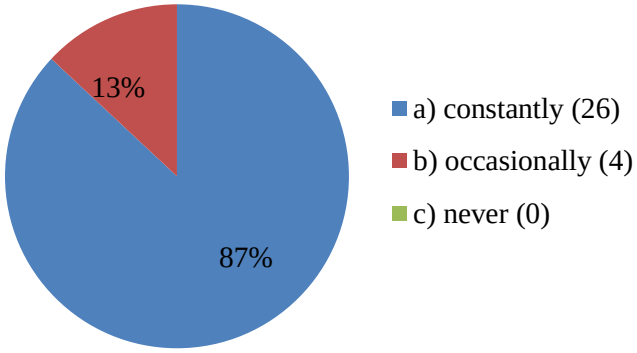


Figure 5. Question no.5. How often do you employ formative assessment in the classroom?

The above stated information could be further confirmed with the result presented in Figure 6. Namely, 87% of the teachers (26 respondents) reckon that formative assessment improves students' achievements to a great extent. Only 13% of them (4 respondents) respond that it has only a slight influence on the improvement. Again, this result indicates the need for further research of the causes that evoke some teachers to reckon that there is only a slight improvement in the students' achievements. The causes may have different character. However, it is very significant to emphasize that there are not teachers who reckon that this kind of assessment does not improve the achievements at all. This is an indisputable indicator of the direct connection of the improved students' achievements with the implementation of the formative assessment.

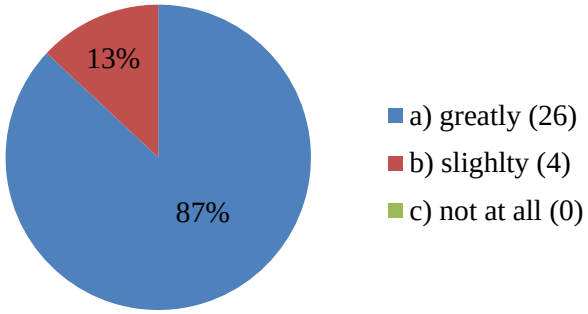


Figure 6. Question no.6. How much does the formative assessment improve student's achievements?

Further, a great number of studies also indicate the importance of incorporating the formative assessment in the yearly and developmental program of schools' work. Consequently, it is very important to incorporate the formative assessment in the yearly, thematic and daily lesson plans, (Black and William, 1998; Black, 2007; Standardi za ocenuvanje na učenicite vo osnovnoto obrazovanie, Nacrt, 2008 [Standards for assessment of students in primary education, Draft, 2008]; Wiggins, 2012). In this sense, the result, presented in Figure 7, confirms the importance of planning the formative assessment in the daily teaching practice. The teachers had the possibility to choose more than one answer. Consequently, 24 respondents (80%) state that they plan and integrate the formative assessment it the thematic and daily lesson plans, 18 respondents (60%) plan the formative assessment in the yearly lesson plans as well. This information depicts the developed competencies of teachers and their understanding that the planning of formative assessment is a very important segment of the teaching practice.

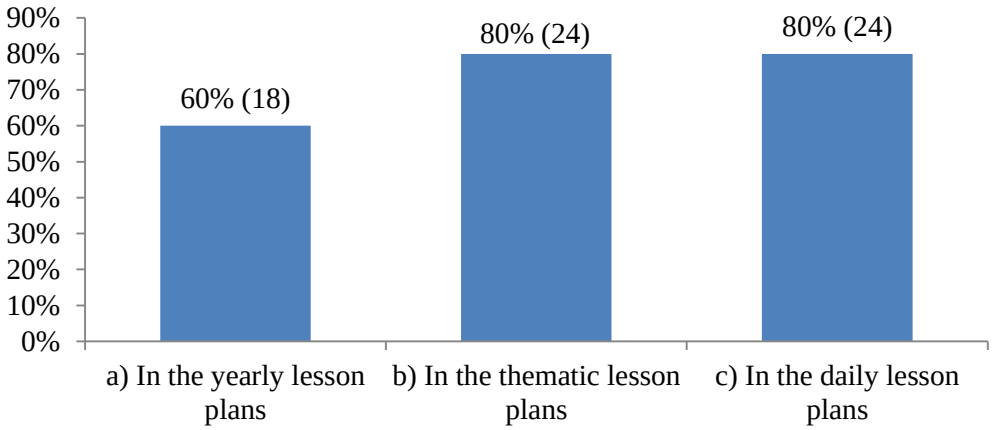


Figure 7. Question no.7. Where do you plan and integrate the formative assessment?

Formative assessment implies the use of various techniques and instruments for assessment of students' achievements. Namely, the use of different techniques provides more valid information about the achievement of each student. Consequently, the teacher could be more confident that the assigned grade is a reflection of the real achievement of the student, (Standardi za ocenuvanje na učenicite vo osnovnoto obrazovanie, Nacrt, 2008 [Standards for assessment of students in primary education, Draft, 2008]). Thus, the result, presented in Figure 8, demonstrates that the three most frequently used techniques and instruments for formative assessment are: check

list 77% (23 respondents); conversations teacher – student 67% (20 respondents); and written exercises 47% (14 respondents). Regarding the other techniques and instruments, the result indicates that equal number of respondents, that is 11, use oral presentation and test; 3 of them use homework as a technique for formative assessment; and 2 of them choose portfolio, self-assessment, mutual assessment and project for the mentioned purpose. None of the respondents uses quizzes, although other studies confirm that the quiz is powerful and creative tool for formative assessment. Consequently, it is evident that teachers use various techniques and instruments for implementing the formative assessment in the teaching practice. This provides more valid grades and allows the teacher to be more objective when assessing students' achievements.

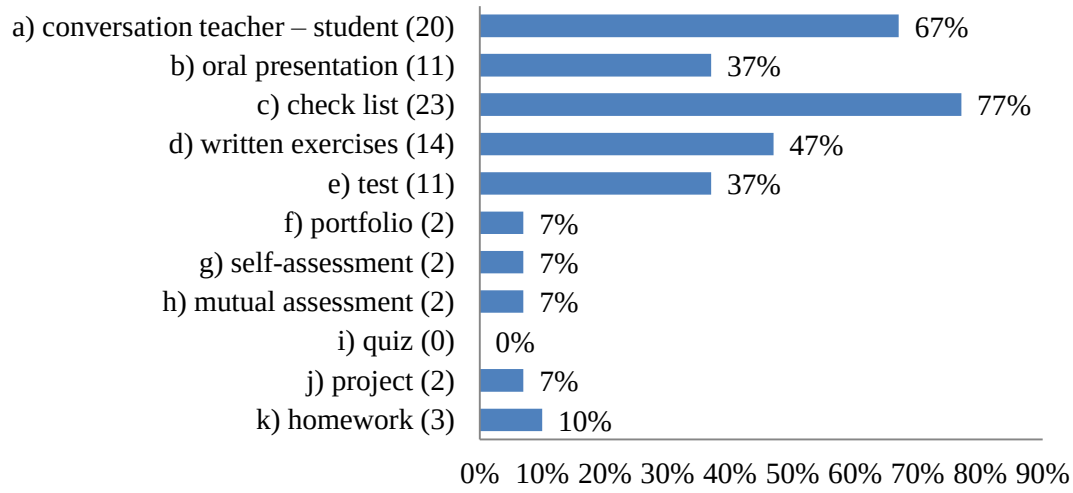


Figure 8. Question no.8. Which three of the following forms for formative assessment do you use most frequently?

It could be said that the techniques and instruments which are presented on workshops for formative assessment are not definite. Thus, it is very important for the teacher to use its creativity and to be in a continuous search of new techniques and instruments. The daily use of the same techniques and instruments could lead to monotony in the teaching process. The surprisingly small number of teachers who use the portfolio for formative assessment should be mentioned in this research. A great deal of attention is given to the portfolio in the analyzed teaching curriculum. That comes as a consequence to the fact that the portfolio is very convenient for students at a younger age. The result concerning the techniques and instruments also depicts the small number of teachers who use self-assessment, mutual assessment and project. Other studies identify them as powerful tools for formative assessment because of the following benefits: each student has a chance to solve authentic tasks related to those in the real world; they help the student to develop the potential for creative and critical thinking; they enable the student to work in team; they enable the student to evaluate his/her achievements independently by creating and implementing personal criteria, (Andrade and Valtcheva, 2009: 13; Black and William, 1998; Chappuis and Stiggins, 2002: 41; McMillan and Hearn 2008: 40–41). Some studies confirm that, in the teaching practice, teachers rarely use activities for self-assessment and mutual assessment. Their presence is more declarative than real because teachers reckon that students do not have developed competencies for these kinds of activities. Teachers believe that students should not interfere into their work as teachers, etc., without considering the fact that, with the participation in the activities for self-

assessment and mutual assessment, the assessment becomes shared partnership activity, and not only their own responsibility, (Talevski et al., 2014).

Further, the new paradigm in the teaching process and assessment indicates a new conceptual methodology for writing textbooks. The content of the textbooks should be suitable and should include the formative assessment. The exercises and tasks should be in correlation with the manner of teaching and assessment and should reflect the higher cognitive processes of thinking. Teachers often have an automatic approach regarding the use of the tasks in the textbooks. They do not take into consideration the cognitive processes which are developed by the task. For instance, regarding the task given in the directions in the Macedonian Language Teaching Curriculum for Sixth Grade, it could not be said that it requires application of verbs. Students should only underline the verbs in a given paragraph of a text and determine the verb form, regarding the person, number, aspect and mood. They should only recognize the indicative, imperative and subjunctive mood without defining them. This is a clear indicator that the task refers to the first two cognitive levels of the revised Bloom's Taxonomy – remembering and understanding. Regarding the applying of knowledge, the tasks given in the Macedonian Language Textbook of Sixth Grade are also questionable. It might be said that only the first task refers to the third level of the Taxonomy, because it requires the student to write sentences for each mood of the verbs. The second task requires identification of the mood of the verbs in the given sentences. The third requires transformation of the verb form from indicative mood into imperative and subjunctive mood. The dilemma refers to the following: To what extent does the student apply the knowledge if he/she is required to recognize the mood of the verbs in the given sentences and to transform the verb form from indicative mood into imperative and subjunctive mood on examples that are already given. The following result presented in Figure 9 refers to the frequency of use of the tasks in the Macedonian language textbooks in general. 70% of the teachers (21 respondents) use them occasionally, 20% (6 respondents) use them constantly, and the remaining 10% (3 respondents) never use them for formative assessment. Certainly, it is easier for teachers to use the already given tasks in the textbooks. They are released from the additional work of creating their own more creative tasks which would be more suitable for development of the higher cognitive processes of students. They are also released from additional expense for paper and photocopies.

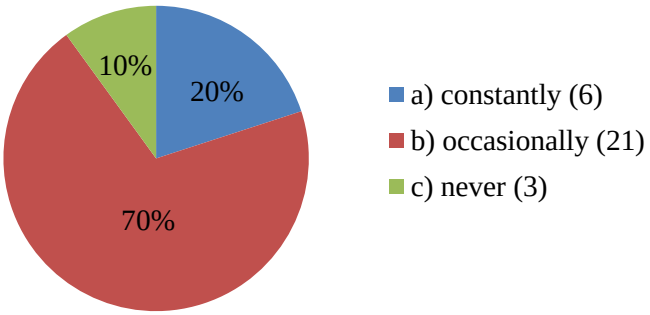


Figure 9. Question no.9. How often do you use the tasks in the textbooks for the purpose of performing formative assessment?

However, as Figure 10 demonstrates, 70% of the teachers (21 respondents) do not consider the tasks to be more practical nor more economical for use in comparison with the tasks that the teacher would create himself/herself. 30% of them (respondents) choose the opposite answer. They find the tasks in the textbooks more practical and more economical for use.

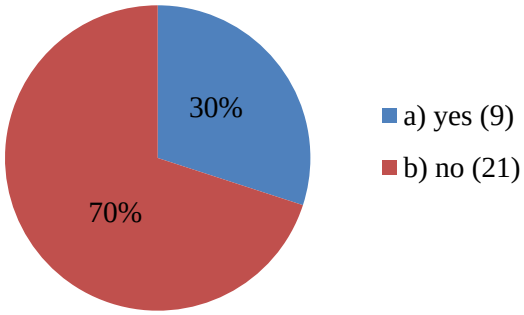


Figure 10. Question no.10. Do you find the use of the tasks in the textbooks more practical and more economical than the use of tasks that you create on your own?

A great number of teachers do not even consider the cognitive processes which the tasks in the textbooks develop. They believe that if the task is given in the textbook, then it must be in accordance with the teaching curriculum and the teaching goals. They also believe that the connection between the teaching goals in the curriculum and the content and tasks in the textbook is always taken into consideration by the authors of the textbooks. When considering the cognitive domain, the difficulty level of the tasks in the textbooks must depend on the age of the students and the teaching goals. Though there are differences, it is obvious that the teaching process, in a great part of its realization, is directed towards the lower (mostly the first two) levels of the Bloom's Taxonomy. The following result of the research, presented in Figure 1, confirms this statement. Namely, 93% (28 respondents) reckon that the tasks in the textbooks mostly refer to the first level – remembering, and about 90% (27 respondents) choose the second level – understanding. Further, the results demonstrate that a relatively small number of respondents, that is, 47% (14 respondents) choose the third level – applying, 37% (11 respondents) reckon that the tasks refer to the fourth level – analyzing, and 23% (7 respondents) choose the fifth level – evaluating. Only 10% (3 respondents) reckon that the tasks refer to the highest, sixth level, of the revised Bloom's Taxonomy which is creating. Regarding the three tasks from the Macedonian Language Textbook for Sixth Grade, as it has previously been stated, the first certainly refers to the third level – applying, but it is questionable whether the second and the third refer to the same level as well. It is evident that they do not refer to the levels analyzing, evaluating and creating in any case.

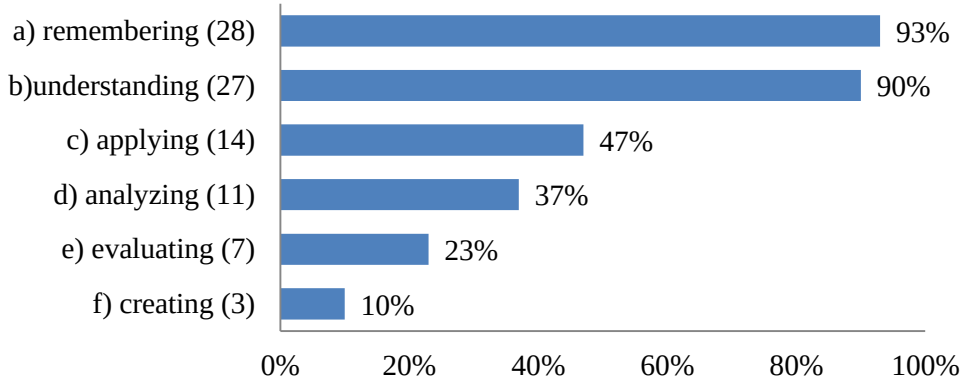


Figure 11. Question no.11. Which three levels of the revised Bloom's Taxonomy are the most prevalent in the tasks (for formative assessment) given in the textbooks?

