Full research paper

IDENTIFYING THE FACTORS INFLUENCING MATHEMATICS TEACHERS' GRADING PRACTICES REGARDING STUDENTS' IN-CLASS PERFORMANCE: A RELIABILITY AND VALIDITY STUDY

Osman Birgin[1]✉
Murat Yılmaz[2]

[1]Usak University, Faculty of Education, Department of Mathmatics and Science Education, Türkiye

[2]100.Yıl Mehmetcik Middle School, Denizli, Türkiye

✉ osman.birgin@usak.edu.tr

ABSTRACT

This study aims to develop a scale to identify the factors influencing mathematics teachers' grading practices regarding students' in-class performance. The study was carried out with 180 secondary and 140 high school mathematics teachers from the southwestern region of Türkiye. The scale's construct validity was determined using item analysis, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA). The EFA results showed that the scale consisted of a 5-point Likert-type scale with 36 items under eight factors (mathematical knowledge and skills, social behaviors, affective skills, effort, homework performance, follow-up test results, academic exam results, and external benchmarks). The scale explained 70.02% of the total variance, with factor loadings ranging from 0.50 to 0.92. The item-total correlations ranged from 0.36 to 0.64, and *t*-test results for the item discrimination index were significant. The Cronbach's alpha coefficient for the overall scale was calculated to be 0.92. The CFA results indicated that the scale model had a good fit (Chi-square/*df* = 1.39; RMSEA = 0.051; IFI = 0.98; NNFI = 0.97; CFI= 0.98; and SRMR = 0.062). Based on the findings, the scale is a valid and reliable instrument that may be used in determining the factors influencing mathematics teachers' grading practices.

KEYWORDS

Grading, mathematics, participation in classroom activities, performance assessment, scale development

HOW TO CITE

Birgin O., Yılmaz M. (2023) 'Identifying the Factors Influencing Mathematics Teachers' Grading Practices Regarding Students' In-Class Performance: A Reliability and Validity Study', *Journal on Efficiency and Responsibility in Education and Science*, vol. 16, no. 3, pp. 196-207. http://dx.doi.org/10.7160/eriesj.2023.160304

Article history
Received
November 25, 2022
Received in revised form
January 23, 2023
Accepted
March 6, 2023
Available on-line
September 23, 2023

Highlights

- *This study aimed to develop a scale to identify the factors influencing the grading of students' in-class math performance.*
- *The study was carried out with 180 secondary and 140 high school mathematics teachers from Türkiye.*
- *The EFA and CFA results confirmed the construct validity of the scale.*
- *A valid and reliable scale with 36 items and eight subfactors (mathematical knowledge and skills, social behaviors, affective skills, effort, homework performance, follow-up test results, academic exam results, and external benchmarks) was created.*

INTRODUCTION

One of the primary functions of education is to prepare students to account for changing individual and societal demands over time. In today's information and technology era of rapid changes and developments, high-level thinking skills such as conceptual learning, making assumptions, problem-solving, criticizing, critical thinking, analyzing, and producing have increased in importance (Birgin, 2011). Many student-centered education approaches, such as cooperative learning, project-based learning, and social constructivist learning, have been proposed in response to these demands as alternatives to the traditional teacher-centered education approach (Shepard, 2000). Connecting concepts to everyday life, actively participating in the teaching process, and allowing students to construct their knowledge in a social learning environment are all critical components of a student-centered education approach (National Council of Teachers of Mathematics [NCTM], 2000). Assessment is an integral and formative part of the teaching process in the student-centered education approach, focusing on both the learning product and the learning process (Stiggins, 1997; Wiliam, 2011). As a result, the student-centered education approach necessitates the use of alternative assessment tools and methods that allow students to determine their learning performance, reveal their strengths and weaknesses, and measure high-level knowledge and skills (Birgin and Baki, 2009).

Assessment in education serves several purposes, including determining student success and deficiencies in the learning-teaching process; determining the effectiveness of teaching methods; revealing the weak and strong aspects of the applied curriculum; and monitoring student development (Ministry of National Education [MoNE], 2018). For effective and efficient assessment, both formative and summative purposes must be considered (Black and Wiliam, 1998; Stiggins, 1997). A summative assessment is performed at the end of the process to evaluate the teaching. Instead of at the end of instruction, formative assessment necessitates process-oriented assessment. Rather than grading, formative assessment seeks to identify instructional deficiencies (Birgin, 2011). As a result, formative assessment focuses on student performance and helps to eliminate deficiencies and organize students' knowledge.

Multiple-choice, matching, true-false, and fill-in-the-blank questions are frequently used in traditional assessment methods. Traditional assessment methods, however, fall short of ensuring students' growth throughout the process and addressing their learning deficits (Bennett, 2011; Guskey, 2011). On the other hand, student-centered assessment strategies provide teachers and students with new responsibilities. In this regard, the student is given the responsibility of actively engaging in the process and assessing both himself and his peers, while the teacher is responsible for planning and organizing the learning environment and assessing both the learning product and the learning processes (Shepard, 2000; Wiliam, 2011). As a result, the student-centered assessment approach requires the use of alternative assessment tools such as performance assessments, projects, portfolios, self- and peer-assessments, observations, interviews, structured grids, and concept maps in addition to traditional assessment tools to track a student's progress during the learning process and provide the necessary feedback.

When reviewing the literature, it is notable that some studies have focused on determining the factors that influence teachers' classroom assessment practices and grading behaviors (e.g., Bursuck et al., 1996; Cizek et al., 1995; Duncan and Noonan, 2007; Frary et al., 1992; Mc Millan et al., 2002; Sun and Cheng, 2014). The results of studies conducted in different countries have shown that many non-academic factors as well as academic achievement are effective in determining student grades. These factors include exam success, attendance, effort, attitude, homework habits, attendance, classroom behaviors, responsibility, bringing textbooks and materials, personal care behaviors, grade distribution in the classroom, grade distribution of other teachers, and school success policy (e.g., Cizek et al., 1995; Cheng and Sun, 2015; Frary et al., 1993; McMillan, 2001; McMillan et al., 2002; Randall and Engelhard, 2010). Furthermore, research has shown that teachers' grading practices differ depending on the curricula in the countries (e.g., Cheng and Sun, 2015; Duncan and Noonan, 2007; Demir et al., 2018; Suurtamm et al., 2010), the school type and grade level (Adams and Hsu, 1998; Cizek et al., 1995; Mc Millan, 2001), the in-service training support that teachers receive (Acar-Erdol and Yıldızlı, 2018; Zhang and Burry-Stock, 2003), school policies (Brookhart, 1994; Veldhuis et al., 2013), and the teachers' beliefs about assessment methods (e.g., Adams and Hsu, 1998; Sun and Cheng, 2014).

The mathematics curriculum in Turkey, which was updated in 2018 based on a student-centered education approach, focuses on the student's performance in the classroom and learning process as well as exam success. Previous studies have revealed that Turkish teachers lack the knowledge to use alternative assessment tools and methods recommended by the curriculum (e.g., Acar-Erdol and Yıldızlı, 2018; Serin, 2015) and prefer traditional assessment methods over alternative assessment methods (e.g., Birgin and Baki, 2009; Demir et al., 2019). Furthermore, teachers' alternative assessment practices were found to be negatively impacted by insufficient in-service training and material support (e.g., Akcadag, 2010; Özenç and Çakir, 2015; Yıldızlı, 2020), a heavy course load, crowded classrooms, resistance to change, and pressure from central exams (e.g., Birgin and Baki, 2012; Demir et al., 2018).

