International Journal of Psychology and Educational Studies, 2023, 10(2), 543-560

www.ijpes.com

International Journal of Psychology and Educational Studies

ISSN: 2148-9378

Teacher-based Evaluation of Students' Problem Solving Skills

İlknur ÖZPINAR[1], Selahattin ARSLAN [2]

[1] *Faculty of Education, Niğde Ömer Halisdemir University, Niğde, Türkiye* 0000-0002-3630-0991

[2] Fatih *Faculty of Education, Trabzon University, Trabzon, Türkiye* 0000-0001-8557-2507

ARTICLE INFO

Article History
Received 09.01.2023
Received in revised form 04.03.2023
Accepted 23.03.2023
Article Type: Research Article

ABSTRACT

The need for improving and evaluating students' cognitive skills, including problem-solving, has been highlighted in both the mathematics education literature and the curricula worldwide. The objective of this study was to assess the problem-solving skills of lower secondary school students using a teacher-based evaluation method, taking into account their grade level, gender, grade-point averages, and mathematics grades. To that end, the Problem-solving Skill Scale was administered to 39 mathematics teachers from 21 schools, for a total of 1010 sixth, seventh, and eighth grade students. Based on the findings, it can be inferred that the students possess an adequate level of problem-solving skills; however, their level of proficiency differed significantly according to their grade levels, gender, grade-point averages, and mathematics grades. This identification of the students' problem-solving skills is expected to guide them towards activities and resources that align with their individual abilities.

Keywords:
Gender, grade, grade in mathematics and grade-point averages, problem-solving skill, secondary school students

1. Introduction

In order to lead a successful life in a society, students should have not only suitable information, but also different types of skills in a variety of areas that will enable them to succeed in today's rapidly changing environments. Problem-solving (PS), with its increasing complexity in different areas of life, is a vital competency for the 21st century (English & Gainsburg, 2015) and it is mainly supported by mathematical PS.

A survey of the history of mathematics reveals that the underlying idea behind its emergence and development is the desire to solve daily life problems. Therefore, PS skills could be seen as critical to humanity's survival. As no one can predict what problems they will face in the future, modern schools educate individuals in a way to solve problems on their own (Altun, 2004). This is one of the reasons why PS skills are noteworthy for students (Lester, 1994).

1.1. Problem and Problem-solving

Different scholars (English & Gainsburg, 2015; Polya, 1957; Schoenfeld, 1985) have similarly characterised a 'problem' as a situation that confuses an individual who has to solve (get rid of) it using previously acquired experiences and knowledge, but the solution path is not directly open (National Council of Teachers of Mathematics [NCTM], 2000; Schoenfeld, 1985). Due to its significance, researchers and theorists developed various notions and methods pertaining to the process of PS. As a result, a variety of PS models have been proposed, including information processing, cognitive science, constructivism, and models describing PS (e.g.

[1]Corresponding author's address: Nigde Omer Halisdemir University-Faculty of Education -Mathematics and Science Education Department /Niğde /Türkiye
e-mail: ilknurozpinar@gmail.com

Citation: Özpınar, İ. & Arslan, S. (2023). Teacher-based evaluation of students' problem solving skills. *International Journal of Psychology and Educational Studies*, *10*(2), 543-560. https://dx.doi.org/10.52380/ijpes.2023.10.2.1160

Anderson, Lee, & Fincham, 2014; Polya, 1957; Schoenfeld, 1985). All of these models involve Polya's (1957) four steps, and mathematical PS theories have emphasised the importance of heuristics (Wilson, Fermandez & Hadaway, 1993). According to Polya (1957), the PS process involves four steps: comprehending the problem, formulating a plan for the solution, executing the plan, and assessing its effectiveness. Another important step in this process, problem posing, is revealed and added by Gonzales (1994) who claims that it is one of the most significant and critical ways to acquire a deeper comprehension of the PS process. In this step, the student amends the problem just solved to obtain a similar one.