This result indicates the need for further improvement of the teachers' competencies regarding the knowledge about the levels of the Bloom's Taxonomy. Also, there is a need for more in-depth consideration of the level applying. In the literature, this level is described in different manners. This gives to teachers' space for different interpretation according to their own understanding of the components that are included in the level applying. For instance, in the explanation about the Bloom's Taxonomy, (Clark, 2014; Huitt, 2011) concerning the level applying, it is said that this level refers to memorizing and applying principles, ideas or theories. The student uses the concept in a new situation, i.e. he/she applies the learned information from the classroom in new situations at the work place. This explanation could be interpreted in two manners, regarding the phrase *new situation*: the first one indicates that the student applies the concept in a new situation in the process of learning, that is, in other situation in the classroom; the second indicates that this level refers to applying the learned information from the classroom in new situations at the work place, that is, in a situation outside the classroom. In the Handbook "Unapreduvanje na ocenuvanjeto vo učilištata" [Enhancement of School Based Assessment, 2007: 64], regarding the level applying, the following example is given: Connect the manner of living of the characters in the novel with the manner of living in the reality. This indicates that the level applying is present when students have to connect the learned information from the classroom with authentic life situations. On the other hand, in the revised Bloom's Taxonomy, (Krathwohl, 2002) it is stated that the level applying is present when students have to use the information in other familiar context. Thus, it could be concluded that the context should be familiar, but the phrases *familiar context* and *given situation* might be interpreted differently as well. An illustration could be made with another example. For instance, we would assume that the student has assimilated the rules for correct use of the accent in the Macedonian standard language in a class at school. What if, later in class, the student is given several new words which have not been mentioned earlier in the exercises, and he/she is asked to determine the place of the accent in them. It is evident that the rule is not applied at the work place, instead it is only applied in the classroom and the context is familiar. Consequently, it is questionable what would be regarded as a new situation. Would the previously mentioned new words be regarded as a new situation? Or a new situation would be correct accenting of words in everyday life which would mean correct use of words regarding the place of the accent while speaking. On the other hand, if the student in class is given, for example, four words with only one of them correctly accented, (the other three words have an accent on incorrect syllable) again, we have double interpretation. It is questionable whether this task refers to the level applying or the level understanding, and whether this is a new situation or it is not,

although it is evident that the context is familiar. The final examination for the mandatory subject Macedonian language and literature, realized at the end of the fourth year of secondary education, imposes different demands on students. What if students are given four new words (one with an accent on the correct syllable and three with an accent on incorrect syllable) and they are asked to choose the word with an accent on the correct syllable? Does this mean that the final examination represents a new situation, an unfamiliar context or the context is familiar? Does the student apply the knowledge, or does he/she know the rule and understands which of them are the correct answer? This paper does not deal with more extensive analysis of the levels of the Bloom's Taxonomy. However, it suggests that there is a need for more precise explanation of the third level – applying, as well as of the other three levels (analyzing, evaluating and creating). In that way, the Bloom's Taxonomy could be used in a more beneficial manner in the teaching practice, i.e. in the process of writing teaching goals and for creation of tasks that would reflect the higher cognitive processes of students.

The following result from the research refers to the previously explained four tasks from the Teaching Curriculum (one task) and the Textbook (three tasks). Namely, 80% of the teachers (24 respondents) state that they would use these tasks. The 24 respondents who gave positive answer, had the opportunity to explain their reasons. Namely, 40% of them (12 respondents) state that the tasks are suitable for estimating whether the required knowledge is assimilated, 37% of them (11 respondents) reckon that they actively involve all students, and 23% of them (7 respondents) reckon that they are suitable for the individual needs of the students. The other 20% (6 respondents) gave negative answer, i.e. they stated that they would not use these four tasks. With regard of the reasons, the result demonstrates that 10% of them (3 respondents) reckon that the tasks are not suitable for estimating whether the required knowledge is assimilated, 7% (2 respondents) state that they do not actively involve all students, and 1 respondent reckons that the tasks do not reflect the individual needs of the students. Those 6 respondents who gave negative answer regarding the use of the four tasks, had an extra space to write what they would use instead. One respondent states that he/she would use tasks for group, individual and individualized work with different level of difficulty suitable for the individual needs of the students. Other respondent states that, apart from the four tasks, there is a necessity for additional assessment of the knowledge with other tasks. Third respondent states that he/she would use tasks which would enable the students to assimilate the material more easily. The fourth respondent states that he/she would use tasks for development of the higher cognitive processes of the students. This information provides evidence of the critical approach of the teachers in regard of the tasks from the Teaching Curriculum and the Textbook. In addition, it is also an indicator of their creativity and the readiness to create their own tasks which would be more suitable for formative assessment and which would enable the students to develop the higher cognitive processes.

4.3 Analysis of the Results of the Questionnaire for Students

The following are the results of the questionnaire for students, along with the analysis and the discussion. The first result, presented in Figure 12, indicates that 84% of the students (76 respondents) learn better when their knowledge is evaluated on a shorter period of time and for a shorter material with different exercises and tasks. On the other hand, 16% of them (14 respondents) learn better when their knowledge is evaluated with half-year tests. They prefer to be tested on an extensive material with one method of assessment. This result demonstrates the importance of formative assessment in the teaching practice and its connection with students' achievements, (Black, 2007; Gojkov, 2003). However, it also implies that students still use the

practice of studying only for tests, i.e. only to get the grade that they want. Therefore, this result suggests the need for change in the manner of studying of the students. That could be achieved by using continuous formative assessment that would significantly contribute to the improvement of their achievements.

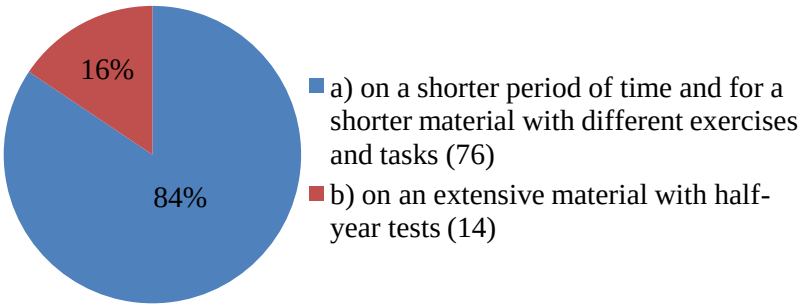


Figure 12. Question no.1. I learn better when my knowledge is evaluated...

The motivation is a key factor related to students' achievements. This means that the amount of development of the internal students' motivation is proportional with the quality of their achievements. The result, presented in Figure 13, confirms the connection between motivation, continuous feedback and numerical grade. 68% of the students (61 respondents) state that they are more motivated for studying, learning and participating in the school activities when they have continuous feedback. This leads to the conclusion that the continuous feedback provided by the teacher also has positive influence on students' achievements. This means that the qualitative, corrective, clear, understandable, indisputable and continuous feedback should be an imperative for every teacher, (Black and William, 1998; Brookhart, 2008; Wiggins, 2012). The remaining 32% of the students (29 respondents) state that they are more motivated by a numerical grade, which means that they prefer numerical grade rather than continuous feedback. This opinion requires further analysis and research regarding the reasons for this kind of motivation. Possibly, the reasons could be found in the present practice of conducting the assessment on a specially selected class for that purpose which is completed by assigning numeral grade. The grade is considered a final result of students' achievements. This practice could be surpassed with a continuous use of formative assessment and in that manner students would be more aware of the long-term effects of their studying.

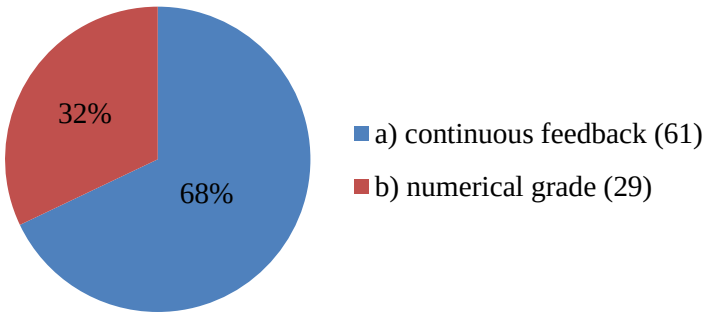


Figure 13. Question no.2. I am more motivated by...

Further, it has already been stated that the teaching goals are closely connected with the daily teaching practice. It is very important for the students to understand why they should assimilate certain material and for what that material would serve them in their future life. Regarding the teaching goals, Figure 14 demonstrates that 93% of the students (84 respondents) state that their Macedonian language teachers inform them about the goals of the lessons on each class. Only 7% (6 respondents) state the opposite, i.e. that the teacher does not inform them about the teaching goals. It is more than evident that the act of sharing the teaching goals with the students presents a positive practice because it is a prerequisite for improvement of students' achievements. The teaching goals along with their concretization and operationalization make a valuable contribution to the teaching process, because the teaching practice of the teachers and the learning of the students depend on them, (Author and Pejchinovska, 2011). However, it should also be stated that there have been and there would always be students who might not articulate the speech of the teacher in a correct manner. The negative practice of not sharing the goals with the students still exists apart from the already established positive effects. Thus, this possibility suggests the need for serious changes in the daily teaching practice of some teachers. Moreover, Figure 14 also demonstrates that 94% of the students (85 respondents) reckon that their Macedonian language teachers observe and note their answers, and only 5 of them, that is, 6% do not agree with it. This result suggests that students are aware that their activities are observed and noted. In other words, it contributes to the awareness of the students that the observed and noted activities would be used as elements for summative assessment of their achievements. Further, the result, also presented in Figure 14, confirms that feedback has very important significance in the teaching process. Namely, 88% of the students (79 respondents) state that they receive continuous feedback from their Macedonian language teacher. Thus, this result is also correlated with the results of the questionnaire for teachers, (see Figure 5, Figure 6). On one hand, it demonstrates the awareness of the teachers that the feedback has an influence on the students' achievements, and that it is very important to provide feedback in the teaching practice. On the other hand, it shows the readiness of the students to recognize the reactions of the teacher for different activities (verbal and non-verbal) as feedback. They should use them in a corrective sense, that is, for improvement of the process of learning. The remaining 12% of the students (11 students) state that they do not receive continuous feedback from their teacher. This might be understood as an indicator for the following aspects: the teacher does not provide continuous feedback; the student is not able to understand or correctly interpret certain reactions of the teacher, and so on. In either case, it is very important to emphasize the importance of corrective and constructive feedback in the teaching process and also to depict the need of its continuity. If the teacher provides feedback occasionally and not constantly, then there is a great possibility that students would not be able to recognize the teacher's reactions as feedback. This would have negative influence on the process of learning and also on the achievements. Further, the variety of techniques and instruments for formative assessment is a necessary imperative in the process of collecting information for students' achievements. Therefore, the result, presented in Figure 14, indicates that 89% of the students (80 respondents) state that their Macedonian language teachers use various techniques and instruments for formative assessment. According to them, they are in accordance with their cognitive individual needs. Only 11% of them (11 respondents) do not agree with the others. Moreover, the participation of students in the assessment is a very important aspect of formative assessment. The assessment becomes mutual activity of the teacher and the student and in that process they have a status of partners. Regarding this aspect of formative assessment, Figure 14 also demonstrates that 71% of the students (64 respondents) state that they are involved in the process of assessment. 29% of them (26 respondents) do not agree with it, because they state that they do not feel involved

in the process of assessment. This result is an illustration of the real situation in the teaching process, because informal conversations with teachers and students also suggest that a great number of teachers only formally include the students in the assessment. This means that they do not really take into consideration the viewpoint of the students. They reckon that the competencies of the students are not developed enough so that they could make realistic assessment of their achievements. It is necessary to indicate that there is a need for greater efforts by the teacher for involvement of the students in the assessment. The teacher could only accomplish this if he is really willing to take into consideration the viewpoint of the students. Only then it could be said that students seriously participate in the activities of this kind. Only then, the teacher would notice the real contributions of formative assessment in the teaching process: independent students who think critically and reconsider their achievements; students who evaluate their achievements properly; students who share the assessment with the teacher so that it becomes a partnership activity; and students who would develop into citizens who take an active responsibility for their community, (Schiel and Kitching, 2007).

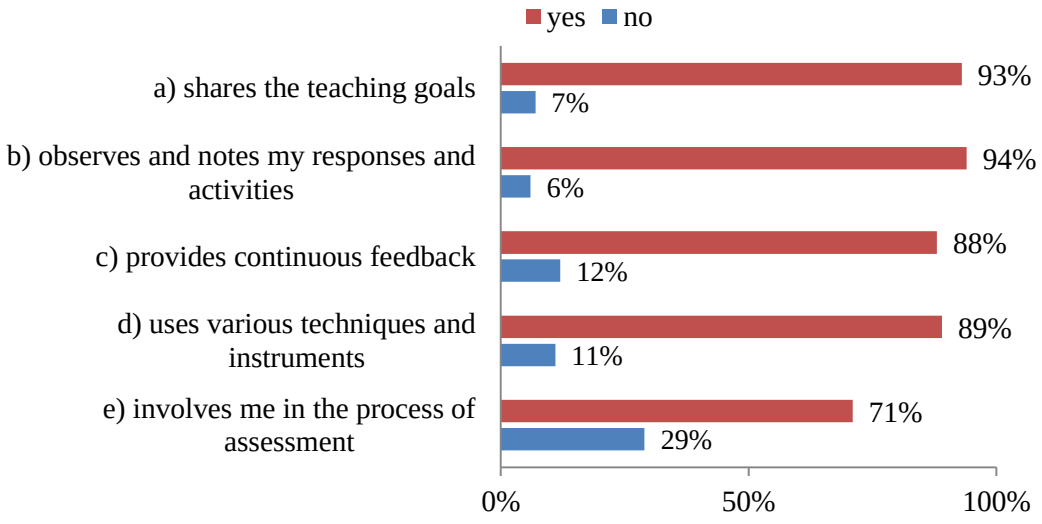


Figure 14. Question no.3. The teacher...

Regarding the techniques and instruments for formative assessment, the result, presented in Figure 15, demonstrates that the viewpoint of the students differs from that of the teachers, (see Figure 8). When asked about the three techniques that help them learn more easily and improve their knowledge, 61% of the students (55 respondents) choose conversation teacher – student, a technique which is ranked second by the teachers. The oral presentation and homework are chosen by 46% of the students (41 respondents) and they are ranked second. In third position is the test with 33% (30 respondents). The result of the questionnaire for teachers demonstrates that these techniques are not among the three most frequently used by the teachers. In fourth position with 28% (25 respondents) are the written exercises (ranked third by the teachers) and the project. On fifth position with 18% (16 respondents) is the quiz, a technique which is not used by any of the teachers. Then there is the self-assessment with 14% (13 respondents); the check list (ranked on first position by the teachers) with 13% (12 respondents); the mutual assessment with 8% (7 respondents) and in last position is the portfolio with 6% (5 respondents). The result indicates that teachers certainly use various techniques and instruments for formative assessment. This practice

has an influence on students' achievements and it also contributes to the objectiveness of the assessment, (Standardi za ocenuvanje na učenicite vo osnovnoto obrazovanie, Nacrt, 2008 586 [Standards for assessment of students in primary education, Draft, 2008]).