Although studies on teachers' grading practices in some countries have been conducted (e.g., Cross and Frary, 1999; Cheng and Wang, 2007; McMillan, 2001; Randall and Engelhard, 2010; Sun and Cheng, 2014; Zhao et al., 2016), it has been noticed that there is no scale to identify the factors influencing the grading practices of mathematics teachers in Türkiye. Therefore, this current study aims to develop a scale to identify the factors affecting Turkish mathematics teachers' grading practices.

Teachers' Grading Practices and Affecting Factors

Classroom assessment has been a complex process. There are several stages to this process, including the development and use of assessment tools, sharing and assessing results, and deciding on instructional strategies (Bennett, 2011). The choice of instructional activities, the observation of the student's growth during the teaching process, and the evaluation of his participation in the course activities are additional requirements for this procedure (Zhang and Burry-Stock, 2003). At this point, teachers' knowledge, abilities, and application strategies for classroom assessment are critical. In this context, some research has concentrated on the factors that influence grading and how teachers grade students in the classroom.

According to research findings (e.g., Adams and Hsu, 1998; Duncan and Noonan, 2007; Zhang and Burry-Stock, 2003), teachers' lack of knowledge and inadequacies about assessment methods raise concerns about the quality of classroom assessment practices. Furthermore, previous research (e.g., Acar-Erdol and Yıldızlı, 2018; Brookhart, 1993; 1994; Cizek et al., 1995; Cheng and Sun, 2015; Cross and Frary, 1999; Frary et al., 1992; Guskey, 2011; McMillan, Myran, and Workman, 2002; McMillan, 2001; Sun and Cheng, 2014) has shown that when determining a student's end-of-term grade, teachers generally do not consider the assessment experts' recommendations but partially do consider non-academic success factors such as homework, class participation, effort, ability, and attitude. Furthermore, it was revealed that teachers' grading practices differ according to the grade level (e.g., Bursuck et al., 1996; Duncan and Noonan, 2007; Gullickson, 1985; McMillan et al., 2002; Randall and Engelhard, 2010), field or course type (e.g., Duncan and Noonan, 2007; Frary et al., 1993; McMillan, 2001; Zhang and Burry-Stock, 2003), and county education systems and cultures (e.g., Cheng and Wang, 2007; Suurtamm et al., 2010; Zhao et al., 2016).

Bursuck et al. (1996) found that when grading students, primary and secondary school teachers consider factors such as effort, bookkeeping, class attendance, class participation, and preparation, with primary school teachers emphasizing student ability more than secondary and high school teachers. Duncan and Noonan (2007) determined in their study of 513 high school teachers that teachers value behaviors that promote success in grading practices (e.g., ability level, student effort, participation in the lesson, development, study habits, and negative behavior in the classroom) more than out-of-class benchmarks (e.g., the school's informal success policy, grade distribution of other teachers, and student success in previous years). Furthermore, mathematics teachers are less concerned with these factors than other branch teachers (social sciences, English, and visual arts).

McMillan et al. (2002) investigated classroom teachers' grading and assessment practices, as well as the factors they consider when grading students and the types of assessments they use. Six factors have been identified as influencing teachers' assessments of student success. Teachers were found to place a higher value on academic achievement and achievement-enhancing behaviors such as effort and development in grading while placing a lower value on factors such as doing homework, comparing with other students, the grade distribution of other teachers, and being on the borderline. Randall and Engelhard (2010) conducted a study with 516 elementary, secondary, and high school teachers in America, and found that teachers primarily follow the school district's official academic success policy and give grades based on success. However, they revealed that for borderline students, some teachers consider non-academic factors such as motivation, effort, and behavior. Similarly, Cizek et al. (1995) found that 61% of teachers use non-academic criteria in grading, such as behavior and effort. Cross and Frary (1999) revealed that 37% of secondary school teachers consider behavior and attitude, and Frary et al. (1992) determined that 31% of teachers agree that student behavior should be considered in grading. Based on previous research, it can be stated that it is necessary to take caution when deciding the grading, as non-academic factors and arbitrary grading practices will have a negative impact on the validity and reliability of students' grades.

Determination of Primary and Secondary School Students' Math Grades in Türkiye

In Turkey, the Ministry of National Education (MoNE) updated the mathematics curriculum after 2005 to reflect student-centered education approaches. The mathematics curriculum (MoNE, 2018) emphasized the importance of monitoring the development of students' cognitive, affective, and psychomotor knowledge and skills, valuing assessment practices for recognition and shaping, and considering both the learning process and the learning product in determining student success. It is encouraged to use alternative assessment tools and methods (project, performance, portfolio, group work, self-assessment, rubrics, etc.) in addition to traditional assessment tools to identify student progress and achievement in the learning process.

The Preschool Education and Primary Education Institutions Regulation (MoNE, 2014: 7) of the Ministry of National Education emphasizes the following factors in determining student success: Exam scores and participation in class activities are used to assess student success in the fourth grade of primary school. Exam scores, participation in course activities, and, if applicable, project work are used to assess the success of middle school students. The regulation defines participation in classroom activities as 'work that students do in the classroom or at school, activities that enable them to use and develop their cognitive, affective, and psychomotor skills such as critical thinking, problem-solving, reading comprehension, and research, as well as to evaluate their performance' (MoNE, 2014: 1). Furthermore, participation in class activities will be awarded three times for the mathematics course that exceeds two hours per week. Students are required to complete at least one performance study per semester and a project for at least one course per academic year. It is suggested that, in addition to academic success, student participation in class activities, as well as cognitive, affective, and psychomotor development, be taken into account when determining student achievement grades.

Following curriculum updates, some studies on teachers' assessment practices have been conducted in Türkiye over the last two decades. Only a few studies have examined the mathematics teachers' decision-making and grading practices in Türkiye (Yıldızlı, 2020). There is a need for a scale to assess the factors that influence students' participation in in-class mathematical activities in deciding their grades, considering that the Turkish mathematics curriculum was revised in 2018 and that in-class activities place a focus on participation. As a result, this study contributes to the development of a scale to identify the factors influencing students' decision-making performance in math class.

METHOD

In scale development studies, it is suggested that participation is based on volunteerism and sampling of the attribute being assessed (DeVellis and Thorne, 2021). The convenient sampling method enables the selection of people who are easily available and eager to volunteer for the research. As a result, participants were informed about this study and provided with an informed consent form.

Participants

The participants in this study are divided into two groups. The exploratory factor analysis of the scale was performed with Study Group I, and the confirmatory factor analysis of the scale was performed with Study Group II. Details of the participants in the study were summarized in Table 1. Data were obtained from a total of 320 mathematics teachers in a province in the southwestern region of Türkiye, 180 in middle schools (5th–8th grade) and 140 in high schools (9th–12th grade). In this study, 49.7% (n = 159) of the participants were female, and 50.3% (n = 161) were male teachers. 50% (n = 160) of the participants work in the city center, 31.6% (n = 101) of whom work in the provincial center, and 18.4% (n = 59) of whom work in rural areas. In terms of professional seniority, 15.3% of participants have a seniority of 1–5 years; 20.6% of them have a seniority of 6–10 years; 15.6% of them have a seniority of 11–15 years; 16.6% of them have a seniority of 16–20 years; and 31.9% of them have a seniority of over 20 years.