1.2. Mathematical Problem-solving

PS is considered extremely significant for mathematics learning since it is the way of thinking used in both learning and doing mathematics (NCTM, 2000; Pimta, Tayraukham, & Nuangchalerm, 2009). PS is defined, and its importance is highlighted in the literature. Lesh and Zawojewski (2007) defined mathematical PS skill as *"the process of interpreting a situation mathematically, which usually involves several iterative cycles of expressing, testing, and revising mathematical interpretation and of sorting out, integrating, modifying, revising or refining clusters of mathematical concepts from various topics within and beyond mathematics."* (p. 782). Polya's theory characterises mathematical PS as a process that includes dynamic and active engagement (Wilson, Fernandez & Hadaway, 1993). It is the ability to perform a comprehensive thinking process (for example, a mathematical algorithm and strategy) in line with the knowledge or skills gained for a particular situation by understanding the mathematical concepts, rules, and principles that are already known. Students' mathematical understanding develops and deepens in the PS process (Han & Kim, 2020; NCTM, 2000). For example, NCTM stated that *"By learning problem-solving in mathematics, students should acquire ways of thinking, habits of persistence and curiosity, and confidence in unfamiliar situations."* (NCTM, 2000, p. 52). PS not only improves students' ability to think but also helps them to acquire the basic skills that they need to overcome daily life problems and develop higher-level skills (Karataş, 2008; Pimta, Tayruakham & Nuangchalerm, 2009). Thus, the core focus of the mathematics curriculum both in Türkiye and in numerous other countries (such as the USA, England, Canada, and Singapore) is on this skill. The Turkish mathematics curriculum for secondary schools emphasises PS skills and expresses that *"One of the main goals of mathematics education is to develop students' problem-solving skills."* (Republic of Turkey Ministry of National Education [RTMNE], 2013**:** I). According to Wilson, Fernandez, and Hadaway (1993) there are multiple reasons why PS skills are prioritised in mathematics, and it is asserted that "*[it] is so essential to understanding mathematics and appreciating mathematics that it must be an instructional goal*." (p. 66)

1.3. Measurement and Evaluation of Problem-Solving

With the growing emphasis on PS in mathematics classrooms, there has been an increased significance placed on both the assessment and teaching of PS skills. Therefore, measuring and evaluating students' skills throughout the education process is a part of mathematics education (Wilson, Fernandez & Hadaway, 1993). NCTM (1989) highlighted the significance of the measurement and evaluation of PS skills .

PS is a mental skill and therefore cannot be observed directly. However, to regulate educational activities and measure and evaluate the resulting products, it is crucial to identify the observable behaviours that make up the PS process (Erden, 1986). Reviewing the literature reveals that PS behaviours skills are frequently highlighted as crucial in evaluating PS (e.g. Kilpatrick, 1985; Reys et al., 1998). Reys et al. (1998) stated that issues related to the indicators that students read the problem carefully, the use of strategies, the frequency of making mistakes, and asking for help should be given due consideration in the evaluation of PS skills. According to Schoenfeld (1985), individuals who are experts in PS have a deep knowledge of heuristic strategies necessary to move forward in unusual situations, evaluate the development of the solutions that have been tried or can be tried and their usefulness to the possible availableness. S(he) does not subscribe to the belief that all problems can be tackled in a standardised manner like in typical research and is therefore open to experimenting with other approaches. Muir, Beswick, and Williamson (2008) investigated the behaviours of expert, routine, and poor problem solvers more comprehensively than Schoenfeld (1985). According to Kilpatrick (1985), a student's success in PS depends on his/her behaviours throughout the PS process, and it can be enhanced through the development of behaviours crucial to PS. Therefore, a significant improvement can be achieved in PS when it becomes possible to identify what behaviours students cannot exhibit in this respect (incomplete behaviours or behaviours that have not yet been developed) and to overcome such deficiencies. That is why it is stressed in the literature that it is necessary to evaluate students' PS skills (NCTM, 2000). Furthermore, an evaluation of the PS process will contribute to teachers' ability to guide their classes. When teachers can observe what their students do throughout the PS process, they will be more likely to see what makes them unsuccessful or what poses a challenge to them and thus arrange those activities that can help them overcome such deficiencies (Arslan, 2007). However, it is rather complicated and difficult to evaluate the PS skill. The evaluation of students' PS abilities is contingent upon monitoring the approach they take to tackle the problem rather than the actual solution (Kyriakides & Gagatsis, 2003; Wilson, Fernandez, & Hadaway, 1993).