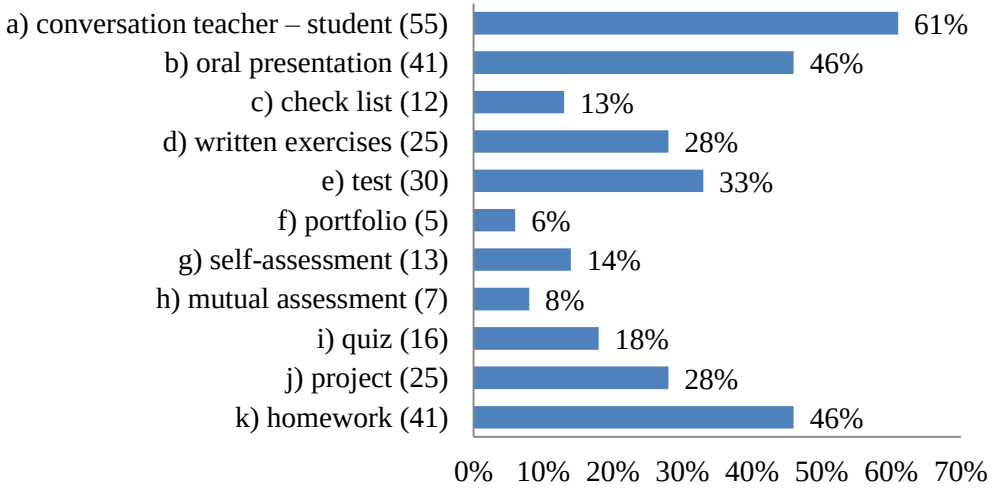


Figure 15. Question no.4. Which three of the following forms help you to learn more easily and to improve your knowledge?

However, regarding the difference between the results of the two questionnaires, (see Figure 8, Figure 15), it should also be suggested that teachers should take into consideration the viewpoint of their students. They should consider using techniques and instruments which help students learn more easily and improve their knowledge. For instance, it is evident that the quiz is not used by the teachers but part of the students prefers it as a technique for formative assessment. The results demonstrate that certain techniques and instruments are preferred by a small number of students, for instance, the self-assessment and the mutual assessment, as well as the project and the portfolio. This indicates that teachers should use them more often in the teaching practice, because studies confirm that they are powerful tools for formative assessment, (Talevski, 2011). Additinally, the importance of the use of other techniques and instruments which are not mentioned here should be emphasized. The constant use of the same techniques could lead to a decrease in students' motivation, to boredom and non-development of their potentials for creative and critical thinking.

The following result, presented in Figure 16, refers to the frequency of use of the tasks in the Macedonian language textbooks in general. Most of the students, i.e. 50% (45 respondents) state that the teacher uses them constantly, and 49% of them (44 respondents) state that the teacher uses them occasionally. Only 1 student (1%) states that the teacher never uses the tasks from the textbooks.

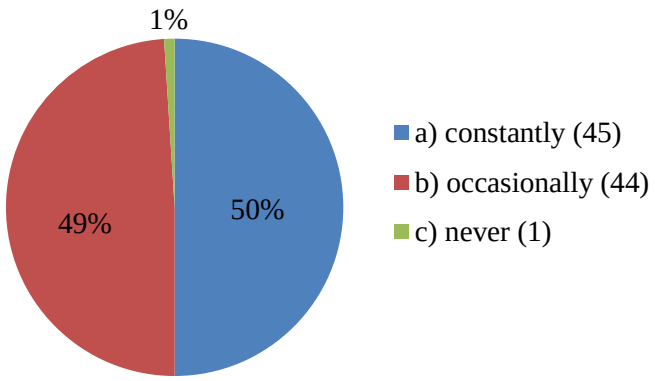


Figure 16. Question no.5. How often does your teacher use the tasks in the Macedonian language textbooks for the purpose of observing and assessing your achievements and development?

This result suggests the existence of certain doubt in the veracity of the teachers' statements, because the results are not entirely correlated, (see Figure 9). Sometimes, a great number of teachers give answers in the questionnaires that do not reflect the real situation in the teaching practice. Their aim is to demonstrate that they know all about the subject of the questionnaire and that they use the directions for formative assessment given by the Ministry of Education and Science and the Bureau for Development of Education. Nonetheless, teachers do not consider the directions when planning their teaching practice, i.e. they plan and realize the teaching practice according to their vision. This is confirmed by informal conversations with teachers. Moreover, regarding this result, the potential weakness of the instrument for obtaining results, i.e. the questionnaire should be taken into consideration. It is clear that the questionnaire could not contain unlimited number of questions. Most of the questions provide choice which means that teachers only have to choose an answer without giving an explanation. Also, there is a need for further research on this topic by using other methods and instruments, such as observation of classes, increase of the number of samples and population, and so on. However, it should be stated that the research serves as an illustration of the situation on national level, even though it has local character (the research is conducted in one region of Republic of Macedonia). This assertion is corroborated by: the results obtained with the conducted research; the information from other studies that have already been mentioned; our experience, observation and findings as teachers of Macedonian language; the findings gained as direct participants (disseminators) in the project by USAID (United States Agency for International Development).

As for the quality of the tasks in the Macedonian language textbooks, regarding their potential for assessment of students' achievements, the students had the opportunity to choose more than one answer. In Figure 17 it could be noticed that 52 respondents (58%) state that the tasks are comprehensible and very easy, 48 respondents (53%) state that they help them recognize their weaknesses and suggest them further directions for improvement, 45 respondents (50%) state that the tasks make the process of assimilating knowledge easier. Only 28 respondents (31%) state that the tasks have medium level of clarity and difficulty. Thereby, this result indicates that a greater number of students have positive viewpoint regarding the tasks in the Macedonian language textbooks in general. Smaller number of students, 8% (7 respondents) state that the tasks are incomprehensible and very difficult, and that they do not make the process of assimilating knowledge easier. Only 6% of the students (5 respondents) state that the tasks do not help them recognize their weaknesses.

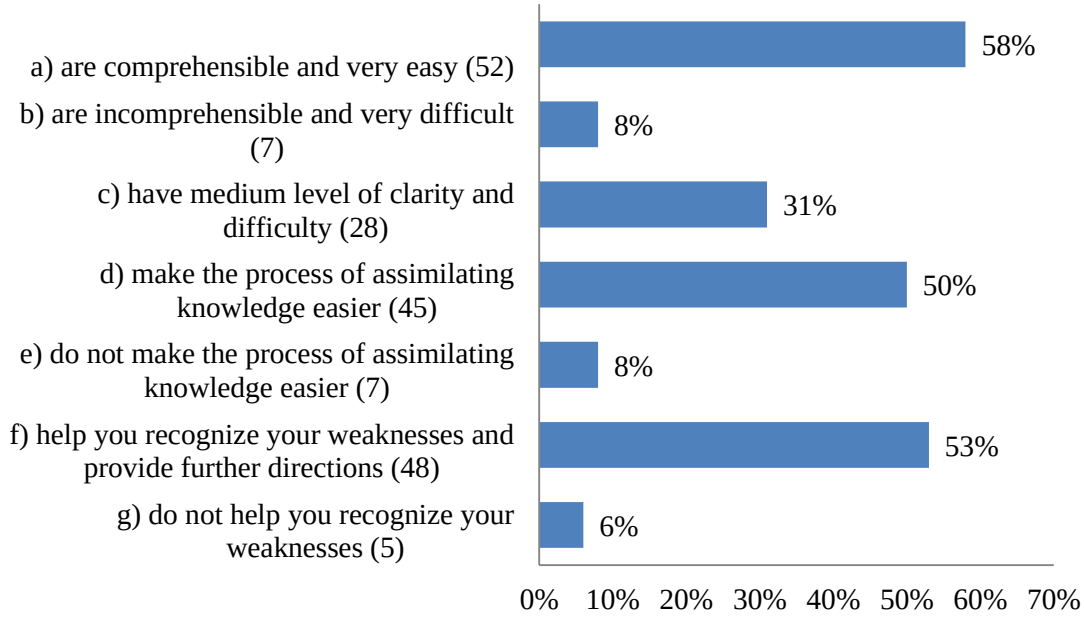


Figure 17. Question no.6. The tasks that are given in the Macedonian language textbook for the purpose of observation and evaluation of your knowledge…

Further, regarding the four tasks from the Curriculum (one task) and the Textbook (three tasks), a greater number of the students, 87% (78 respondents) state that they are suitable for their individual needs. The other 13% (12 respondents) do not agree with it. 83% of them (75 respondents) state that these four tasks actively involve them in the teaching process and the other 17% (15 respondents) do not agree with this. 77 respondents (86%) state that the tasks are suitable for estimating whether they have assimilated the required knowledge, and the other 13 respondents (14%) do not agree with it. It is more than certain that students differ regarding the individual abilities and needs, the style of studying and learning, the cognitive, affective, and psychomotor skills, and so on. They do not represent a homogenous group. However, these results indisputably indicate the need for reconsidering the following: the degree of suitability of the tasks intended for realization of formative assessment in the Macedonian language textbooks; the extent to which the tasks respond to the individual needs of the students; the need for incorporating tasks which would activate the students' higher cognitive processes; as well the need for compatibility of the textbooks with the directions and the prepositions for successful realization of formative assessment.

5. CONCLUSION

The assessment of students' achievements should be understood as a continuous process which is directly connected with the development of the entire personality of the student. That means that the emphasis should be on the process of learning which leads to support of the process of assimilating knowledge, and not on the results.

The results from the analysis of the Macedonian language teaching curriculum, from first to ninth grade of the nine-year primary education in the Republic of Macedonia, demonstrate that the formative assessment has a very strong institutional support. According to them, all prerequisites

and prepositions are ensured for correct interpretation of the directions for implementation of formative assessment in the teaching process.

The results from the questionnaires for teachers and students confirm the awareness of the teachers that the formative assessment should be part of the planning of the teaching practice. They also confirm that formative assessment is directly connected with students' achievements. It motivates the students to participate in the activities and it helps them achieve better results. That means that a great number of segments of formative assessment are implemented into the teaching practice to a great extent: sharing the teaching goals with the students; observation of and taking notes of students' achievements; continuous feedback; and use of tasks which are reflection of the individual needs of each student. A weaker segment is the involvement of the students in the process of assessment, which depicts the need for further improvement of the teachers' competencies regarding this segment. Studies confirm that students' involvement in the assessment significantly contributes to the quality of the teaching process. It also provides an increase in the number of students who think critically and students who are able to make realistic evaluation. Thereby, the purpose of education is to create citizens who accept responsibility for the improvement of the community. Further, the analyses demonstrate that teachers use various techniques and instruments for formative assessment. These methods provide greater validity of the collected information, so teachers could become more objective in the assessment of the students' achievements. Regarding the tasks from the Curriculum and the Textbook, the analyses indicates a correlation between the viewpoints of teachers and those of students. It demonstrates that the tasks from the Textbook refer to the first two levels of Bloom's Taxonomy, and that the third level (applying) is questionable when referring to the task from the Teaching Curriculum and the two tasks from the Textbook. Consequently, the following should be emphasized: the great carefulness when choosing suitable tasks with regard of the Bloom's Taxonomy; the need for creating tasks which would develop students' higher cognitive processes; and the need for more precise explanation and further clarification of the concept of the third level (applying), as well as of the other three levels (analysing, evaluating and creating). The other levels and especially the third one (applying) are interpreted in different manner by teachers and other people involved in the process of assessment of students' achievements. For instance, it is questionable whether the multiple choice questions could refer to the level applying, and so on.

It could be concluded that the implementation of formative assessment contributes to the improvement of the quality of the teaching process. The classroom becomes an environment that nurtures a culture which stimulates the interaction between teachers and students. This leads to partnership which promotes learning in an environment that gives to the students the possibility to state their viewpoints, and to the teachers the possibility to evaluate their performance and the quality of their activities.

6. REFERENCES

Almanah na proektot na USAID za osnovno obrazovanie. (2011). *[Almanac of the USAID project for primary education]*. Team of editors: Mojsoski, N., Dimitrovska, V., Miceva, E., Dimovska, A. Skopje: Project by USAID for primary education. http://www.scribd.com/doc/82280276/Проект-за-основно-образование-Алманах#scribd. Accessed 20 April 2015.

Andrade, H., and Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. Theory into Practice, 48: 1, 12–19. doi: 10.1080/00405840802577544

Author and Pejchinovska, M. (2011). Formative assessment in the teaching practice through the prism of the teachers. In Proceeding of International Science Conference: "Educational Technologies" 2011 on TU – Sofija, IPF – Sliven. Announcements of Union of Scientists – Sliven, Vol. 19: 71–76.

Black, P. (2007). *Formative Assessment: Promises or problems?*. King's College London [Adobe Acrobat Document]. Retrieved from https://www.learntogether.org.uk/resources/Documents/Formative%20Assessment%20in%20Music.pdf on 15 September 2014.

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, Vol. 5, Issue 1* [Microsoft Office Word Document]. Retrieved from http://area.fc.ul.pt/artigos%20publicados%20internacionais/Assessment%20and%20classroom%20learning.doc on 15 August 2014.

Black, P., and Wiliam, D. (2001). *Inside the Black Box: Raising Standards Through Classroom Assessment*. King's College London School of Education [Adobe Acrobat Document]. Retrieved from http://weaeducation.typepad.co.uk/files/blackbox-1.pdf on 15 August 2014.

Brookhart, M. S. (2008). Chapter 1: Feedback: An Overview. *How to Give Effective Feedback to Your Students*. ASCD Learn, Teach, Lead. http://www.ascd.org/publications/books/108019/chapters/Feedback@-An-Overview.aspx. Accessed on 2 October 2014.

Brookhart, M. S. (2008). Chapter 2: Types of Feedback and Their Purposes. *How to Give Effective Feedback to Your Students*. ASCD Learn, Teach, Lead. http://www.ascd.org/publications/books/108019/chapters/Types-of-Feedback-and-Their-Purposes.aspx. Accessed on 2 October 2014.

Brookhart, M., S. and Nitko, A. J. (2014). Chapter 1: Classroom Decision Making and Using Assessment, Chapter 2: Describing the Goals of Instruction. *Educational Assessment of Students*, Seventh Edition. Pearson Publisher [Adobe Acrobat Document]. Retrieved from http://www.pearsonhighered.com/assets/hip/us/hip_us_pearsonhighered/samplechapter/0133830268.pdf on 13 October 2014.

Bureau for Development of Education, official website http://bro.gov.mk/.

Clark, D. *Bloom's Taxonomy of Learning Domains*, updated July 2014. http://www.nwlink.com/~donclark/hrd/bloom.html. Accessed 13 October 2014.