Gender	Study Group I (n = 170)		Study Group II (n = 150)		Total (n = 320)	
	Middle School	High School	Middle School	High School	n	%
Female	48	35	44	32	159	49.7
Male	45	34	43	39	161	50.3
Seniority						
1–5 years	12	11	14	12	49	15.3
6–10 years	20	14	17	15	66	20.6
11–15 years	16	10	14	10	50	15.6
16–20 years	17	12	15	9	53	16.6
+ 20 years	28	22	27	25	102	31.9

Table 1: Participants

The Development Process of the Scale

Table 2 summarizes the development processes of the scale to determine the factors that mathematics teachers consider when grading students' in-class math performance.

To begin the process of developing scale items, the literature on teachers' grading practices was reviewed (e.g., Brookhart, 1993; 1994; Cheng and Sun, 2015; Cizek et al., 1995; Duncan and Noonan, 2007; Frary et al., 1993; McMillan, 2001; McMillan et al., 2002; Sun and Cheng, 2014; Suurtamm et al., 2013). Furthermore, the assessment practices recommended in the Turkish primary and secondary school mathematics curricula were investigated. Moreover, the grading procedures and principles stated in the regulation of the primary and secondary education institutions in Türkiye (MoNE, 2014) were reviewed. Following the literature review, a pool of 45 items was created by taking into account the characteristics and factors used to grade the student's participation in classroom activities and performance.

Stage	Procedure
Making an item pool:	• Reviewing national and international studies and creating a draft item pool
Assuring the face and content validity:	• Obtaining comments from subject matter experts and math teachers on the draft items' content, language, and expression suitability • Making changes to the scale items based on the suggestions
Implementation of the draft scale:	• Applying the draft scale to the mathematics teacher to ensure the scale's validity and reliability
Item analysis:	• Calculating item-total correlation and item discrimination
Analysis construct validity:	• Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tests were used to assess construct validity. • Performing exploratory factor analysis (EFA) and confirmatory factor analysis (CFA)
Reliability:	• Calculating Cronbach's alpha coefficient

Table 2: Development stages of the scale

The draft scale was reviewed by two mathematics education experts, a measurement and assessment expert, and two educational sciences experts to ensure its content and face validity. Five items were removed, three new items were added, and some items were corrected in response to expert opinions and suggestions. Furthermore, the opinions of eight experienced mathematics teachers were solicited in terms of content, language, and expression, and corrections were made to the three items following their suggestions. There are 43 items in total in the final draft scale to determine the features and factors that are taken into account in the grading of the student's in-class mathematics performance. A 5-point Likert scale was used to respond to the scale item (1 = not at all, 5 = completely).

The validity and reliability of the scale were carried out in two stages. In the first stage, the scale was administered to 170 mathematics teachers for item analysis of the draft scale. Both the item discrimination index and the item-total correlation were computed. The item discrimination index was determined using 27% lower and 27% upper group methods. To assess the scale's conformance with the construct validity, the Kaiser-Meyer-Olkin (KMO) sample fit test and Bartlett's sphericity test were first performed (Field, 2005). To examine the factor structure of the scale, exploratory factor analysis (EFA) was performed using principal component analysis and oblique rotation method (direct oblimin), which allow correlations between the factors. The Cronbach's alpha coefficient was used to assess the reliability of the scale as well as that of its sub-dimensions. In the second stage, confirmatory factor analysis (CFA) was carried out on 150 mathematics teachers using LISREL 8.8 to assess the model's fit for the scale's factors.

Data Analysis

The scale was applied to the participants, and their responses to each item were scored from 1 (not at all) to 5 (completely). The statistical software programs LISREL 8.5 and SPSS 17.0 were used to conduct the data analysis. For the convenience of the analysis and the assumptions, the evaluation of the blank data, the normality test, the linearity, and the determination of extreme values (excluding the -3 and +3 values of the standard deviation) were evaluated before the data analysis. Outlier data scores were removed from the analysis. The answers provided by 140 high school math teachers and 180 middle school math teachers were evaluated as a result of the data analysis. It was found that the kurtosis (-0.724) and skewness (-0.294) values for the scale, which ranged from +1 to -1, indicated a normal distribution. The discrimination indices of the items were examined for the 27% lower and 27% upper groups, as well as the item-total correlation values. Also, EFA and CFA were performed to confirm the construct validity of the scale. Cronbach's alpha coefficient was calculated to determine the reliability of the scale.

RESULTS

This section presents the item analysis test results for the construct validity and reliability of the scale, the EFA and CFA, and reliability test findings.

Findings Relating to the Scale's Item Analysis

Item-total correlation and *t*-test scores for 27% of the upper and lower groups were calculated for the scale's item analysis. A positive and high item-total correlation indicates that the test's internal consistency is good and that the items exhibit similar behaviors. The degree to which items identify individuals in terms of the measured feature can also be determined using item-total correlation. According to the literature, highly discriminating items have an item-total correlation of 0.30 or higher; items between 0.20 and 0.30 can be included in the scale or modified if they are judged essential; and items less than 0.20 should be eliminated from the scale (DeVellis and Thorne, 2021). For this reason, when selecting the items for the scale, it was assumed that the item-item correlation was not extremely high ($r > 0.90$) and that the item-total correlation was above 0.30. One item with an item-total correlation between 0.20 and 0.30 ($r = 0.237$ for B13) and two items with item-total correlations below 0.20 ($r = 0.108$ for B36 and $r = 0.194$ for B37) were excluded from the scale. The item-item correlation values calculated for the remaining 40 items in the scale were found to be not very high ($r < 0.90$), and the significant item-total correlation values ranged from 0.36 to 0.64 (Table 3).

To assess the distinctiveness of the items on the scale, the item mean scores of the 27% upper and lower groups were compared using the *t*-test. The significant difference is considered evidence of the internal consistency of the scale (DeVellis and Thorne, 2021). As shown in Table 3, the *t*-test results for the 40 items on the scale were determined to be significant ($p < 0.01$). Based on these findings, it was accepted that the items on the scale were distinctive.

Item No	Item-Total Correlation	*t*–value	Item No	Item-Total Correlation	*t*–value
B1	0.508**	5.464**	B22	0.640**	9.249**
B2	0.570**	6.690**	B23	0.509**	6.477**
B3	0.417**	6.883**	B24	0.529**	7.852**
B4	0.438**	7.635**	B25	0.537**	6.758**
B5	0.493**	7.829**	B26	0.493**	7.283**
B6	0.485**	7.241**	B27	0.557**	6.628**
B7	0.493**	7.805**	B28	0.592**	7.332**
B8	0.511**	7.075**	B29	0.552**	8.419**
B9	0.496**	4.740**	B30	0.596**	8.557**
B10	0.501**	5.608**	B31	0.494**	8.617**
B11	0.544**	5.706**	B32	0.602**	8.290**
B12	0.512**	5.245**	B33	0.597**	8.247**
B14	0.542**	7.823**	B34	0.599**	7.816**
B15	0.519**	7.078**	B35	0.570**	7.017**
B16	0.509**	7.470**	B38	0.479**	7.046**
B17	0.524**	7.154**	B39	0.354**	5.008**
B18	0.561**	6.578**	B40	0.455**	6.562**
B19	0.588**	7.358**	B41	0.544**	8.314**
B20	0.507**	6.951**	B42	0.510**	7.937**
B21	0.519**	6.946**	B43	0.367*	5.022**

**p < 0.05 **p < 0.01*

Table 3: Item analysis results for the scale

Findings Relating to the Exploratory Factor Analysis (EFA)

Factor analysis is a statistical technique for combining variables that assess the same structure (Field, 2005). To determine the scale's factor structure, EFA was performed. Bartlett's sphericity test and Kaiser-Meyer-Olkin (KMO) proficiency test results were examined to determine whether the data were adequate for factor analysis. It is emphasized that for the data to be suitable for factor analysis, the result of the KMO test should be greater than 0.60 and the result of the Bartlett's sphericity test should be significant (Field, 2005). In this study, the KMO test result was found to be 0.867, and the Bartlett's sphericity test result was significant (χ^2 = 4899.104, p < 0.001). These findings indicated that the data were suitable for factor analysis.