It is well known that effective assessment is essential to the progression of student learning. It is also common knowledge that standardised assessments using pencil and paper are the most widely used assessment instruments. However, many scholars argue that tests are not intended to be formative assessment tools and are based on the literature. Vendlinksi and Stevens (2002, p.3) state that "*many performance-based assessments suffer from validity, pedagogical, logistic, time and cost problems*." Therefore, other forms of evaluation are needed both in general and in PS as well. Such a fact has led researchers to inaugurate new forms of evaluation, including teacher-based assessment. Studies in the literature indicate that various variables (such as age, class, gender, and students' academic success) have an effect on students' mathematical PS success (Bunar, 2011; Güven & Çabakçor, 2013; Özgen, Aydın, Geçici & Bayram, 2017; Özsoy, 2005). In addition to the necessities of daily life, PS skills are required to be successful in the whole mathematics courses. Students with high PS skills are also expected to be successful in general mathematics course (Özsoy, 2005). In fact, the literature contains studies that explore the correlation between students' PS skills and their academic and mathematics course achievements, generally through an achievement test prepared by researchers (e.g. Bunar, 2011; Güven & Çabakçor, 2013; Özsoy, 2005). In these cases, general items were used to assess academic achievement in mathematics when it was necessary to evaluate students' mathematical PS skills. It is important to distinguish between mathematical PS skill and mathematical performance measured by tests in mathematics. Han and Kim (2020, p. 99) expresses that*"Whereas test items measuring mathematical performances are relatively more likely to emphasize students' mathematical knowledge and skills, mathematical problem solving competence is a concept that goes beyond simple mathematical knowledge and skills. Therefore, the measurement tool for mathematical problem solving competence must be different from tests for mathematics knowledge and skills."* However, when viewing PS as a process, emphasis should be placed on the process of evaluation, and more clearly, evaluations should be made by considering the experiences of the students instead of evaluating what they are doing as right or wrong (Karataş, 2008; Wilson, Fernandez, & Hadaway, 1993). Also, the literature indicates that researchers also take into consideration diverse factors, including gender and academic level, when examining PS abilities. (e.g. Gallagher et al., 2000; Nurhayanti, Riyadi & Usodo 2020; Zhu, 2007). In addition to these studies, there are some PS scale development studies and studies using these scales in the literature (e.g. Altun, 2019; Heppner & Peterson, 1982). However, it is seen that these scales and their evaluations are not related to mathematics. Various scale studies have been realized, such as the belief scale for mathematical PS (Kloosterman & Stage, 1992), attitude scale for PS (Çanakçı & Özdemir, 2011) reflective thinking skills scale for PS (Kızılkaya & Aşkar, 2009). However, as these studies are not specifically designed to assess PS skills, it seems that the development of measurement tools for mathematical PS skills is restricted. Numerous studies assessing students' mathematical PS abilities have neglected to verify the reliability of their measuring instruments. Additionally, many of the instruments utilised failed to effectively measure PS skills across all stages of the process (Han & Kim, 2020). Similarly, D'Zurilla, Maydeu-Olivaresb, and Kant (1998) stated that performance tests were utilised to measure PS skills in all empirical studies, but the interpretability of the results was difficult due to the questionable construct validity of the PS performance tests, which are not based on any particular theory or PS model (D'Zurilla, Maydeu-Olivares, & Kant, 1998). Research on the evaluation of mathematical PS skills has been conducted either via qualitative research methods or based on small samples (e.g. Maulyda, Hidayati, Rosyidah, & Nurmawanti, 2019). Most of these studies focused either on the methods used for measuring and evaluating students' PS skills (Işık & Kar, 2011; Nurkaeti, 2018) or on the identification of the behaviours students exhibit during the PS process (Erden, 1986; Muir, Beswick & Williamson, 2008; Peranginangin, & Surya, 2017; Rahmawati & Sulisworo, 2020). A review of the literature indicates that a portion of the research in mathematical PS has been directed towards determining PS strategies and individuals' thinking processes (Baraké, El-Rouadi, & Musharrafieh, 2015; Bunar, 2011; Gökkurt, Örnek, Hayat & Soylu, 2015; Lester, 1994; Schoenfeld, 1985). In these studies, it was revealed that students' PS skills did not meet the expected standards (Bozkurt & Karslıgil-Ergin, 2017; Rahmawati & Sulisworo, 2020). Upon examining these studies, it became apparent that the emergence of this situation is caused by the problems of students in understanding and associating mathematical concepts, a lack of logical reasoning, strategic knowledge, and problem posing. If students' PS processes are evaluated in this direction, more qualified information can be obtained about the students.