Chappuis, S. and Stiggins, J. R. (2002). Classroom Assessment for Learning. Association for Supervision and Curriculum Development. Educational Leadership, pp. 40–43 [Adobe Acrobat Document]. Retrieved from http://hssdnewteachers.pbworks.com/w/file/fetch/50394085/Classroom.Assessment.for.Learning.Chappuis.pdf on 2 December 2014.

Gojkov, G. (2003). Dokimologija, Priručnik. [Dokimology, Handbook]. Second altered edition. Vršac: Viša škola za obrazovanje vaspitača "Mihailo Palov" [Preschool Teacher Training College "Mihailo Palov"].

Huitt, W. (2011). Bloom et al.'s taxonomy of the cognitive domain. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. [Adobe Acrobat Document]. Retrieved from http://www.edpsycinteractive.org/topics/cognition/bloom.html on 14 October 2014.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An Overview. *Theory into Practice, 41 (4)*, 212–218 [Adobe Acrobat Document]. Retrieved from http://ocw.metu.edu.tr/pluginfile.php/9009/mod_resource/content/1/s15430421tip4104_2.pdf on 24 November 2014.

McMillan, H. J. and Hearn, J. (2008). Student Self-Assessment: The Key to Stronger Student Motivation and Higher Achievements. Educational Horizons [Adobe Acrobat Document]. Retrieved from http://files.eric.ed.gov/fulltext/EJ815370.pdf on 24 November 2014.

Ministry of Education and Science, official website http://www.mon.gov.mk/.

Nastavna programa po makedonski jazik za I oddelenie vo devetgodišnoto osnovno obrazovanie. (2007). *[Macedonian language teaching program for first grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/osnovno-obrazovanie/nastavni_programi/I_odd_nastavna_programa_MK-ALB.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za II oddelenie vo devetgodišnoto osnovno obrazovanie. (2007). *[Macedonian language teaching program for second grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/osnovno-obrazovanie/nastavni_programi/II_odd_nastavna_programa_MK-ALB.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za III oddelenie vo devetgodišnoto osnovno obrazovanie. (2007). *[Macedonian language teaching program for third grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/osnovno-obrazovanie/nastavni_programi/III_odd_nastavna_programa_MK-ALB.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za IV oddelenie vo devetgodišnoto osnovno obrazovanie. (2009). *[Macedonian language teaching program for fourth grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/4_nastavni_programi.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za V oddelenie vo devetgodišnoto osnovno obrazovanie. (2008). *[Macedonian language teaching program for fifth grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/V_oddelenie_programi.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za VI oddelenie vo devetgodišnoto osnovno obrazovanie. (2008). *[Macedonian language teaching program for sixth grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/Nastavni%20programi%20VI%20odd.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za VII oddelenie vo devetgodišnoto osnovno obrazovanie. (2008). *[Macedonian language teaching program for seventh grade of the*

nine-year primary education]. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/osnovno-obrazovanie/6odd/nastavni%20programi/makedonski_jazik.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za VIII oddelenie vo devetgodišnoto osnovno obrazovanie. (2008). *[Macedonian language teaching program for eighth grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/osnovno-obrazovanie/VII%20oddelenie/Microsoft%20Word%20-%20makedonski%20jazik.pdf on 20 December 2014.

Nastavna programa po makedonski jazik za IX oddelenie vo devetgodišnoto osnovno obrazovanie. (2008). *[Macedonian language teaching program for ninth grade of the nine-year primary education]*. Skopje: Bureau for Development of Education [Adobe Acrobat Document]. Retrieved from http://bro.gov.mk/docs/makedonski_jazik.pdf on 20 December 2014.

Pejchinovska, M. and Author. (2014). *Diagnostic and Informative Function of Assessment in the Teaching Practice*. In Conference Proceedings of the 1st International Conference: Education Across Borders, pp. 543–551, Greece, Florina, ISBN: 978-618-81385-0-6, ISSN: 2241-8881. Retrieved from http://www.edu.uowm.gr/site/sites/default/files/crossborder_proceedings_2012_issn.pdf on 10 August 2014.

Petroska-Beška, V. (2007). *Ocenuvanje so testovi na znaenje*, Priračnik**.** *[Assessment with knowledge tests, Handbook]*. Skopje: Filozofski fakultet.

Poposki, K. (2005). *Učilišna dokimologija: sledenje, proveruvanje i ocenuvanje na postiganjata na učenicite. [School dokimology: observation, evaluation and assessment of students' achievements]*. Skopje: Kitano.

Priračnik: Vrednuvanje na znaenjata na postiganjata na ucenicite. (2007). *[Handbook: Assessment of the knowledge and achievements of students]*. Skopje: USAID, SEA. Retrieved from http://www.scribd.com/doc/61719535/1-Priracnik-Vrednuvanje-i-Ocenuvanje on 20 April 2015.

Priračnik: Unapreduvanje na ocenuvanjeto vo učilištata. (2007). *[Handbook: Enhancement of the assessment in schools]*. Skopje: USAID, SEA.

Questionnaire for students, Google form

https://docs.google.com/forms/d/1PXrvQQkjEu7RDomhhPjpO246b8gmgYX522PAxgYdD_Q/viewform.

Questionnaire for teachers, Google form

https://docs.google.com/forms/d/1fHm4W7VO3p4w2mNNdGSnCUharUNUQP7qQgJkFy22kV0/viewform.

Schiel, G. T. and Kitching, K. (2007). *Improvement of School Based Assessment*. Skopje: USAID [Adobe Acrobat Document]. Retrieved from

https://cora.ucc.ie/bitstream/handle/10468/441/Improvment_of_school.pdf?sequence=1 on 3 July 2015.

Standardi za ocenuvanje na učenicite vo osnovnoto obrazovanie, Nacrt: Upatstvo za ocenuvanje na učenicite vo osnovnoto učilište. (2008). *[Standards for assessment of students in primary education, Draft: Manual for assessment of students in primary education]*. Skopje: Ministry

of Education and Science. http://www.scribd.com/doc/240078301/Standardi-Za-Ocenuvanje-Upatstvo#scribd. Accessed on 20 April 2015.

Talevski, D., J. (2011). *Samoocenuvanje i ocenuvanje od součenici*. *[Self-assessment and mutual assessment]*. Proceedings from the International Conference: "Ocenuvanje za učenje vo 21 vek" ["Assessment for learning in the 21 century"]. Skopje: USAID, PEP.

Talevski, D. J., Author, Pejčinovska, M. (2014). *Formative Assessment and its Effects in the Teaching Practice*, WyzsaSzkolaPeadogiczna w Warszawie, University "St Kliment Ohridski" – Bitola, Modern Social and Educational Challenges and Phenomena – Polish and Macedonian Perspective, pp. 212–221, Warszawa – Bitola.

Velkova, S. and Jovanovska, S. (2011). *Učebnik po makedonski jazik za VI oddelenie*. *[Macedonian language teaching textbook for sixth grade]*. Skopje: Ministry of Education and Science [Adobe Acrobat Document]. Retrieved from http://www.e-ucebnici.mon.gov.mk/pdf/Makedonski_6.pdf on 2 October 2013.

Wiggins, G. (2012). *Seven Keys to Effective Feedback, Educational Leadership, Feedback for Learning*, v. 70, no.1, pp.10–16. ASCD Learn, Teach, Lead. http://www.ascd.org/publications/educational-leadership/sept12/vol70/num01/Seven-Keys-to-Effective-Feedback.aspx. Accessed on 1 October 2014.

Zbornik od megjunarodnata konferencija: Ocenuvanje za ucenje vo 21 vek. (2011). *[Proceeding of the international conference: Assessment for learning in the 21 century]*. Skopje: USAID, PEP.

International Journal of Assessment Tools in Education: Vol. 4, Issue 1, (2017) pp. 79-95

http://www.ijate.ne
e-ISSN: 2148-7456 © IJATI
INTERNATIONAL JOURNAL of ASSESSMENT TOOLS in EDUCATON

"Research Article"

The Development Study of Thoughts Scale Towards Measurement and Assessment Course on High Education

Serhat SÜRAL[1,*]

[1]Pamukkale University, Faculty of Education, 20070, Denizli, Turkey

Abstract

Considering that measurement and assessment, the most critical component of the education process, should be properly carried out by educators, it is necessary to make sure that education faculties deliver the measurement and assessment course meticulously and more importantly, positive and negative views of teacher candidates, who will teach future generations, should be determined with respect to measurement and evaluation. The present paper aims to develop "Opinions Scale" for measurement and assessment to see attitudes of teacher candidates of education faculties in terms of different perspectives. In this vein, it was aimed to reach all students attending the department of education faculty in 2016-2017 academic years and the study was carried out with 433 teacher candidates. To test teacher candidates' attitudes towards measurement and assessment, the Annals of Factor Analysis (AFA) and Confirmatory Factor Analysis (DFA) were employed. The scale was based on 3 sub-levels namely the requirement of the course, the course content and the instructor.

Article Info

Received
08 September 2016

Revised:
06 October 2016

Accepted
26 October 2016

Keywords:
Measurement and Assessment, Thoughts Scale, Exploratory Factor Analysis, Factor Analysis, Confirmatory Factor Analysis

Yükseköğretimde Ölçme Değerlendirme Dersine Yönelik Düşünceler Ölçeği Geliştirme Çalışması

Özet

Sürecin en kritik öğesi durumunda olan ölçme değerlendirme, öğretmenler tarafından sorunsuzca uygulanabilmesi gerekliliği göz önünde bulundurulduğunda eğitim fakültelerinde bu dersin sağlıklı bir şekilde yürütülmesi, daha da önemlisi gelecek nesilleri yetiştirecek olan öğretmen adaylarının ölçme değerlendirmeye yönelik olumlu ya da olumsuz düşüncelerinin neler olduğunun belirlenmesi gerekmektedir. Eğitim fakültelerinde öğrenim görmekte olan öğretmen adaylarının ölçme değerlendirme dersine yönelik farklı boyutlardan ele alınarak ne gibi düşüncelere sahip olduğunu görebilmek adına ölçme değerlendirme dersine yönelik düşünceler ölçeği geliştirilmesi bu araştırmanın amacını oluşturmaktadır. Araştırma kapsamında 2016-2017 eğitim öğretim yılında eğitim fakültesindeki öğrenim görmekte olan tüm anabilim dallarındaki öğrencilere ulaşılmaya çalışılmış ve 433 öğretmen adayına ulaşılmıştır. Çalışmada Açımlayıcı Faktör Analizi (AFA) ve Doğrulayıcı Faktör Analizi (DFA) analizi yapılmıştır. Öğretmen adaylarının ölçme değerlendirme dersine yönelik düşüncelerin ölçülmek istendiği ölçekte, dersin gerekliliği dersin içeriği ve son olarak öğretim elemanı şeklinde 3 alt boyut yer almaktadır.

Makale Bilgisi

Makale Gönderim
08 Eylül 2016

Makale Düzeltme:
06 Ekim 2016

Makale Kabul
26 Ekim 2016

Anahtar Kelimeler:
Ölçme Değerlendirme, Düşünceler Ölçeği, Açımlayıcı Faktör Analizi, Faktör Analizi, Doğrulayıcı Faktör Analizi

[*]Corresponding Author E-mail: ssural@pau.edu.tr

2148-7456 /© 2017 DOI: 10.21449/ijate.270300

1.GİRİŞ

Öğrenme öğretme süreci içerisinde öğretim programları tasarlanırken ortaya konulan en önemli amaç, öğrenci başarısını sağlamaya çalışmaktır. Öğrenci başarısını en doğru şekilde ortaya koyabilmek için etkili, kaliteli ya da bir başka deyişle güvenilir ve geçerli ölçme değerlendirme işlemi gerçekleştirmek gerekir. Öğrenciler üzerinde en önemli öğrenme ürünü olarak görülen akademik başarı günümüz eğitim anlayışında tek başına yeterli gelmemekte; öğrencileri diğer yönleriyle de tanımak, ölçmek ve ona göre değerlendirmek gerekmektedir. Kart ve Gülleroğlu (2013) geçmişten günümüze kadar öğrenci başarısı ve öğrenci başarısı ile ilişkili faktörlerin, önemle üzerinde durulan bir konu olduğunu söylerken; geleceğin nesilleri yetiştiren öğretmen, veli ve ilgili tarafların tüm çabaları ile birlikte çevresel koşulların öğrenci başarısı üzerindeki rolünün de yadsınamayacağını belirterek, öğrenci başarısının çok yönlü ele alınması gerektiğine vurgu yapmışlardır. Öğrencilerin bilişsel özelliklerinin yanında duyuşsal özelliklerinin de dikkate alınması büyük önem taşır. Öğrencilerin eğitiminde büyük rol üstlenen öğretmenlerin yetiştirildiği eğitim fakültelerinde, öğretmen adaylarının ölçme değerlendirme dersine yönelik duyuşsal özelliklerinin ne yönde şekillendiğine dikkat edilmelidir. Gelecek nesilleri yetiştirecek öğretmen adaylarının meslek hayatlarına başlamadan ölçme değerlendirme dersine yönelik düşüncelerin ne yönde geliştiğinin incelenip analiz edilmesi, hem öğretmen kalitesini artırmaya hem de daha kaliteli ve etkili bir eğitim sistemi oluşturmaya katkı sağlayacaktır.

Goodwin'e (2000) göre öğretmen kalitesini yükseltmeye yönelik gayretlerin yanında, son zamanlarda öğretim süreçlerine yönelik gerçekleştirilen yenilikler sonucunda özellikle şu üç öğenin ön plana çıktığı görülmektedir. Bunlar güçlü bir öğretim programının net bir şekilde ortaya konulması, öğretim programının etkili ve verimli öğretim uygulamaları içermesi ve son olarak doğru ölçme değerlendirme yöntemlerinin kullanılmasıdır. Öğretmen kalitesini yükseltmek, öğretmen yetiştirme programlarının ve buna bağlı olarak verilen eğitimin kalitesiyle doğru orantılı olduğu şüphesiz bir gerçektir. Öğretmen adaylarının lisans eğitimiyle birlikte kendilerinde oluşacakları yeterlik algısı, kaliteli öğretmen olma yolunda atacakları ilk adımdır. Woolfolk ve Spero'ya (2005) göre de öğretmenlerin yeterlik algılarındaki değişiklikler büyük oranda eğitim fakültelerindeki öğretmen yetiştirme programlarında öğrenim gördükleri süre içinde gelişmektedir.

Öğretmen adaylarının lisans eğitimiyle birlikte almış oldukları mesleki eğitim dersleri içerisinde ölçme değerlendirme dersine yönelik edindikleri düşünceler, yeterlik algıları üzerinde farklı etkiler yarattığı gözlenmektedir. Derse karşı ortaya konulan tutum, dersin içeriği, gerekliliği konusundaki düşünceler her zaman mesleki hayatlarında bu dersi sorgulama sebebi olmaktadır. Popham (2009) öğretmenlerin sınıfta içinde gerçekleştirilen ölçme değerlendirme uygulamaları ve buna dayalı ölçme-değerlendirme alanlarında kendilerini yetersiz hissetmeleri, verilen eğitimin kalitesini olumsuz yönde etkileyebilmektedir. Johnston'un (1992) da belirttiği gibi öğretmen adaylarının inançları ve tutumlarının; kullandıkları öğrenme ve öğretme yöntemini, öğrencilerin sınıf içindeki algılarını, kararlarını ve davranışlarını etkilediği sonucuna varılmaktadır.