Principal component analysis and the oblique rotation method (direct oblimin) were used to conduct EFA. The oblique rotation method allows the factors to correlate (Tabachnick and Fidell, 2007). In the process of factor analysis, it was required that the factor loadings of the items be at least 0.40 and that, when the items were collected under different factors, the difference between the factor loadings be at least 0.10. According to the first EFA results, the scale was categorized into nine factors with an eigenvalue above 1.00, which could account for 71.63% of the total variance. Four items (B14, B15, B35, and B40) were found to have high values in several factors, with the difference between factor loadings being smaller than 0.10. Due to this, it was decided to remove four items from the scale and repeat factor analysis on the remaining 36 items. The results of repeated factor analysis indicated eight factors with eigenvalues greater than one, accounting for 70.02% of the total variance, with factor loadings ranging from 0.50 to 0.92. The factor analysis results are shown in Figure 1 as a line graph of the eigenvalues, and the sub-factor loadings are shown in Table 4.

According to the results of the principal component analysis, the common factor variances of the items varied from 0.545 to 0.854, and the total variances explained by each factor after oblique rotation were 29.35%, 8.82%, 8.60%, 6.67%, 5.41%, 4.19%, 3.62%, and 3.32%, respectively. As seen in Table 3, the first-factor loadings ranged from 0.509 to 0.733, the second-factor loadings ranged from 0.626 to 0.808, the third-factor loadings ranged from 0.800 to 0.923, the fourth-factor loadings ranged from 0.789 to 0.852, the fifth-factor loadings ranged from 0.709 to 0.836, the sixth-factor loadings ranged from 0.575 to 0.830, the seventh-factor loadings ranged from 0.839 to 0.893, and the eighth-factor loadings ranged from 0.501 to 0.873.

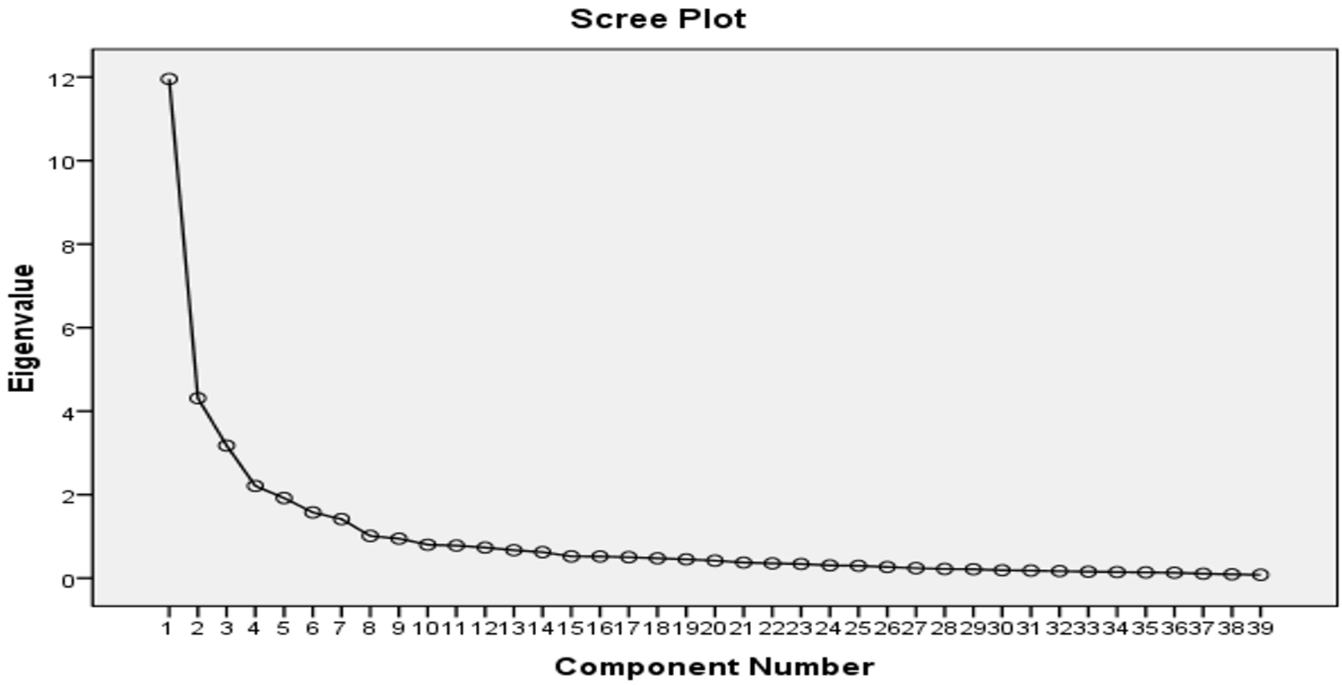

Figure 1: Eigenvalue screen plot graph of the scale

The sub-factors that resulted from the EFA were given names by considering the properties of the items they include. The sub-factor items of the scale were given in Appendix. The first factor, which contains seven items, was given the name 'mathematical knowledge and skills'. Due to external benchmark factors influencing the student's academic performance, the second factor, which contains five items, was given the name 'external benchmarks'. Because they reflect the student's social attitudes and behaviors in the classroom and at school, the third factor, which contains six items, was given the name 'social behaviors'. The fourth factor, which contains three items, was given the name 'homework performance'. The fifth factor, which contains four items, was given the name 'effort'. The sixth factor, which contains six items, was given the name 'affective skills'. The seventh factor, which contains two items, was given the name 'academic exam results'. The eighth factor, which contains three items, was given the name 'follow-up test results'.

The Pearson correlations between the scale's overall scores and its sub-factors were calculated as another construct validity

Item	Factor Common Variance	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
B17	0.643	0.733							
B18	0.728	0.728							
B16	0.557	0.713							
B21	0.573	0.689							
B19	0.715	0.577							
B20	0.661	0.529							
B22	0.597	0.509							
B43	0.668		0.808						
B41	0.668		0.764						
B38	0.585		0.757						
B39	0.583		0.742						
B42	0.545		0.626						
B31	0.806			-0.923					
B34	0.775			-0.863					
B32	0.854			-0.861					
B30	0.813			-0.856					
B33	0.688			-0.804					
B29	0.694			-0.800					
B4	0.817				0.852				
B3	0.796				0.852				
B5	0.770				0.789				
B9	0.746					0.836			
B12	0.725					0.826			
B11	0.759					0.744			
B10	0.611					0.709			
B24	0.668						-0.830		
B26	0.706						-0.740		
B27	0.749						-0.734		
B28	0.630						-0.616		
B23	0.646						-0.589		
B25	0.679						-0.575		
B2	0.834							0.893	
B1	0.735							0.839	
B7	0.777								-0.873
B8	0.816								-0.843
B6	0.591								-0.501
Eigenvalue:		10.56	3.17	3.09	2.40	1.95	1.51	1.31	1.20
Explained Variance:		29.35%	8.82%	8.61%	6.68%	5.42%	4.19%	3.63%	3.32%
Cronbach's alpha (α):		0.90	0.85	0.93	0.83	0.90	0.86	0.83	0.80