The objective of mathematics education in the 21st century has shifted from simply teaching individuals to apply the rules of mathematics to teaching them how to apply their knowledge to solve problems, transfer their skills, and develop strategies (Incebacak & Ersoy, 2016). Therefore, there is a consensus among scholars that the development of PS skills should be a key objective of instruction (Fülöp, 2021; Lester & Cai, 2016). Education programmes should establish assessment guidelines as well as specify how to evaluate students' basic knowledge, skills, mathematical thinking, and PS (Zhao, Van den Heuvel-Panhuizen, & Veldhuis, 2018). Improving the quality of teaching is dependent on the students ability to demonstrate the skills they having acquired. Therefore, it is crucial to identify and develop students' potential PS skills. Evaluating students' PS skills is essential to have information about students' mathematical knowledge and to obtaining information that has a clue that can guide education programmes (Karataş, 2002). Therefore, determining students' actual PS skill level is crucial, in other words, how much of PS skill they actually possess. In line with the specified points, incorporating such problems into the education system will become less challenging over time (Incebacak & Ersoy, 2016).

1.4. Why using Teacher-based Assessment instead of Skill-based Assessment?

Teacher-based assessment has been employed for many years and in many countries (Gutierrez & Vignoles, 2015). For example, Marcenaro-Gutierrez and Vignoles (2015, pp. 1-2) state that the Spanish educational system relies heavily on teacher-based assessment in the sense that *"teachers form their view of the student's achievement based on observation of the pupil in class, their classwork, and teacher records of student performance"* and this assessment is even considerably important in determining a pupil's transition from elementary to secondary education. Since test and teacher-based assessments measure different characteristics of the cognitive or non-cognitive skills of pupils, teacher-based assessment is used instead of skill-based assessment. As a result, it is possible that they have different concentrations. For example, teacher-based assessments value pupils more than test-based assessments while the latter measure way students' performance in a particular subject in a more accurate way than teacher-based assessments. In sum, both types of assessments have their advantages and disadvantages. For example, according to certain research, instructors have a tendency to underestimate or overgrade students' actual level of achievement, maybe due to prejudice towards or against particular types of students (Hay & Macdonald, 2008; Marcenaro-Gutierrez & Vignoles, 2015). However, other studies highlighted that a completely 'fair' test does not exist and there are biases in various methods of students' assessment, implying that test scores could not be seen as (the only) true tool measure of the students' genuine achievement and should therefore be cautiously referred to (Brennan, 1996; Rothstein 2005; Sackett, Borneman, & Connelly 2008; Chiu, 2001). This assumption suggests that any disparity between assessments made by teachers and those made by tests should not be attributed to biases held by teachers. (Marcenaro-Gutierrez & Vignoles, 2015). We contend with Marcenaro-Gutierrez and Vignoles (2015, p.3) who argue that *"it is important to investigate how non-test-based forms of assessment, such as teacher assessment, produce different results from test-based measures and how these differences may be related to pupils' characteristics."* On the other hand, although some studies found a discrepancy between assessments made by teachers and those made by tests, others reveal consistency between both types of assessments (for example, Chui, 2016; Hoge & Coladarci, 1989; Reeves et al. 2001). In a study conducted on the determination of whether children's academic self-concepts come from teacher assessment or their own cognitive abilities, Chui (2016) combined eight data sets (teacher assessment, student cognitive ability test results, student self-concepts, and student backgrounds) relating to students aged 5, 7, and 11. The teachers are asked to rate students on a 5-point scale from 1 (well above average) to 5 (well below average). According to structural equation modelling, this study revealed that the internal/external frame of the reference model in general backed up teacher assessment as a pointer of achievement more than those with cognitive ability. Similarly, Hoge and Coladarci (1989) also aimed to reveal the match between teacher-based assessments of learners' achievement levels and objective measurement of their learning, and this study yielded a result indicating that teacher-based assessment has high level validity. Zhao, Van den Heuvel-Panhuizen, and Veldhuis (2018) stated that teachers typically employ observational assessment methods such as inquiry and observation for formative purposes and that they rely on tool-based methods. It is imperative to diagnose the cognitive changes occurring in students if they are to succeed in learning mathematics. Also, teachers state that it is difficult to prepare and incorporate PS activities in the process of evaluating students' PS skills, to devote enough time to practise, and therefore to measure and test PS skills (Wilson, Fernandez, & Hadaway, 1993). Therefore, it is seen that students' PS skills are not limited to PS activities, but instruments like scales are needed to measure this skill by observing students' behaviour for a long time with various tools in all activities carried out in the mathematics teaching process (Özpınar, 2012). In this context, it is significant for mathematics education in general and PS education in particular to carry out studies on the evaluation of students' PS skills by keeping students under observation over the long term on larger samples and with reference to various variables. In addition, there are no studies in the literature in which teachers evaluate their students' PS skills through scales by observing their behaviours for a long time. Indeed, Lester (2013) emphasises the importance of making extensive and systematic observations of classrooms for the PS process. Along with measuring the student's PS skills through this scale, by observing what (s)he has done during the process, his/her failures or deficiencies can be identified more easily, and efforts can be made to eliminate these deficiencies. In this way, a significant contribution will be made to teachers in guiding their lessons, and students will be informed and guided in line with their skills. The main objective of this study was to investigate the PS skills of lower secondary school students and analyse these skills based on several variables, including grade, gender, grade-point averages (GPA), and grades in mathematics (GIM).