Aktaş ve Alıcı'nın (2013) öğretmen adayı öğrencilerin ölçme değerlendirme alanındaki yeterliklere sahip olma düzeylerini etkileyen faktörlerin, bu alandaki çalışmaların gerekliliği ve önemi konusundaki duygu ve düşünceleridir. Yine Aktaş ve Alıcı (2013) ölçme değerlendirme alanındaki çalışmalarının önemine inanan, bu çalışmalara değer veren, bu çalışmaların ancak bilimsel ilke ve yöntemlere uygun bir biçimde yapılması durumunda nitelikli olabileceğini düşünen bir öğretmen ya da öğretmen adayının bu alandaki yeterlikler konusunda kendisini yetiştirme, geliştirme çabası içinde olacağını ve bu konuda mesleki anlamda kendisinden

beklenenleri nitelikli bir biçimde yerine getirebileceğini ifade ederek başta söylenenleri desteklemektedir.

Bir öğretmen adayının ölçme değerlendirme dersine yönelik ortaya koyacağı düşünceler, bu derse yönelik onda oluşacak olan tutumlar ile şekillenebileceği ifade edilebilir. Bireylerin bir duruma karşı göstereceği tutum onun düşünce yapısının şekillenmesine etki etmektedir. Üstüner'in (2006) de belirttiği gibi bir bireyin bir obje ya da uyarana karşı tutumunun ne olduğunun bilinmesi, o bireyin ilgili uyarana karşı davranışının da ne olacağını tahmin edilebilmesini sağlar. Bu durum uygulamada son derece önemli olmaktadır.

Bu ifadeden yola çıkılarak denilebilir ki; eğitimde ölçme-değerlendirmenin ana konularından birisi de öğretmenlerin inançları ve uygulamalarıdır (Richardson, 1996). Yaşar'ın (2014) aktardığına göre yapılan çalışmalarda tutum sadece öğrencilerin başarılarını etkilememekte bunun yanında gelecekteki mesleki yaşantılarını da etkileme gücüne sahip olduğu belirtilmektedir (Auzmendi, 1991; Gal ve Ginsgurg, 1994). Bir başka ifadede ise; öğretmenlerin ölçme değerlendirmenin doğasına ve amacına ilişkin inançlarının, ölçme değerlendirme yönelik hazırlanan yöntemleri ve uygulamaları da etkilediğini ortaya koymaktadır (Brown, 2002; Coll ve Remeasal, 2009).

Kilmen ve Çıkrıkçı Demirtaşlı'ya (2009) göre öğretmenlerin ölçme ve değerlendirme uygulamalarını olması gerektiği şekilde gerçekleştirememelerinin temel sebebinin; öğretmenlerin lisans eğitimleri sırasında almış oldukları ölçme ve değerlendirme alanında temel bilgi ve becerileri kazanmayı sağlayıcı ders ve uygulamaların yeterli olmadığıdır. Bunun yanında öğretmen adaylarının ölçme değerlendirme dersine yönelik, bu dersin işe vurukluk düzeyinin düşük olduğu düşüncesini benimsemeleri, dersin sayısal içerikli bölümünün pek çok öğrenciye hitap etmemesi de öğrencilerin bu derse yönelik akademik başarı sağlamada engel teşkil ettiği görülmektedir. Öğretmen adayları, mesleğe başladıklarında ölçme değerlendirmenin sadece sınav yapıp, not vermekten ibaret olduğuna inanmaları bu derse yeteri kadar önemin verilmemesine sebep olmaktadır.

Cronbach'ın (1990) da ifade ettiği gibi sınavlar öğrenciler hakkında değerlendirme yapabilmek için kriter konumunda olmasına karşılık kararlar alabilme açısından tek başına yeterli değildir. Kart ve Gülleroğlu'na (2013) göre de başarıyı değerlendirebilmek için ilgi, tutum ve motivasyon gibi niteliklerin de psikometrik ölçme araçları ile belirlenip, değerlendirilmesi gerekli görülmektedir. Ogan Bekiroğlu (2009) fizik öğretmeni adaylarının ölçme değerlendirmeye yönelik tutumlarını incelediği bir araştırmada, öğretmen adaylarının öncelikle ölçme değerlendirme bilgilerinin ve sonra da öz yeterliliklerinin, sınıflarında yaptıkları uygulamaları şekillendiren tutumları üzerinde etkili olduğunu ortaya koymuştur. Yine Karaman'ın (2014) yapmış olduğu bir çalışmada ise fen bilgisi öğretmenlerinin mikro öğretim yoluyla gerçekleştirilen öğretimin sonucunda ölçme değerlendirmeye yönelik tutumlarının olumlu yönde bir etkiye sahip olduğu sonucuna ulaşılmıştır.

Öğrencileri sadece bilişsel düzeyde başarılarını değerlendirmek bir öğretmenin tam anlamıyla ölçme değerlendirme yaptığı anlamlına gelmemelidir. Öğrencilerin tüm yönleriyle değerlendirilmeleri, sadece ürüne yönelik değil süreç içerisindeki gelişiminin izlenmesi doğru ve gerçek ölçme değerlendirme sürecini tanımlamaktadır. Öğretmen adaylarına yükseköğretim düzeyinde aldıkları eğitimde bu görüş ve tutumları kazandırmak bu dersi veren öğretim elemanlarının temel kazanımları arasında olmalıdır. Bu çalışmada yapılmak istenen öğretmen adaylarının ölçme değerlendirmeye yönelik tutumları üzerinden bu derse yönelik ne yönde düşüncelere sahip olduğunun belirlenmesi planlanmaktadır. Ölçme değerlendirme dersine yönelik

öğretim elemanları açısından, dersin içeriği ve bu dersin ne kadar gerekli olduğunu düşündükleri konularında öğretmen adaylarının edindikleri tutumdan yola çıkarak ne düşündüklerini ölçülebilir düzeye çıkarılması, ölçme değerlendirme dersine yönelik öğretmen adaylarının duyuşsal kazanımlarına farklı bir bakış açısı kazandıracaktır.

Sonuç olarak; öğrenme öğretme süreci içerisinde öğrenci başarısının ne düzeyde değiştiğini görebilmek, program içeriğinde verilen konuyu öğrencilere aktarabilmek için uygulanan yöntemin, kullanılan materyalin öğrenciler üzerinde ne denli kalıcı bir etki yarattığını anlayabilmek, öğretmen performansının ne düzeyde olduğunu görüp ona göre öz değerlendirmesini yapabilmesi için ölçme değerlendirmeye ihtiyaç vardır. Sürecin en kritik öğesi durumunda olan ölçme değerlendirme, öğretmenler tarafından sorunsuzca uygulanabilmesi gerekliliği göz önünde bulundurulduğunda eğitim fakültelerinde bu dersin sağlıklı bir şekilde yürütülmesi, daha da önemlisi gelecek nesilleri yetiştirecek olan öğretmen adaylarının ölçme değerlendirmeye yönelik olumlu ya da olumsuz düşüncelerinin neler olduğunun belirlenmesi gerekmektedir. Öğretmen adayının ölçme değerlendirmeye yönelik ortaya koyacağı düşünceler, onun bu alandaki başarısını olumlu yönde etkileyeceği ve daha kaliteli bir öğretmen olma yolunda önemli bir yeterliliğe sahip olacağı gerçeği düşüncesinden yola çıkarak geliştirilecek ölçek bu araştırmanın önemini oluşturmaktadır.

Eğitim fakültelerinde öğrenim görmekte olan öğretmen adaylarının ölçme değerlendirme dersine yönelik farklı boyutlardan ele alınarak ne gibi düşüncelere sahip olduğunu görebilmek adına ölçme değerlendirme dersine yönelik düşünceler ölçeği geliştirilmesi bu araştırmanın amacını oluşturmaktadır.

2. YÖNTEM

2.1. Araştırma Deseni

Bu çalışma, ölçek geliştirme amacıyla yapılması planlanan geçerlik ve güvenirlik çalışması üzerine desenlenmiştir.

2.2. Çalışma Grubu

Araştırma kapsamında çalışmaya eğitim fakültesindeki öğrenim görmekte olan tüm anabilim dallarındaki öğrencilere ulaşılmaya çalışılmıştır. Araştırmanın evreni olarak çalışma kapsamına alınan Eğitim Fakültesi'ndeki öğrenci sayısının 4908 olduğu; ancak örneklem grubu içerisinde 433 öğretmen adayına ulaşılmıştır. Çalışma grubunu faktör analizi tekniğinin kullanımı için önerilen madde sayısının beş katı örneklem büyüklüğü ölçütünü karşıladığı söylenebilir (Child, 2006).

Tablo 1. Örneklem Grubunda Yer Alan Öğretmen Adaylarının Frekans Dağılımı

Anabilim Dalları	Cinsiyet				Kişi Sayısı	
	Kız		Erkek		N	%
	N	%	N	%		
Sınıf Öğretmenliği	76	27,5	26	16,5	102	23,6
Okul Öncesi Öğretmenliği	25	9,05	19	12,1	44	10,2
Sosyal Bilgiler Öğretmenliği	49	17,7	16	10,1	65	15,0
İlköğretim Matematik Öğretmenliği	32	11,5	6	3,82	38	8,8
Fen Bilimleri Öğretmenliği	31	11,2	26	16,5	57	13,2
Türkçe Öğretmenliği	27	9,78	17	10,8	44	10,2
Güzel Sanatlar	19	6,88	27	17,1	46	10,6
Psikolojik Danışmanlık ve Rehberlik	17	6,15	20	12,7	37	8,5
TOPLAM	276	100	157	100	433	100

2.3. Ölçme Aracının Hazırlanması

Araştırmada kullanılan Ölçme Değerlendirme Dersine Yönelik Düşünceler Ölçeği öğretmen adaylarının ölçme değerlendirme dersine yönelik ne gibi düşüncelerinin olduğunu belirlemek amacıyla geliştirilmiştir. Ölçme değerlendirme dersini daha önceden almış 25 öğretmen adaylarına "Ölçme değerlendirme dersinin öğretmenlik mesleğindeki önemi, kendi alanınızdaki gerekliliği, bu dersi veren öğretim elemanının nasıl olması gerektiği, derse karşı şu ana kadar sizde oluşan olumlu ve olumsuz duygular konusundaki görüşleriniz, düşünceleriniz nelerdir?" şeklinde açık uçlu bir soru yöneltilmiştir. Bu soruya ilişkin yazmış oldukları cevaplardan yola çıkarak ölçek maddesi olabilecek cümleler çıkartılmıştır. Başta 44 maddeyle başlanan çalışmada gerçekleştirilen güvenirlik ve geçerlik analizleri sonucunda 23 maddeden oluşan bu ölçek 4 (4.,5.,7. ve 18. maddeler) olumsuz 19 (1.,2.,3.,6.,7.,8.,9.,10.,11.,12.,13.,14.,15.,16.,17.,19.,20.,21.22 ve 23. maddeler) olumlu ifade içermektedir ve 5'li Likert tipindedir. Ölçme aracı çalışma grubuna uygulandıktan sonra öğretmen adaylarının cevapları ifadelerin olumlu ve olumsuz olması da dikkate alınarak puanlanmıştır. Öğretmen adaylarının içtenlikle cevap verip vermediklerini kontrol etmek için aynı ifade ölçeğin başında ve sonunda tekrar sorularak kontrol maddesi işlevi sağlanmıştır (3. ve 37. maddeler).

Ölçekten madde atılması işlemine başlanmadan önce ilk olarak kontrol maddeleri arasındaki korelasyon ise 0,79 olarak hesaplanmış ve bu değerin 0,01 düzeyinde anlamlı olduğu tespit edilmiştir. Kontrol maddeleri arasındaki korelasyonun anlamlı çıkması öğretmen adaylarının ölçeğe cevap verirken içten davrandıkları konusunda fikir verebilmektedir. Verilerin analizi sırasında kontrol maddesi olarak kullanılan 37. Madde analizlerden çıkarılmıştır. Ölçek, 3 alt boyuttan oluşmaktadır. Öğretmen adaylarının ölçme değerlendirme dersine yönelik düşüncelerin ölçülmek istendiği ölçekte, dersin gerekliliği (4.,5.,7.,10.,11.,19. ve 20. maddeler), dersin içeriği (14., 15., 16. ve 18. maddeler) ve son olarak öğretim elemanı (1.,2.,3.,6.,8.,9.,12.,13.,17.,21.,22. ve 23.maddeler) şeklinde alt boyutlar yer almaktadır.

Tablo 2. Ölçeğin ve Alt Boyutlarının Güvenirlik Katsayıları

Faktörler	Cronbach Alpha Değerleri
1.Gereklilik Alt Boyutu	,794
2.İçerik Alt Boyutu	,722
3.Öğretim Elemanı Boyutu	,931
Genel	,817

Kullanılacak ölçeklerde; ön deneme çalışmaları için 0,60, temel çalışmalar için 0,80 ve uygulamalı çalışmalar için 0,90-0,95 güvenirlilik oranlarının gerekli olduğu belirtilmiş; sosyal bilimlerde yapılan araştırmanın türüne göre güvenirlilik katsayıları değişmekle birlikte, bilimsel içerikli çalışmalarda 0,70 ve yetenek, ilgi ve beceri gerektiren çalışmalarda kullanılacak ölçekler için ise 0,85 düzeyinde bir güvenirlik katsayısı istenmektedir (Şencan, 2005). Yapılan çalışmada ölçeğin tüm maddelerine yer verilerek yapılan güvenirlik analizi sonucuna göre Cronbach Alpha değeri .817 olarak ölçülmüştür.

2.4. Verilerin Analizi

İlk olarak hazırlanan taslak ölçekte yer alan maddeler 432 öğretmen adayının verdiği cevaplara göre bilgisayar ortamına alınmış ve öğretmen adaylarının hem madde düzeyinde hem de toplamda elde ettikleri puanlar hesaplanmıştır. Ölçeğin yapı geçerliliği için açımlayıcı faktör analizi (AFA) kullanılmış ve daha sonrasında ortaya çıkan faktörlerin uyum indeksine bakabilmek

için Doğrulayıcı Faktör Analizi (DFA) kullanılmıştır. Verilerin faktör analizine ve örneklem büyüklüğünün uygunluğu için Kaiser-Meyer-Olkin (KMO) ve Bartlett's testi kullanılmıştır. Ölçme aracından elde edilen verilerin faktör analizine uygunluğunun bir başka kanıtı olarak da Anti-imaj korelasyon matrisi analizi gerçekleştirilmiştir.

3. BULGULAR

Faktör analizi uygulamasında ilk olarak anti imaj korelasyon değerleri analizi gerçekleştirilmiştir. Anti-imaj korelasyon katsayılarına ait köşegen değerlerin .50'den büyük olması gerekmektedir (Can, 2014). Bu koşulun sağlanmadığı durumda ilgili maddeler ölçme aracından çıkarılarak kalan maddelerle tekrar faktör analizi yapılmıştır. Bu çalışmada yapılan Anti-imaj korelasyon matrisi sonuçlarına göre Tablo 3'te yer alan köşegen değerleri .590 (11. madde) ile .901 (1. madde) arasında değişmektedir.