Table 4: Pattern matrix of scale and factor loadings

Variables	1	2	3	4	5	6	7	8
Total Scale Score	0.708**	0.335**	0.642**	0.557**	0.585**	0.621**	0.643**	0.565**
1. Math Knowledge and Skills	-	0.216**	0.334**	0.351**	0.393**	0.547**	0.454**	0.405**
2. External benchmarks		-	0.284**	0.072	0.011	0.243**	0.008	0.228**
3. Social Behaviors			-	0.341**	0.329**	0.521**	0.193*	0.256**
4. Homework performance				-	0.410**	0.365**	0.127	0.438**
5. Effort					-	0.438**	0.200**	0.279**
6. Affective skills						-	0.371**	0.279**
7. Academic exam results							-	0.208**
8. Follow-up test results								-

*$p < 0.05$ **$p < 0.01$

Table 5: Pearson correlation coefficient for scale sub-factors (n = 170)

indicator (Table 5). Table 5 revealed a moderately positive and significant relationship between the overall scale and the sub-factors ($p < 0.01$), ranging from 0.335 to 0.708. Additionally, it was found that the scale sub-factors did not have a high correlation.

Findings Relating to the Confirmatory Factor Analysis (CFA)

CFA is another way to test the scale's construct validity (Field, 2005). The most common goodness-of-fit indices of the model in the CFA were used to decide whether a model should be accepted or not. Fit indices used in this current study include Chi-square (χ^2), Chi-square/degrees-of-freedom ratio (χ^2/df), Standardized Root of Squared Residual (SRMR), Goodness of Fit Index (GFI), Non-Normed Fit Index (NNFI), Incremental Fit Index (IFI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Adjusted Goodness-of-Fit Index (AGFI). For model fit, the ratio of χ^2/df should be less than 3, but less than 5 is also acceptable. When CFI, NNFI, IFI, and GFI are higher than 0.85, AGFI is higher than 0.80, and SRMR and RMSEA are less than 0.08, the model is a good fit (DeVellis and Thorne, 2021; Kline, 2016).

The items on the scale in this study were loaded under eight factors based on EFA analysis. The 36-item scale was administered to 150 math teachers, and 1st-order CFA was performed with the LISREL 8.8. The Chi-square test for goodness-of-fit in the CFA was first calculated to be $\chi^2 = 908.46$ ($df = 566$, $p < 0.001$), and the ratio of χ^2/df was found as 1.60, which was less than 3. There were correlations found between the measurement errors of items B16 and B17, B18 and B19, B20 and B21, B30 and B31, and B33 and B34. Following the necessary modifications based on the analysis results, a path diagram with 36 items consisting of this eight-factor structure was created (Figure 2). The standardized factor loadings of the model items ranged from 0.49 to 0.94, and the *t*-test results for items were significant ($p < 0.01$).

According to Table 6, the fit indices for the 1st-order CFA after structure modifications were found to be $\chi^2/df = 1.39$ ($\chi^2 = 782.01$, $df = 561$, $p < 0.001$), which was less than 3. This result showed the model had a good fit (Kline, 2016). Other fit indices for the final model were determined to be RMSEA = 0.051, NNFI = 0.97, IFI = 0.98, GFI = 0.77, CFI = 0.98, AGFI = 0.73, and SRMR = 0.062, respectively. Some fit indices (RMSEA, NNFI, IFI, and SRMR) indicated that the scale model had a good fit, whereas AGFI and GFI fit indices indicated that the model had an acceptable fit (DeVellis and Thorne, 2021; Kline, 2016). These findings confirm the scale's factor structure.

CFA	χ^2	*df*	χ^2/df	RMSEA	NNFI	IFI	GFI	CFI	AGFI	SRMR
1st-order after structure modification	782.01	561	1.39	0.051	0.97	0.98	0.77	0.98	0.73	0.062
2nd-order after structure modification	816.71	581	1.41	0.052	0.97	0.98	0.77	0.98	0.73	0.070

Table 6: Fit indices of the 1st- and 2nd- order CFA of the scale

A second-order CFA was also performed to determine the structural relationship between the scale and the sub-factors. The results in Table 7 indicated that the standardized factor loadings of the 1st-order latent variables in the model ranged from 0.21 to 0.86, and the *t*-test results were significant ($p < 0.01$). All of the fit indices for the 2nd-order after structure modification in Table 6 ($\chi^2 = 816.71$, $df = 581$, $\chi^2/df = 1.41$, RMSEA = 0.052, NNFI = 0.97, IFI = 0.98, GFI = 0.77, CFI = 0.98, AGFI = 0.73, and SRMR = 0.070) confirmed that the model had a good and acceptable fit (DeVellis and Thorne, 2021; Kline, 2016).

2nd-order Variable	1st-Order Variable	Factor loading	Error Variance	*t*-value	R^2
Grading	1. Social Behaviors	0.70	0.50	7.67	0.50
	2. Math Knowledge and Skills	0.86	0.25	9.16	0.74
	3. Effort	0.73	0.47	8.79	0.53
	4. Affective skills	0.80	0.36	7.64	0.64
	5. External benchmarks	0.21	1.02	2.13	0.04
	6. Homework performance	0.74	0.46	9.34	0.54
	7. Follow-up test results	0.58	0.47	5.34	0.53
	8. Academic exam results	0.62	0.60	6.76	0.39

Table 7: Factor loadings in the 2nd-order CFA for the scale model

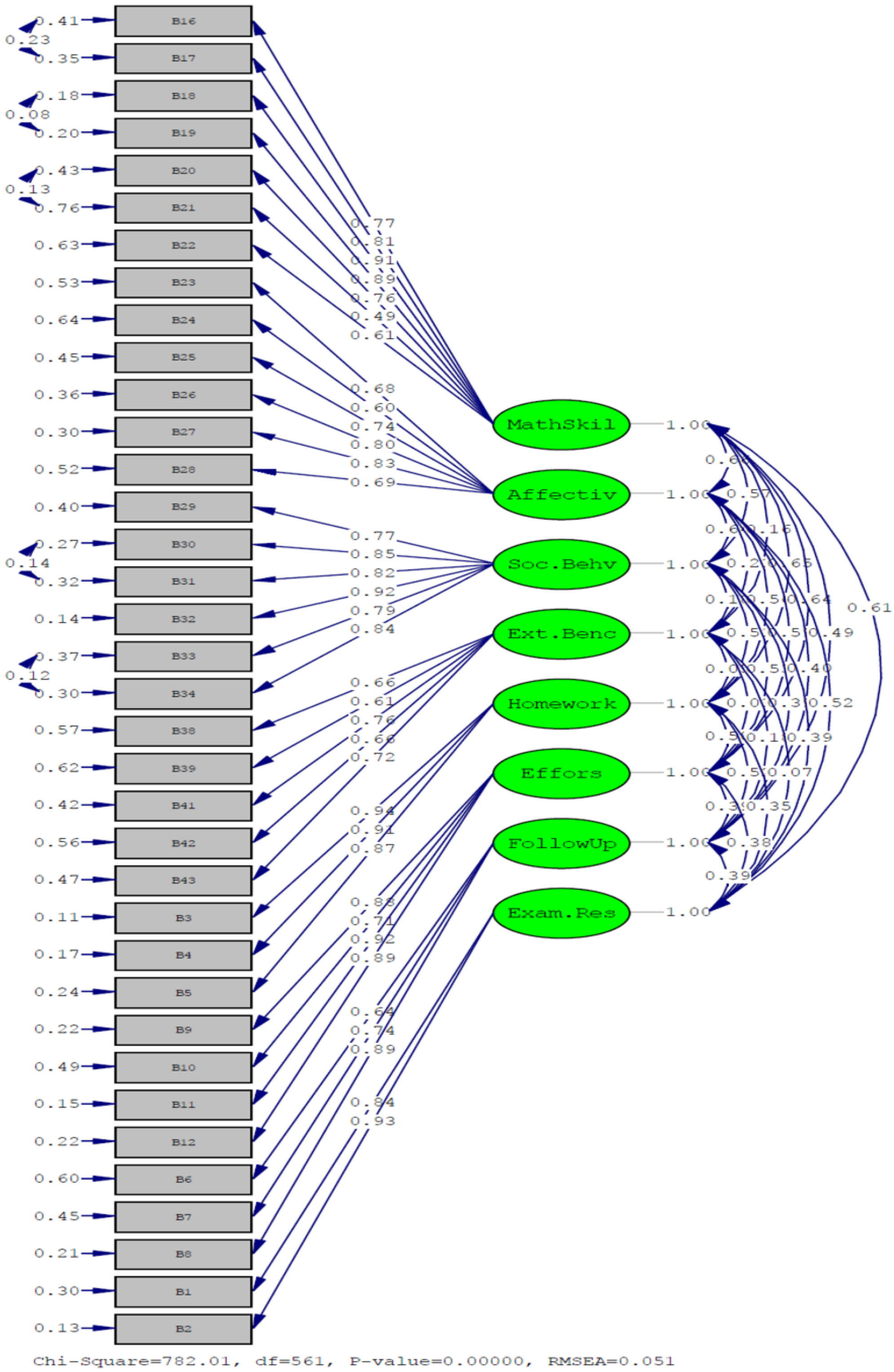

Figure 2: Factor loadings in the 1st-order CFA (standardized factor loadings)

Findings Relating to the Reliability Test

The reliability coefficient for scales measuring affective characteristics is recommended to be higher than 0.70 (Field, 2005). The overall Cronbach's alpha (α) coefficient of the scale was calculated as $\alpha = 0.92$ in this study. The Cronbach's alpha (α) coefficients for sub-factors were found to be 0.93 (Social behaviors), 0.89 (Math knowledge and skills), 0.90 (Effort), 0.86 (Affective skills), 0.85 (External benchmarks), 0.83 (Homework), 0.80 (Follow-up test results), and 0.83 (Academic exam results), respectively (Table 4).

DISCUSSION AND CONCLUSION

The study aimed to develop a scale to determine the factors that mathematics teachers consider when grading their students' in-class math performance. Item-total correlation and a *t*-test for 27% of the upper and lower groups were calculated for item analysis. It is suggested that the item-total correlation was greater than 0.30, and any items with a score between 0.20 and 0.30 were modified and improved (DeVellis and Thorne, 2021). In this study, three items were excluded because the item-total correlation of one item was less than 0.20, and two items were between 0.20 and 0.30. The item-total correlation for the remaining items ranged from 0.36 to 0.64, and the *t*-test results for the upper and lower groups were significant ($p < 0.01$). These findings indicated that the item discrimination of the scale was within an acceptable level (Field, 2005).

In the first stage of this study, EFA was performed on the scale's construct validity using principal component analysis and oblique rotation method (Oblimin), where factors are allowed to correlate. In this study, the Kaiser-Meyer-Olkin (KMO) value was 0.867, and Bartlett's sphericity test chi-square value was 4899.104 ($p < 0.001$). These findings demonstrated the feasibility of EFA (Tabachnick and Fidel, 2007). The first-factor analysis results revealed a nine-factor scale with eigenvalues greater than one. Four items were removed from the analysis since their factor loadings were less than.40 or had cross-loading. After performing the factor analysis for the remaining items, it was found that the scale composed of eight factors with 36 items, with factor loadings ranging from 0.50 to 0.92, accounting for 70.02% of the total variance. In social sciences, total variances ranging from 40% to 60% are considered sufficient on multi-factorial scales (Tabachnick and Fidell, 2007). These findings demonstrated that the scale's sub-factors and related items were sufficient to explain the factors influencing math teachers' grading practices. The correlation between the sub-factors and the overall scale was moderately significant, ranging from 0.34 to 0.71.

In this study, the items in the sub-factors were examined, and the sub-factors were labeled as 'mathematical knowledge and skills', 'external benchmarks', 'social behaviors', 'homework performance', 'effort', 'affective skills', 'academic exam success', and 'follow-up test results', respectively. Some of the grading factors that emerged in this study are consistent with previous studies (e.g., Cheng and Sun, 2015; Cizek et al., 1995; Duncan and Noonan, 2007; Frary et al., 1993; McMillan, 2001; McMillan et al., 2002). McMillan (2001) identified six factors for grading practice (i.e., academic enabling behaviors, use of external benchmarks, academic achievement, use of extra credit and borderline cases, use of graded homework, and use of homework not graded). Duncan and Noonan (2007) determined two factors for grading practices: 'academic enabling behaviors' (e.g., ability level, student effort, participation in the lesson, development, study habits, and negative behavior in the classroom) and 'use of external benchmarks' (e.g., the school's informal success policy, grade distribution of other teachers, and student success in previous years). Chen and Sun (2015) revealed three factors for grading practice: the 'norm/objective-references factor' (e.g., learning goal, other students' grades), the 'efforts factor' (e.g., homework, effort, improvement, work habit), and the 'performance factor' (e.g., academic, non-academic performance, academic ability). Previous research has shown that certain non-academic factors are effective in determining students' grades, and these factors differ depending on grade level and country. In contrast to previous research, the academic success factor for grading in this study emerged as two different sub-factors (follow-up test results and academic exam results). This finding could be explained by the fact that academic exams are an official requirement in the Turkish education system, whereas the follow-up test results are optional.

CFA is another method for determining scale construct validity (Field, 2005). First- and second-order CFA was performed on 150 math teachers who were not part of the EFA group in this study. The results of the 1st-order CFA after structure modifications indicated that the standardized factor loadings for each item ranged from 0.51 to 0.94, and the *t*-test results were significant. The model fit indices were found to be at a good and acceptable level (Kline, 2016; Tabachnick and Fidell, 2007). Moreover, the results of second-order CFA showed that the model had good fit indices. These findings confirmed the scale's sub-factors structure.

The reliability coefficient for the scales should be greater than 0.70 (DeVellis and Thorne, 2021). In this study, Cronbach's alpha coefficient for the overall scale was calculated to be 0.92, and Cronbach's alpha coefficients for the sub-factors ranged from 0.80 to 0.93. The reliability analysis revealed that the overall reliability of the scale and its sub-factors was satisfactory.

The findings of this study revealed that the scale, which consists of 36 items and eight sub-factors, was a valid and reliable instrument for identifying the factors influencing mathematics teachers' grading practices regarding students' in-class performance. As a result, it can be stated that the scale will contribute to future research in this field.