Tablo 3. Anti İmaj Korelasyon Matrisi

	Madde 1	**Madde 2**	**Madde 3**	**Madde 4**	**Madde 5**	**Madde 6**	**Madde 7**	**Madde 8**	**Madde 9**	**Madde 10**	**Madde 11**	**Madde 12**	**Madde 13**	**Madde 14**	**Madde 15**	**Madde 16**	**Madde 17**	**Madde 18**	**Madde 19**	**Madde 20**	**Madde 21**	**Madde 22**	**Madde 23**
M1	**,901[a]**	-,459	-,123	,291	-,070	-,124	-,061	-,068	-,001	-,115	,137	,005	-,275	-,035	,072	-,104	,112	-,106	,139	-,088	,000	,054	,012
M2	-,459	**,882[a]**	-,214	-,084	,066	,046	-,074	-,077	-,215	-,156	,019	,072	,032	-,084	-,112	,096	-,318	,114	-,005	,359	,028	,000	-,149
M3	-,123	-,214	**,862[a]**	-,375	,102	-,184	,216	-,229	-,111	,331	,107	-,114	-,048	-,153	,008	-,026	,068	-,060	,036	-,316	,003	-,166	,202
M4	,291	-,084	-,375	**,694[a]**	-,272	,312	-,338	,349	,034	-,404	,156	-,002	-,211	,071	-,171	,129	-,152	-,024	-,113	,210	,087	,044	,051
M5	-,070	,066	,102	-,272	**,696[a]**	-,019	-,283	-,087	-,205	,059	-,295	,383	,078	-,025	,050	,141	,014	-,181	,002	-,274	-,028	,104	-,212
M6	-,124	,046	-,184	,312	-,019	**,887[a]**	,037	,020	-,192	-,125	-,073	,206	-,041	,093	,160	-,001	-,061	,020	-,230	,030	-,051	-,092	-,038
M7	-,061	-,074	,216	-,338	-,283	,037	**,766[a]**	-,131	,027	-,027	-,004	-,089	,066	-,044	,232	-,267	,180	,146	-,060	-,063	-,182	-,009	,006
M8	-,068	-,077	-,229	,349	-,087	,020	-,131	**,881[a]**	-,205	-,115	,040	-,129	,051	,020	-,184	,207	-,106	,080	-,049	,128	,175	-,106	,118
M9	-,001	-,215	-,111	,034	-,205	-,192	,027	-,205	**,897[a]**	-,064	,110	-,231	-,300	-,059	-,285	-,087	,236	,014	,176	-,003	,082	-,172	,091
M10	-,115	-,156	,331	-,404	,059	-,125	-,027	-,064	-,064	**,643[a]**	-,450	-,067	,185	,029	,120	-,047	,031	-,054	-,206	-,423	,056	-,132	,120
M11	,137	,019	,107	,156	-,295	-,073	-,004	,040	,110	-,450	**,590[a]**	-,198	-,341	-,013	-,214	,006	-,064	,060	,062	,284	-,031	,001	,159

	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23
M12	,005	-,275	-,035	,072	-,104	,112	-,106	,139	-,088	,000	,054	,012
M13	,072	,032	-,084	-,112	,096	-,318	,114	-,005	,359	,028	,000	-,149
M14	-,114	-,048	-,153	,008	-,026	,068	-,060	,036	-,316	,003	-,166	,202
M15	-,002	-,211	,071	-,171	,129	-,152	-,024	-,113	,210	,087	,044	,051
M16	,383	,078	-,025	,050	,141	,014	-,181	,002	-,274	-,028	,104	-,212
M17	,206	-,041	,093	,160	-,001	-,061	,020	-,230	,030	-,051	-,092	-,038
M18	-,089	,066	-,044	,232	-,267	,180	,146	-,060	-,063	-,182	-,009	,006
M19	-,129	,051	,020	-,184	,207	-,106	,080	-,049	,128	,175	-,106	,118
M20	-,231	-,300	-,059	-,285	-,087	,236	,014	,176	-,003	,082	-,172	,091
M21	-,067	,185	,029	,120	-,047	,031	-,054	-,206	-,423	,056	-,132	,120
M22	-,198	-,341	-,013	-,214	,006	-,064	,060	,062	,284	-,031	,001	,159
M23	**,774[a]**	-,029	-,285	,292	,232	-,264	,134	-,130	-,166	-,147	,128	-,345
M24	-,029	**,869[a]**	,183	,136	,014	-,189	-,028	-,050	-,149	-,206	,002	-,039
M25	-,285	,183	**,794[a]**	-,070	-,272	,275	-,318	,033	,051	-,103	-,054	,124
M26	,292	,136	-,070	**,657[a]**	-,422	-,060	,064	-,073	-,270	-,248	-,139	-,128
M27	,232	,014	-,272	-,422	**,598[a]**	-,230	-,138	-,239	,123	-,171	-,001	-,106
M28	-,264	,189	,275	-,060	-,230	**,816[a]**	-,281	,110	-,078	-,020	-,114	-,016
M29	,134	-,028	-,318	,064	-,138	-,281	**,740[a]**	-,028	,016	-,059	-,279	-,155
M30	-,130	-,050	,033	-,073	-,239	,110	-,028	**,686[a]**	-,059	,154	,154	-,344
M31	-,166	-,149	,051	-,270	-,001	-,078	,016	-,059	**,727[a]**	,023	-,279	-,114
M32	-,147	-,206	-,103	-,248	-,001	-,020	-,344	,154	,023	**,853[a]**	-,365	-,114
M33	-,129	-,039	,124	-,128	-,106	,016	-,100	-,255	-,142	-,104	-,711	**,812[a]**

3.1. Ölçme Aracının Yapı Geçerliği (Açımlayıcı Faktör Analizi)

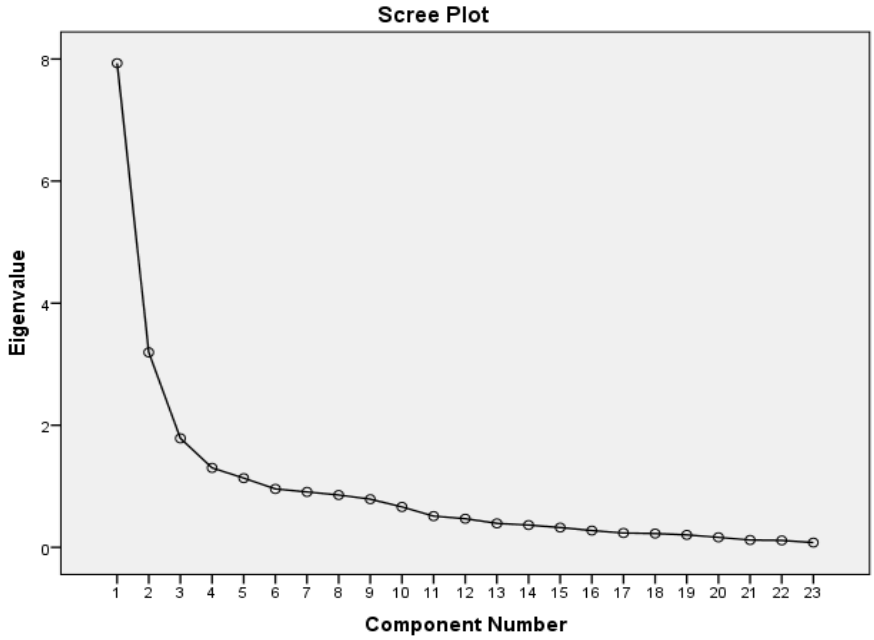
Verilerin faktör analizine uygunluğu ve çalışmanın yapıldığı örneklem sayısının büyüklüğünün yeterliliğini test etmek için yapılan Kaiser-Meyer-Olkin (KMO) testi sonucunda 0.803 değeri elde edilmiştir. Bu değer mükemmele yakın bir değer olarak görülmekte ve örneklem uygunluğunun çok yüksek düzeyde olduğunu göstermektedir (Şencan, 2005; Tavşancıl, 2006; Kalaycı, 2010). Bartlett's testine ait x^2 değeri ise 6652.284 ($p< .01$) olarak hesaplanmıştır. Bu sonuçlar, değişkenler arasında yüksek düzeyde korelasyon olduğunu göstermektedir. Bu değerler Tablo 4'te aşağıdaki gibidir.

Tablo 4. KMO ve Bartlett's Testi Sonuçları

Kaiser-Mayer-Olkin (KMO) Örneklem Ölçüm Değer Yeterliliği		0,803
Bartlett Testi	Ki-Kare	6652.284
	Sd	253
	Anlamlılık Düzeyi (p)	,000

p<0.01 düzeyinde anlamlı

Varimax faktöriyel döndürme yapıldıktan sonra faktör yük değeri 0.33'ün altında kalan ve birden fazla faktörde yer alan ve faktör yük değerleri arasındaki 0.10'dan küçük maddeler ölçekten çıkarılmıştır. Yavuz (2005) ve Bütüner ve Gür (2007) maddelerin birden fazla faktöre girmemesinin gerektiğini; birden fazla faktöre girme ile ilgili alınabilecek ölçüt faktör yükleri arasında en az 0,10 fark olmasına dikkat edilmesi ve iki faktördeki yük değerleri arasında 0,10'dan az fark olan maddeler binişik maddeler olarak adlandırıldığını ifade etmektedirler.



Şekil 1. Özdeğerlere Ait Çizgi Grafiği

Ölçeğin 3 faktör altında toplandığı Şekil 1'de verilen özdeğerlere ait çizgi grafiğinden de anlaşılmaktadır. Grafikte, birinci faktörden sonra yüksek ivmeli bir düşüşün olması ölçeğin genel bir faktöre sahip olduğunu göstermektedir. Birinci faktörün ölçek varyansının % 34.489'unu açıklaması bu sonucu destekler niteliktedir. Üçüncü faktörden sonra grafiğin genel gidişinin yatay olduğu görülürken, önemli bir düşüş eğilimi gözlenmediği söylenebilir. Sonuç olarak ölçeğin üç faktörlü bir yapıya sahip olduğu söyleyebilir. Ayrıca faktör sayısının belirlenirken faktör öz değerinin (eigenvalue) 1.00'den büyük olan faktörlerin dikkate alındığı da Tablo 5'teki değerlerle gösterilmiştir.

Faktör analizi sonucuna göre faktörlere ait öz-değeri 1'den büyük üç faktörün olduğu sonucuna varılmıştır. Dolayısıyla "Yükseköğretimde Ölçme Değerlendirme Dersine Yönelik Düşünceler Ölçeği" üç faktörden oluşan bir ölçek olarak tanımlanabilir. Tablo 5'e bakıldığında ortaya çıkarılan üç faktöre ait öz-değerler (eigenvalue) ve açıkladıkları varyans oranları gösterilmektedir. Ölçeğin birinci faktörü olan "öğretim elemanı" 12 maddeden (1, 2, 3, 6, 8, 9, 12, 13, 17, 21, 22 ve 23 no'lu maddeler) oluşurken, faktör öz değeri 7.932 ve ölçme değerlendirme dersine yönelik düşüncelerdeki değişimin ise % 34.489'unu açıklamaktadır. "Gereklilik" olarak adlandırılan ikinci faktör ise 7 maddeden (4, 5, 7, 10, 11, 19 ve 20 no'lu maddeler) oluşmakta ve faktör öz değeri 3.194 olup ölçme değerlendirme dersine yönelik düşüncelere ilişkin değişkenliğin de % 13.888'sini açıklamaktadır. Son olarak ortaya konan üçüncü faktör 4 (14, 15, 16 ve 18 no'lu maddeler) maddeden oluşmakta ve faktör öz değeri 1.785 ve açıkladığı varyans ise % 7.763'tür.

Tablo 5. Ölçme Değerlendirme Dersine Yönelik Düşünceler Ölçeği'ne Ait Faktör Öz Değerleri (Eigenvalue)

Faktörler	Faktör Başlangıç Öz Değerleri *(Initial Eigenvalues)*			Faktör Yükleri Kareler Toplamı *(Extraction Sums of Squared Loadings)*			Faktörlere Ait Betimsel İstatistikler		
	Toplam	Açıkladığı Varyans (%)	Kümülatif Açıklanan Varyans	Toplam	Açıkladığı Varyans (%)	Kümülatif Açıklanan Varyans	Faktör Ortalamaları	Faktör Standart Sapma Değeri	Güvenirlik Katsayısı α
1. Öğretim Elemanı	7,932	34,489	34,489	7,932	34,489	34,489	47,56	9,207	,931
2. Gereklilik	3,194	13,888	48,376	3,194	13,888	48,376	18,11	5,477	,794
3. İçerik	1,785	7,763	56,139	1,785	7,763	56,139	13,58	3,355	,722

Açımlayıcı faktör analizi (AFA) sonucunda, elde edilen üç faktörün açıkladığı toplam varyans % 56.139'dur. Şencan'a (2005) ve Can'a (2014) göre bu üç faktörün açıkladığı varyans oranı kabul edilebilir açıklama oranının üstünde bir oran olduğu kabul edilebilir. Ölçekte yer alan üç faktörün, ölçek toplamı ve faktörler arası ilişki düzeyini ortaya koyabilmek için pearson koreleasyon yöntemi uygulaması yapılmış olup elde edilen korelasyon katsayıları Tablo 6'da verilmiştir. Tablo 6 incelendiğinde ölçeğe ait üç faktörün hem faktörler arası hem de ölçek geneli ile anlamlı düzeyde bir ilişkiye sahip olduğu görülmektedir. Korelasyon katsayısı, 0.70 - 1.00 aralığında yüksek düzey; 0.69-0.30 aralığında orta düzey; 0.29-0.00 aralığında ise, düşük düzeyde bir ilişki olarak kabul edilmektedir (Büyüköztürk, 2006).

Tablo 6. Faktörler Arası Korelasyon Katsayıları

Faktörler		Faktör 1	Faktör 2	Faktör 3	Genel
Öğretim Elemanı	(Faktör 1)	*			
Gereklilik	(Faktör 2)	.612	*		
İçerik	(Faktör 3)	.596	.603	*	
Genel		.815	.735	.752	*

*p< 0.01 düzeyinde anlamlı

Faktörler arası ve faktörler ile genel ölçek arasında yapılan korelasyon analizi sonuçlarına göre ölçekten elde edilen toplam puan ile öğretim elemanı (faktör 1) alt boyutu arasında r=,815, gereklilik (faktör 2) alt boyutu arasında r=.735, içerik (faktör 3) alt boyutu arasında ise r=.752 düzeyinde korelasyon katsayısı tespit edilmiştir. Sonuç olarak ölçekteki 3 faktör ile genel ölçek puanları arasında yüksek düzeyde bir ilişkinin olması, ölçeğin yapı geçerliğinin yüksek olduğu sonucunu, KMO değeri ve Bartlett testi sonuçlarını da destekler nitelikte olduğunu ortaya koymaktadır.

Tablo 7. Elde Edilen Ölçekte Yer Alan Maddelerin Ortalaması, Standart Sapma Değerleri, Madde-Toplam Korelasyon Katsayısı, Madde Ortak Varyansı, Faktör Yük Değeri

	Ölçek Maddeleri	Madde Ortalaması	Madde Standart Sapma Değerleri	Madde-Toplam Korelasyonu	Madde Ortak Varyansı	Faktör Yük Değerleri
Faktör 1. Ölçme Değerlendirme Dersine Yönelik Düşünceler "Öğretim Elemanı"				*α =.931*		
1.	Bir öğretim elemanının ölçme değerlendirme alanındaki bilgi yönünden donanımı çok önemlidir.	4,10	1,194	,633	,697	,832
2.	Öğretim elemanının ölçme değerlendirme alanındaki akademik bilgisi, öğrencinin derse yönelik motivasyonunu olumlu ya da olumsuz yönde etkiler.	4,03	,957	,703	,729	,808
3.	Öğretim elemanının gerektiğinde teorik bilgilerin dışına çıkması, pratik bilgiler noktasında öğretmen adaylarını yönlendirebilmesi gerekir.	4,19	,895	,587	,650	,806
6.	Dersin içeriğinden ziyade, dersi veren öğretim elemanının dersi sevdirebileceğine inanıyorum.	2,65	1,309	,548	,624	,790
8.	Öğretim elemanının ölçme değerlendirme alanındaki akademik bilgisi, öğrencinin derse yönelik yeteneklerini, eksikleri doğrultusunda neyin öğrenilmesi gerektiğini olumlu ya da olumsuz yönde etkiler.	2,10	1,101	,686	,527	,763
9.	Bir öğretmenin ölçme değerlendirme bilgisi ne kadar fazlaysa, öğrencilerin akademik başarını o denli doğru ölçeceğine inanıyorum.	3,95	1,140	,597	,555	,751
12.	Ölçme değerlendirme dersinin içeriği çok soyut olduğu için öğretim elemanının bilgiyi aktarabilmesinde belli bir yeteneğinin olması gerektiğine inanıyorum.	2,50	1,225	,798	,585	,741
13.	Ölçme değerlendirme dersindeki bilgilerin nerede ve ne şekilde kullanacağımızın çok iyi öğretilmesi gerekir.	3,85	,959	,680	,553	,701
17.	Ölçme değerlendirme hocası olsaydım, özellikle istatistik konusunda daha somut örnekler kullanırdım.	3,88	,932	,735	,739	,696
21.	Öğretim elemanı ölçme değerlendirme dersini eğlenceli ve anlaşılır bir hale getirip dersi buna göre anlatmalıdır.	2,78	1,185	,686	,708	,657
22.	Ölçme değerlendirme dersi sayesinde öğrencilere geri bildirim verebilme açısından son derece önemli bir derstir.	2,75	1,267	,647	,592	,639
23.	Ölçme değerlendirme dersinde öğrenilmesi gereken ölçme araçlarının, uygulamalı olarak öğretilmesi gerekir.	3,94	,975	,606	,574	,617
Faktör 2. Ölçme Değerlendirme Dersine Yönelik Düşünceler "Gereklilik"				*α = .794*		
4.	Ölçme değerlendirme dersini ilk aldığım zaman "Evet" bu bilgileri meslek hayatımda kullanabilirim düşüncesi kendimde oluşmadı.	4,09	,907	,699	,655	,835
5.	Ölçme değerlendirme dersi, ne işe yaradığını bile anlamadığım işlem karmaşasından ibaret bir derstir.	3,69	1,163	,663	,510	,722
7.	Ölçme değerlendirme dersindeki konuların bir çoğunu meslek hayatımda kullanabileceğimi düşünmüyorum.	3,30	1,178	,733	,577	,694
10.	Ölçme değerlendirme dersinde sadece ölçme araçlarının ne işe yaradığını öğrensek bizim için yeterlidir.	3,34	1,001	,633	,634	,682

11.	Ölçme değerlendirme dersi, pedagojik anlamda kendimi çok iyi geliştirebileceğimi düşündüğüm bir derstir.	3,83	1,018	,620	,469	,656
19.	Ölçme değerlendirme dersinde öğrencilerin müzik, spor ya da sanatsal yeteneklerinin nasıl ölçüleceği konusunda hiçbir fikrim yok.	3,24	1,192	,698	,507	,522
20.	Ölçme değerlendirme dersinin meslek hayatımda ne gibi bir öneme sahip olduğu konusunda hiçbir fikrim yok.	2,26	1,069	,727	,595	,405
Faktör 3. Ölçme Değerlendirme Dersine Yönelik Düşünceler "İçerik"				*α = .722*		
14.	Sayısal içerikli derslere yatkınlığın ölçme değerlendirme dersindeki başarıyı arttıracağına inanıyorum.	3,06	,998	,770	,579	,789
15.	Ölçme değerlendirme dersi üniversitede bir dönemde verilecek kadar kolay bir ders değildir.	3,87	1,006	,650	,517	,723
16.	KPSS'nin eğitim bilimleri sınavındaki en zor soruların yer aldığı ders, ölçme değerlendirme dersidir.	3,87	1,103	,699	,726	,676
18.	Ölçme değerlendirme dersinin teorik kısmı ayrı, istatistik kısmı ayrı bir dönemde verilirse daha kolay anlaşılır.	3,95	1,061	,618	,610	,635

Tablo 7 incelendiğinde olumsuz anlamdaki maddelerin puanlanmasında ters çevirme yapıldıktan sonra en düşük ortalamanın 8.maddeye ($X_{ort.}$=2.10), en yüksek ortalamanın ise 3.maddeye ($X_{ort.}$= 4,19) ait olduğu görülmektedir. Yapı geçerliği çalışmasında oldukça büyük öneme sahip olan madde - toplam korelasyon değerlerine bakılmış ve maddelerin genellikle orta ve yüksek düzeyde ilişkilere sahip olduğu görülmektedir. Bu durum, maddelerin birbirleriyle ve geneliyle tutarlı olduğunu ve yapı geçerliğinin sağlandığı sonucunu ortaya koymaktadır. En yüksek korelasyon değerinin madde 12'ye (r= .798), en düşük korelasyonun ise madde 6'ya (r=.548) ait olduğu söylenebilir.

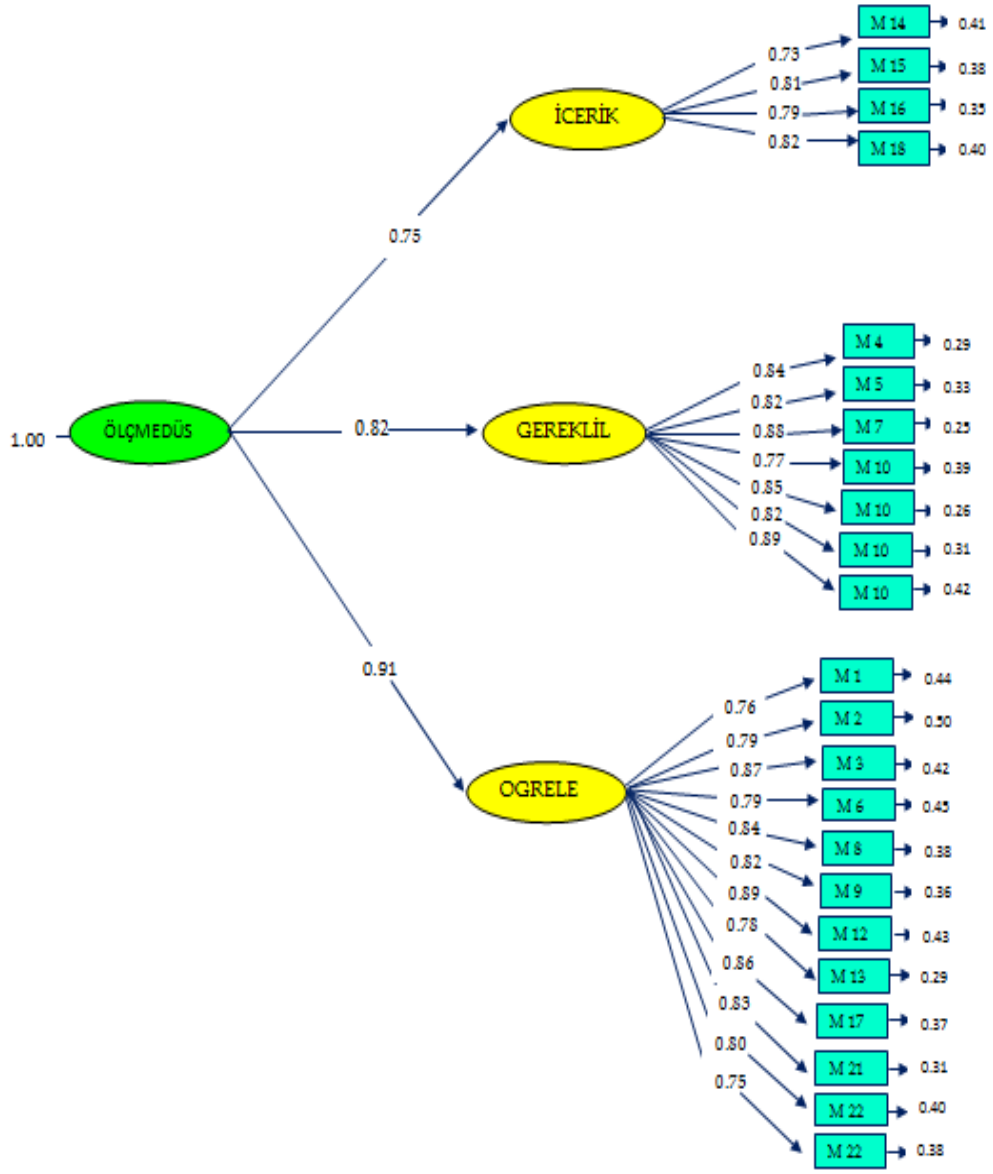
Tablo 8. Doğrulayıcı Faktör Analizi

Uyum İndeksleri	Uyum Aralığı	Araştırma Modeli
		Üç Faktörlü Model
Genel Uyum İndeksi		
χ^2/sd	$0 \leq \chi^2/sd \leq 3$	475.09 / 388= 1.22
Karşılaştırma Uyum İndeksleri		
NFI	.90 ≥ - ≥ .94	.91
NNFI	.90 ≥ - ≥ .94	.93
IFI	.90 ≥ - ≥ .94	.92
CFI	≥ ,95	.95
RMSEA	0.05 ≤ - ≤ 0.08	0.065
Mutlak Uyum İndeksleri		
GFI	≥ .90	.91
AGFI	≥ .85	.86
Artık Temelli Uyum İndeksleri		
SRMR	.06 ≤ - ≤ .08	.065
RMR		.071

Ölçeğin geçerliliğini belirlemek üzere yapılan açımlayıcı faktör analizi sonuçlarına göre belirlenen üç boyutlu yapının doğruluğunu sınamak için doğrulayıcı faktör analizine başvurulmuştur. Doğrulayıcı faktör analizi sonucunda Ki Kare (χ^2=475.09), serbestlik derecesi

(df=388, p=0.00) oranının χ^2/df=1.22; SRMR= .065, RMR=.071; AGFI= .86; GFI=.91; RMSEA= 0,065, CFI=.95, NNFI=.93, NFI=.91, IFI=.92, olduğu görülmektedir. Ki kare serbestlik derecesi oranında 3 den düşük, Karşılaştırma uyum indeksleri NFI, NNFI ve IFI değerlerinin .90 ile .94 arasında olması, CFI değerinin .95'ten büyük olması, RMSEA değerinin .05 ile .08 arasında olması, mutlak uyum indeksleri GFI'nin .90'dan, AGFI değerinin ise .85'ten büyük olması, son olarak artık temelli uyum indeksleri olan SRMR ve RMR değerlerinin ise .06 ile .08 arasında olması kabul edilebilir uyum iyiliğini göstermektedir (Şimşek, 2007; Yılmaz ve Çelik, 2009).

Bu verilerden yola çıkarak yükseköğretimdeki öğretmen adaylarının ölçme değerlendirme dersine yönelik düşünceler ölçeğiyle ilgili üç boyutlu yapının uygun olduğu söylenebilir. Doğrulayıcı faktör analizi ile hesaplanan standardize edilmiş madde faktör katsayıları Şekil 1' de sunulmuştur. Görüldüğü gibi madde-faktör doğrudan ilişki katsayıları .40 ile .83 arasında maddelerin hata varyansları .31 ile .84 arasında değişmektedir. Gözlenen madde ölçek ilişkilerinin anlamlı olduğu belirlenmiştir.



Şekil 2. ÖDDY-DÖ Üç Faktörlü Path Diyagramı

4. SONUÇ VE TARTIŞMA

Bu araştırmayla birlikte öğretmen adaylarının lisans eğitiminde almış oldukları meslek bilgisi dersleri içerisinde büyük öneme sahip olan ancak öğretmen adayları tarafından her zaman dersin içeriğine, mesleki hayatlarındaki gerekliliğine yönelik tartışma konusu olan ölçme değerlendirme dersine yönelik bir ölçek geliştirilmesi amaçlanmıştır. Bu amaçtan yola çıkarak öğretmen adaylarının ölçme değerlendirme dersine yönelik düşüncelerinin ne yönde olduğunun ölçülebilir bir hale getirerek, bu konu hakkında nicel veriler elde edilebilme fırsatı yaratılmaya çalışılmıştır. Bu bağlamda "Yükseköğretimde Öğretmen Adaylarının Ölçme Değerlendirme Dersine Yönelik Düşünceler Ölçeği" geliştirilmiş ve buna ilişkin geçerlik ve güvenirlik analizleri gerçekleştirilmiştir.

Ölçek geliştirme sürecinin en başında öğretmen adaylarına yöneltilen açık uçlu sorudan alınan cevaplar sonucunda, ölçme değerlendirme dersine yönelik öğretmen adaylarının ne gibi düşüncelere sahip olduğu analiz edilirken bu düşünce yapılarının farklı alt boyutlara bölünebileceği sonucuna ulaşılmıştır. Böylece ölçme değerlendirme dersine yönelik geliştirilen düşünceler ölçeği içerisinde bu dersi veren öğretim elemanlarının akademik ve kişisel özelliklerinin ölçüldüğü "öğretim elemanı", bu dersin öğretmen adaylarının meslek hayatlarına ne kadar katkı sağladığı ve işe vurukluk düzeyinin ne yönde olduğunun ölçüldüğü "gereklilik" ve ölçme değerlendirme dersi hakkında ortaya konulan düşüncelere dersin içeriği açısından da bakma fırsatı sağlayan "içerik" alt boyutları olmak üzere toplam üç alt boyutun oluştuğu sonucuna varılmaktadır.