ACKNOWLEDGEMENT

This study was derived from the second author's Master's thesis.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

Acar-Erdol, T. and Yıldızlı, H. (2018) 'Classroom assessment practices of teachers in Turkey', *International Journal of Instruction*, Vol. 11, No. 3, pp. 587–602. https://doi.org/10.12973/iji.2018.11340a

Adams, T. L. and Hsu, J.-W. Y. (1998) 'Classroom assessment: Teachers' conceptions and practices in mathematics', *School Science and Mathematics*, Vol. 98, No. 4, pp. 174–180. https://doi.org/10.1111/j.1949-8594.1998.tb17413.x

Akcadag, T. (2010) 'Öğretmenlerin ilköğretim programındaki yöntem teknik, ölçme ve değerlendirme konularına ilişkin eğitim ihtiyaçları [The training needs of teachers concerning methods & techniques and assessment & evaluation in the primary school curriculum]', *Bilig*, Vol. 53, pp. 29–50.

Birgin, O. and Baki, A. (2009) 'An investigation of primary school teachers' proficiency perceptions about measurement and assessment methods: The case of Turkey', *Procedia Social and Behavioral Sciences*, Vol. 1, No. 1, pp. 681–685. https://doi.org/10.1016/j.sbspro.2009.01.119

Birgin, O. (2011) 'Pre-service mathematics teachers' views on the use of portfolios in their education as an alternative assessment method', *Educational Research and Reviews*, Vol. 6, No. 11, pp. 710–721.

Birgin, O. and Baki, A. (2012) 'Investigation of the purposes of the measurement and assessment practice of primary school teachers within the context of the new mathematics curriculum', *Education and Science*, Vol. 37, No. 165, pp. 152-167.

Bennett, R. E. (2011) 'Formative assessment: A critical review', *Assessment in Education Principles, Policy and Practice*, Vol. 18, No. 1, pp. 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P. and Wiliam, D. (1998) 'Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, Vol. 5, No. 1, pp. 7–74. https://doi.org/10.1080/0969595980050102

Brookhart, S. M. (1993) 'Teachers' grading practices: Meaning and values', *Journal of Educational Measurement*, Vol. 30, No. 2, pp. 123–142. https://doi.org/10.1111/j.1745-3984.1993.tb01070.x

Brookhart, S. M. (1994) 'Teacher's grading: Practice and theory', *Applied Measurement in Education,* Vol. 7, No. 4, pp. 279–301. https://doi.org/10.1207/s15324818ame0704_2

Bursuck, W., Polloway, E. A., Plante, L., Epstein, M. H., Jayanthi, M., and McConeghy, J. (1996) 'Report card grading and adaptations: A national survey of classroom practices', *Exceptional Children*, Vol. 62, No. 4, pp. 301–318. https://doi.org/10.1177/001440299606200402

Cheng, L. and Sun, Y. (2015) 'Teachers' grading decision making: Multiple influencing factors and methods', *Language Assessment Quarterly*, Vol. 12, No. 2, pp. 213–233. https://doi.org/10.1080/15434303.2015.1010726

Cheng, L. and Wang, X. (2007) 'Grading, feedback, and reporting in ESL/EFL classrooms', *Language Assessment Quarterly*, Vol. 4, No. 1, pp. 85–107. https://doi.org/10.1080/15434300701348409

Cizek, G. J., Fitzgerald, S. M. and Rachor, R. A. (1995) 'Teachers' assessment practices: preparation, isolation, and the kitchen sink', *Educational Assessment*, Vol. 3, No. 2, pp. 159–179. https://doi.org/10.1207/s15326977ea0302_3

Cross, L. H. and Frary, R. B. (1999) 'Hodgepodge grading: Endorsed by students and teachers alike', *Applied Measurement in Education,* Vol. 12, No. 1, pp. 53–92. https://doi.org/10.1207/s15324818ame1201_4

Demir, M., Tananis, C. A. and Basbogaoglu, U. (2018) 'Comparative investigation of alternative assessment methods used in Turkey and United States elementary 4th grade mathematics curriculum', *International Journal of Educational Administration and Policy Studies*, Vol. 10, No. 7, pp. 72–82. https://doi.org/10.5897/IJEAPS2018.0561

Demir, M., Tanasis, C. A. and Trahan, K. W. (2019) 'Evaluation of alternative assessment methods used in elementary schools', *Education and Science*, Vol. 44, No. 197, pp. 223–238. http://dx.doi.org/10.15390/EB.2019.6272

DeVellis, R. F. and Thorpe, C. T. (2021) *Scale development: Theory and applications*, London: Sage Publications.

Duncan, C. R. and Noonan. B. (2007) 'Factors affecting teachers' grading and assessment practices', *Alberta Journal of Educational Research*, Vol. 53, No. 1, pp. 1–21. https://doi.org/10.11575/ajer.v53i1.55195

Field, A. (2005) *Discovering statistic using SPSS for Windows*, 2nd edition, London: Sage Publication.

Frary, R. B., Cross, L. H. and Weber, L. J. (1993) 'Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement', *Educational Measurement: Issues and Practice*, Vol. 12, No. 3, pp. 23–30. https://doi.org/10.1111/j.1745-3992.1993.tb00539.x

Gullickson, A. R. (1985) 'Student evaluation techniques and their relationship to grade and curriculum', *Journal of Educational Research*, Vol. 79, No. 2, pp. 96–100. https://doi.org/10.1080/00220671.1985.10885657

Guskey, T. (2011) 'Five obstacles to grading reform', *Educational Leadership*, Vol. 69, No. 3, pp. 17–21.

Kline, R. B. (2016) *Principles and practice of structural equation modeling,* 4th edition, New York: The Guilford Press.

McMillan, J. H. (2001) 'Secondary teachers' classroom assessment and grading practices', *Educational Measurement: Issues and Practice,* Vol. 20, No. 1, pp. 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

McMillan, J. H., Myran, S. and Workman, D. (2002) 'Elementary teachers' classroom assessment and grading practices', *The Journal of Educational Research*, Vol. 95, No. 4, pp. 203–213. https://doi.org/10.1080/00220670209596593

Ministry of National Education (MoNE) (2014) *Milli Eğitim Bakanlığı okul öncesi eğitim ve ilköğretim kurumları yönetmeliği* [*Ministry of National Education regulation on pre-school education and primary education institutions*], Ankara: MoNE

Ministry of National Education (MoNE) (2018) *Matematik dersi öğretim programı* [*Primary school mathematics curriculum (1st-8th grade)*], Ankara: MoNE

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

Özenç, M. and Çakir, M. (2015) 'Exploring primary school teachers' competencies of alternative assessment and evaluation', *Elementary Education Online*, Vol. 14, No. 3, pp. 914–933. https://doi.org/10.17051/io.2015.22900

Randall, J. and Engelhard, G. (2010) 'Examining the grading practices of teachers', *Teaching and Teacher Education*, Vol. 26, No. 7, pp. 1372–1380. https://doi.org/10.1016/j.tate.2010.03.008

Serin, G. (2015) 'Alternative assessment practices of a classroom teacher: alignment with reform-based science curriculum' *Eurasia Journal of Mathematics, Science and Technology Education*, Vol. 11, No. 2, pp. 277–297. https://doi.org/10.12973/eurasia.2015.1330a

Shepard, L. A. (2000) 'The role of assessment in a learning culture', *Educational Researcher,* Vol. 29, No. 7, pp. 4–14. https://doi.org/10.3102/0013189X029007004

Stiggins, R. J. (1997) *Student-centered classroom assessment*, 2nd edition, Old Tappan, NJ: Prentice-Hall.