Bir öğretim programı tasarlanırken sürecin son basamağı olarak görülen ölçme değerlendirmenin aslında sürecin her aşamasında aktif olarak kullanıldığı, kullanılması gerektiği tüm öğretmenler tarafından bilinmesi gereken bir durumdur. Öğretim sürecinin başında öğrencilerin hazır bulunuşluk düzeylerini tespit etmek amacıyla, içeriğin hazırlanıp uygulandığı aşamada öğrenme eksikliklerini görebilmek amacıyla yine uygulanan yöntemin etkililiğini test amacıyla ölçme değerlendirmenin yapılması gerektiği, en sonunda da hedeflenen kazanımların ne kadarına ulaşıldığını görmek adına ölçme değerlendirmenin gerekliliği son derece önemli bir öğretmenlik yeterliliği olduğu unutulmamalıdır. Yaşar (2014) yapmış olduğu benzer bir çalışmanın sonucunda "eğitim sistemi içinde yer alan milyonlarca öğrencinin geleceklerini şekillendiren öğretmenlerin, öğrencilerin gelecek yaşantılarının rotasını belirleme noktasında oldukları görülmektedir. Öğrencilerin geleceğini yönlendiren öğretmenlerin öğrencilerin mevcut ve geliştirilebilir özelliklerinin neler olduğunu ve öğrencilerin potansiyellerinin nasıl olduğunu ancak sağlam ölçme ve değerlendirme dersinden edindikleri bilgi ve becerileri kullanarak belirleyebilirler." şeklinde bir sonuca ulaşırken Balcı ve Tekkaya (2000) ölçme ve Değerlendirme konusunda Eğitim Fakültelerinde alınan derslerin yanı sıra hizmet içi eğitim kursları düzenlenerek öğretmenlerin bu konudaki bilgilerinin arttırılmaları ve yeni tekniklerden haberdar edilmesi yerinde olacağını ifade ederek hem lisans eğitimlerinde hem de hizmet içi eğitimlerinde öğretmen adaylarının ve daha sonrasında öğretmenlerin ölçme değerlendirmeye yönelik akademik anlamda kendilerini yeterli hissedecekleri her türlü çalışmanın yapılması gerektiği sonucuna ulaşılmaktadır.

5. KAYNAKLAR

Aktaş, M., alıcı, D. (2013). Eğitimde Ölçme Değerlendirme Dersine Yönelik Tutum Ölçeğinin Geliştirilmesi. *Journal of Qafqaz University,* 33. (66-73).

Auzmendi, E. (1991). Factors related to attitudes toward statistics: a study with a Spanish sample. *Paperpresented at the annual meeting of the American Educational Research Association*, Chicago, IL.

Balcı, E., Tekkaya, C. (2000). Ölçme değerlendirme tekniklerine yönelik bir ölçeğin geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 18, (42-50).

Baştürk, R. (2010). *Bütün yönleriyle spss örnekli non parametrik istatistiksel yöntemler*. Ankara: Anı Yayıncılık.

Brown, G.T.L. (2002). *Teachers'conceptions of assessment*. Unpublished doctoral dissertation, University of Auckland, New Zealand.

Bütüner, Ö.S ve GÜR, H. (2007). V diyagramına yönelik bir tutum ölçeğinin geliştirilme çalışması, *Milli Eğitim Dergisi*, 176 (1), 72–85.

Büyüköztürk, Ş. (2006). *Sosyal Bilimler İçin Veri Analizi El Kitabı*. Ankara: Pegem A Yayıncılık.

Can, A. (2014). *SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi.* Ankara: Pegem A Yayıncılık.

Child, D. (2006). *The Essentials Of Factor Analysis.* Third Ed. Continuum, London.

Coll, C., Remesal, A. (2009). Mathematics teachers'conceptions of learning assessment in compulsory education: dimensions and functions. *Infanciay Aprendizaje*, 32, 3, (391–404).

Cronbach, L.J. (1990). Essentials of psychological testing. Harper *and Row Publishers*. New York.

Çalışkan, H. ve Yazıcı, K. (2013). Ölçme değerlendirmeye yönelik tutum ölçeğinin geliştirilmesi ve sosyal bilgiler öğretmenlerinin tutum düzeylerinin çeşitli değişkenlere göre incelenmesi. *Internetioanl Journal of Human Science*, 10 (1), 398-415.

Gal, I., Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education*, 2, 1-54.

Gall, J., Gall, M.D., Borg, W.R. (1999). *Appling educational research*. New York: Longman.

Gay, L.R. (1987). Educational Research Compentencies For Analysis And Application. New York: Macmillan Publishing Company.

Goodwin, B. (2000). Raising the Achievement of low-Performing Students (McREL Policy Brief). *Mid-Continent Research for Education and Learning*. Aurora, CO.

Johnston, S. (1992). Images: A way of understanding the practical knowledge of student teachers. *Teaching and Teacher Education*, 8, 123- 136.

Kalaycı, Ş. (2010). *Faktör Analizi. SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*. (Edt: Ş. Kalaycı) Ankara: Asil Yayın Dağıtım.

Karaman, P. (2014). *Öğretmen adaylarının ölçme-değerlendirme okuryazarlıklarının belirlenmesi ve mikro-öğretim yoluyla geliştirilmesi*, yayımlanmamış doktora tezi. Çanakkale Onsekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü. Çanakkale.

Kart, A., Gülleroğlu, H.D. (2013). Demografik ve duyuşsal değişkenlerin ölçme ve değerlendirme dersi başarısını yordama gücü. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi* 10 (1), 11-30.

Kilmen, S. ve Çıkrıkçı Demirtaşlı, N. (2009). Sınıf öğretmenlerinin ölçme ve değerlendirme ilkelerini uygulama düzeylerine ilişkin görüşleri. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 42 (2), 27-55.

Mcmillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20 (1), 20-32.

Ogan Bekiroglu, F. (2009). Assessing assessment: Examination of pre-service teachers' attitudes towards assessment and factors affecting their attitudes. *International Journal of Science Education,* 31 (1), 1-39.

Quilter, S.M., Gallilini, J.K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36 (2).

Popham, W.J. (2008). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice* 48, (1), 4-11

Richardson, V. (1996). *The role of attitudes and beliefs in learning to teach*. In J. Sikula, T. J. Buttery, and E. Guyton (Eds.), Handbook of Research on Teacher Education, 2nd Edition Macmillan:New York.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

Şimşek, Ö. F. (2007). *Yapısal Eşitlik Modellemesine Giriş Temel İlkeler ve LISREL Uygulamaları.* Ankara: Ekinoks Yayınları.

Tavşancıl, E. (2006). *Tutumların Ölçülmesi ve SPSS İle Veri Analizi*. (3. Baskı). Ankara: Nobel Yayınları.

Tezbaşaran, A. (1996). Likert Tipi Ölçek Geliştirme Kılavuzu. *Tük Psikologlar Derneği Yayınları*. Ankara.

Üstüner, M. (2006). Öğretmenlik mesleğine yönelik tutum ölçeğinin geçerlik ve güvenirlik çalışması. Kuram ve Uygulamada Eğitim Yönetimi, 45, 109-127.

Woolfolk Hoy, A., Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21 (4), 343 - 356.

Yaşar, M. (2014). Eğitimde ölçme değerlendirme dersine yönelik tutum ölçeğinin geliştirilmesi. *Journal of Educational Science Research*, 4 (1), 259-279.

Yavuz, S.(2005). Developing a technology attitude scale for pre-service chemistry teachers, *The Turkish Online Journal of Educational Technology*, 4(1).

Yiğit, N., Bütüner, S.Ö., Dertlioğlu, K. (2008). Öğretim amaçlı örütbağ sitesi değerlendirme ölçeği geliştirme. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi,* 2 (2), 38-51.

Yılmaz, V., Çelik, E. (2013). *LİSREL 9.1 ile Yapısal eşitlik modellemesi: Temel kavramlar – uygulamalar – programlama*. Ankara: Anı Yayıncılık.

Summary

Inroduction

Only cognitive assessment of students' achievement does not necessarily mean that measurement and assessment are conducted thoroughly. Student assessment should be carried out in every aspect. In addition to product-based approach, students' improvement should be monitored for a proper and accurate measurement and assessment. Cultivating such views and attitudes in teacher candidates should be among instructors' basic course outcomes. In this study, it is planned to examine students' opinions on the course considering their attitudes towards measurement and assessment. Measuring students' attitudes in terms of the instructor, the course content and whether the course is necessary will give a different view of point to teacher candidates' affective acquisition. Measurement and assessment are essential to follow student progress in teaching/ learning process, to find out how permanently methods and material utilized for instructing topics within the course content influence students and to understand teacher performance allowing self-assessment. Considering that measurement and assessment, the most critical component of the education process, should be properly carried out by educators, it is necessary to make sure that education faculties deliver the measurement and assessment course meticulously and more importantly, positive and negative views of teacher candidates, who will teach future generations, should be determined with respect to measurement and assessment. The significance of the research is the scale to be developed viewing of the fact that student candidates' opinions towards measurement and assessment will positively impact on their success in this field and they will get a key competence to be more qualified.

Methodology

It was aimed to reach all students attending the department of education faculty. The number of the sampling group, students of education faculty, was calculated as 4908, however 433 teacher candidates participated the study. "Opinions Scale" was developed to determine teacher candidates' opinions about the measurement and assessment course. 25 teacher candidates who previously took the course were asked to answer the following open ended question: "What are your opinions and views regarding the role of the measurement and evaluation course in teaching profession, whether it is necessary in your own career, what skills the course instructor should be equipped with and positive and negative emotions that you have felt so far?". Based on the candidates' responses, sentences that could possibly be scale items were identified. To test construct validity, the Annals of Functional Analysis (AFA) was performed and Confirmatory Factor Analysis (DFA) were used to analyze fit indices in light of the factors obtained.

Results and Discussion

The study aimed to develop a scale regarding measurement and assessment course which is among the most significant teaching knowledge courses taken by teacher candidates in their undergraduate educations and is always debatable issue for teacher candidates in terms of course content and its requirement in practice. For this purpose, it has been tried to measure candidates' attitudes towards measurement and assessment course and to collect quantitative data. As a result, "Opinions Scale for Teacher Candidates' Attitudes towards the Measurement and Assessment Course in Higher Education" were developed and accordingly, validity and reliability analyses were conducted.

EK. ÖLÇME VE DEĞERLENDİRME DERSİNE YÖNELİK DÜŞÜNCELER ÖLÇEĞİ

	MADDELER	Kesinlikle Katılmıyorum	Katılıyorum	Az Katılıyorum	Katılıyorum	Kesinlikle Katılıyorum
1	Bir öğretim elemanının ölçme değerlendirme alanındaki bilgi yönünden donanımı çok önemlidir.	(1)	(2)	(3)	(4)	(5)
2	Öğretim elemanının ölçme değerlendirme alanındaki akademik bilgisi, öğrencinin derse yönelik motivasyonunu olumlu ya da olumsuz yönde etkiler.	(1)	(2)	(3)	(4)	(5)
3	Öğretim elemanının gerektiğinde teorik bilgilerin dışına çıkması, pratik bilgiler noktasında öğretmen adaylarını yönlendirebilmesi gerekir.	(1)	(2)	(3)	(4)	(5)
4	Ölçme değerlendirme dersini ilk aldığım zaman "Evet" bu bilgileri meslek hayatımda kullanabilirim düşüncesi kendimde oluşmadı.	(1)	(2)	(3)	(4)	(5)
5	Ölçme değerlendirme dersi, ne işe yaradığını bile anlamadığım işlem karmaşasından ibaret bir derstir.	(1)	(2)	(3)	(4)	(5)
6	Dersin içeriğinden ziyade, dersi veren öğretim elemanının dersi sevdirebileceğine inanıyorum.	(1)	(2)	(3)	(4)	(5)
7	Ölçme değerlendirme dersindeki konuların bir çoğunu meslek hayatımda kullanabileceğimi düşünmüyorum.	(1)	(2)	(3)	(4)	(5)
8	Öğretim elemanının ölçme değerlendirme alanındaki akademik bilgisi, öğrencinin derse yönelik yeteneklerini, eksikleri doğrultusunda neyin öğrenilmesi gerektiğini olumlu ya da olumsuz yönde etkiler.	(1)	(2)	(3)	(4)	(5)
9	Bir öğretmenin ölçme değerlendirme bilgisi ne kadar fazlaysa, öğrencilerin akademik başarını o denli doğru ölçeceğine inanıyorum.	(1)	(2)	(3)	(4)	(5)
10	Ölçme değerlendirme dersinde sadece ölçme araçlarının ne işe yaradığını öğrensek bizim için yeterlidir.	(1)	(2)	(3)	(4)	(5)
11	Ölçme değerlendirme dersi, pedagojik anlamda kendimi çok iyi geliştirebileceğimi düşündüğüm bir derstir.	(1)	(2)	(3)	(4)	(5)
12	Ölçme değerlendirme dersinin içeriği çok soyut olduğu için öğretim elemanının bilgiyi aktarabilmesinde belli bir yeteneğinin olması gerektiğine inanıyorum.	(1)	(2)	(3)	(4)	(5)
13	Ölçme değerlendirme dersindeki bilgilerin nerede ve ne şekilde kullanacağımızın çok iyi öğretilmesi gerekir.	(1)	(2)	(3)	(4)	(5)
14	Sayısal içerikli derslere yatkınlığın ölçme değerlendirme dersindeki başarıyı arttıracağına inanıyorum.	(1)	(2)	(3)	(4)	(5)
15	Ölçme değerlendirme dersi üniversitede bir dönemde verilecek kadar kolay bir ders değildir.	(1)	(2)	(3)	(4)	(5)
16	KPSS'nin eğitim bilimleri sınavındaki en zor soruların yer aldığı ders, ölçme değerlendirme dersidir.	(1)	(2)	(3)	(4)	(5)
17	Ölçme değerlendirme hocası olsaydım, özellikle istatistik konusunda daha somut örnekler kullanırdım.	(1)	(2)	(3)	(4)	(5)
18	Ölçme değerlendirme dersinin teorik kısmı ayrı, istatistik kısmı ayrı bir dönemde verilirse daha kolay anlaşılır.	(1)	(2)	(3)	(4)	(5)
19	Ölçme değerlendirme dersinde öğrencilerin müzik, spor ya da sanatsal yeteneklerinin nasıl ölçüleceği konusunda hiçbir fikrim yok.	(1)	(2)	(3)	(4)	(5)
20	Ölçme değerlendirme dersinin meslek hayatımda ne gibi bir öneme sahip olduğu konusunda hiçbir fikrim yok.	(1)	(2)	(3)	(4)	(5)
21	Öğretim elemanı ölçme değerlendirme dersini eğlenceli ve anlaşılır bir hale getirip dersi buna göre anlatmalıdır.	(1)	(2)	(3)	(4)	(5)
22	Ölçme değerlendirme dersi sayesinde öğrencilere geri bildirim verebilme açısından son derece önemli bir derstir.	(1)	(2)	(3)	(4)	(5)
23	Ölçme değerlendirme dersinde öğrenilmesi gereken ölçme araçlarının, uygulamalı olarak öğretilmesi gerekir.	(1)	(2)	(3)	(4)	(5)