Sun, Y. and Cheng, L. (2014) 'Teachers' grading practices: Meaning and values assigned', *Assessment in Education: Principles, Policy and Practice*, Vol. 21, No. 3, pp. 326–343. https://doi.org/10.1080/0969594X.2013.768207

Suurtamm, C., Koch, M., and Arden, A. (2010) 'Teachers' assessment practices in mathematics: Classrooms in the context of reform', *Assessment in Education: Principles, Policy and Practice*, Vol. 17, No. 4, pp. 399–417. http://dx.doi.org/10.1080/0969594X.2010.497469

Tabachnick, B. G. and Fidell, L. S. (2007) *Using multivariate statistics*, 5th edition, Boston: Pearson/Allyn and Bacon.

Veldhuis, M., Van Den Heuvel-Panhuizen, M., Vermeulen, J. A. and Eggen, T. J. H. M. (2013) 'Teachers' use of classroom assessment in primary school mathematics education in the Netherlands', Cadmo, Vol. 21, No. 2, pp. 35–53. http://dx.doi.org/10.3280/CAD2013-002004

Wiliam, D. (2011) 'What is assessment for learning?', *Studies in Educational Evaluation*, Vol. 37, No. 1, pp. 3–14. http://dx.doi.org/10.1016/j.stueduc.2011.03.001

Yıldızlı, H. (2020) 'Classroom assessment practices and student goal orientations in mathematics classes', *Journal of Qualitative Research in Education*, Vol. 8, No. 1, pp. 294–323. https://doi.org/10.14689/issn.2148-2624.1.8c.1s.13m

Zhang, Z. and Burry-Stock, J. A. (2003) 'Classroom assessment practices and teachers' self-perceived assessment skills', *Applied Measurement in Education*, Vol. 16, No. 4, pp. 323–342. https://doi.org/10.1207/S15324818AME1604_4

Zhao, X., den Heuvel-Panhuizen, V., and Veldhuis, M. (2016) 'Teachers' use of classroom assessment techniques in primary mathematics education-an explorative study with six Chinese teachers', *International Journal of STEM Education*, Vol. 3, No. 19, pp. 1–18. https://doi.org/10.1186/s40594-016-0051-2

APPENDIX

SCALE ITEMS

To what extent do you consider the following characteristics or factors in determining a student's participation and performance in math course activities? (1 = not at all, 5 = completely)
Factor 1: *Mathematics Knowledge and Skills* *(7 items, Cronbach's alpha = 0.89)*
Item 7 (B20). Matematiksel dil ve sembolleri kullanma becerisini (Ability to use mathematical language and symbols)
Item 8 (B18). Problem çözmede öğrendiği bilgileri kullanmasını (Using the learned knowledge to solve problems
Item 9 (B17). Matematiksel bilgileri kavramasını (Understand mathematical knowledge)
Item 10 (B21). Sahip olduğu matematik yetenek düzeyini (Mathematics ability)
Item 11 (B19). Muhakeme etme ve akıl yürütme becerisini (Reasoning skills)
Item 12 (B16). Matematiksel kural, formül ve bilgileri hatırlamasını (Recall mathematical rules, formulas and information)
Item 13 (B22). Derslerde gösterilen pratik zeka göstergeleri (Practical intelligence indicators shown in the lessons)
Factor 2: *External Benchmarks* *(5 items, Cronbach's alpha = 0.85)*
Item 24 (B38). Diğer derslerdeki başarısını (Achievement level in other courses)
Item 25 (B43). Diğer okulların not/başarı politikasını (Other school' grade/achievement policy)
Item 26 (B39). Diğer öğrencilere göre başarı durumunu (Achievement status compared to other students)
Item 27 (B41). Okul idaresinin başarı politikasını (The achievement policy of the school administration)
Item 28 (B42). Sınıfın matematik başarı düzeyini (Mathematics achievement level of the class)
Factor 3: *Social Behaviors* *(6 items, Cronbach's alpha = 0.93)*
Item 1 (B32). Sınıf içindeki sosyal ve ahlaki davranışları (Social and moral behavior in the classroom)
Item 2 (B31). Diğer öğretmenler yönelik saygısı ve iletişimini (Respect and communication with other teachers)
Item 3 (B34). Sınıf içi ve okul kurallarına uyma (To comply with classroom and school rules)
Item 4 (B30). Ders öğretmenine yönelik saygısı ve iletişimi (Respect and communication towards the course teacher)
Item 5 (B29). Sınıf arkadaşlarına yönelik saygısı ve iletişimi (Respect and communication towards classmates)
Item 6 (B33). Sınıf dışındaki sosyal ve ahlaki davranışları (Social and moral behavior outside the classroom)
Factor 4: *Homework Performance* *(3 items, Cronbach's alpha = 0.83)*
Item 29 (B4). Ev ödevlerinin niteliği ve kalitesini (The quality of homework)
Item 30 (B3). Ev ödevlerinin zamanında yapılmasını (Timely completion of homework)
Item 31 (B5). Ekstra yapılan araştırma ödevlerini (Extra research assignments)
Factor 5: *Effort* *(4 items, Cronbach's alpha = 0.90)*
Item 14 (B9). Derste göstermiş olduğu gayret ve çabasını (Effort in the lesson)
Item 15 (B12). Derslere katılım düzeyi ve sıklığını (Level and frequency of participation in classes)
Item 16 (B11). Süreç içinde gösterilen gelişim performansını (The improvement performance shown in the process)
Item 17 (B10). Sınıf içinde sorulara doğru cevap verme sıklığını (Frequency of answering questions correctly in class)
Factor 6: *Affective Skills* *(6 items, Cronbach's alpha = 0.86)*
Item 18 (B24). Dersi sevme (Loving the lesson)
Item 19 (B26). Matematik çalışmalarına gönüllü olmasını (Volunteering in mathematics studies)
Item 20 (B27). Derse yönelik tutum (Attitude towards the lesson)
Item 21 (B23). Derse karşı ilgisi ve önem verme düzeyi (Interest and giving importance to the lesson)
Item 22 (B25). Öğrenme istekliliği (Willingness to lesson)
Item 23 (B28). Derse yönelik öz-güvenini (Confidence in the lesson)
Factor 7: *Academic Exam Results* *(2 items, Cronbach's alpha = 0.83)*
Item 35 (B1). Yazılı sınavlardaki performansını (Performance in written exams)
Item 36 (B2). Derse ilişkin proje görev performansını (Project task performance related to the course)
Factor 8: *Follow-up Test Results* *(3 items, Cronbach's alpha = 0.80)*
Item 32 (B8). Ünite/İzleme test sonuçlarını (Unit test results)
Item 33 (B7). Deneme sınav sonuçlarını (Trial exam results)
Item 34 (B6). Yapılan ara sınav (quiz) sonuçlarını (The results of the pop quizzes)
Items removed from the scale:
B13. Derse devam-devamsızlık durumunu (Course attendance-absence status)
B14. Ders araç-gereçlerini düzenli olarak getirmesini (Bring the course equipment regularly)
B15. Verilen görev ve sorumluluklarını yerine getirmesini (To fulfill the assigned duties and responsibilities)
B35. Kişisel bakım, giyim ve temizliğini (Personal care, clothing and cleaning)
B36. Velinin statüsünü (Parent's status)
B37. Diğer öğretmenlerin not dağılımını (Distribution of grades of other teachers)
B40. Velinin öğretmen ile iletişim düzeyi (Parent's level of communication with the teacher)