2. Methodology

2.1. Research Model

The present study utilises a descriptive survey method, which seeks to depict a past or current situation in its actuality. The object, individual, or event being investigated is described in its own conditions (Karasar, 2005). This model is suitable for research (as in this study) that intends to "*gather data at a particular point in time with the intention of describing the nature of existing conditions, or identifying standards against which existing conditions can be compared, or determining the relationships that exist between specific events*" (Cohen, Manion & Morrison, 2000, p. 169).

2.2. Research Sample

The research group for the present study was comprised of 39 volunteer mathematics teachers (23 females and 13 males) who worked at 21 different schools in a big city in Turkey and 1010 students of theirs. There are 4 teachers with 1-5 years of professional experience, 12 with 6-10 years, 11 with 11-15 years, 8 with 16-20 years, and 4 with more than 20 years of professional experience.

Table 1. *Distribution of the Students according to Gender and Grade*

Gender→ / Grade Level ↓	Female (%)	Male (%)	Total
6	121 (12)	117 (11.6)	238
7	228 (22.6)	191 (18.9)	419
8	187 (18.5)	166 (16.4)	353
Total	536 (53.1)	474 (46.9)	1010

Table 1 presents the distribution of the students according to gender and grade. GIM and GPA scores of the students were examined in five groups, as seen in Table 2.

Table 2. *Students' GPA and GIM*

Level of Achievement	GIM		GPA	
	f	%	f	%
1-Very weak (0-44)	119	11.8	31	3.1
2-Weak (45-54)	191	18.9	122	12.1
3-Moderate (55-69)	216	21.4	225	22.3
4-Good (70-84)	226	22.4	331	32.8
5-Very Good (85-100)	258	25.5	301	29.8
Total	1010	100	1010	100

2.3. Data Collection Tools and Data Analysis

The 18-item Problem-solving Skill Scale (PSSS) developed by authors (Özpınar, 2012; Özpınar & Arslan, 2016) was used to evaluate the behaviours of students against the problems in the classroom environment by their teachers. The factor structure of PSSS was analysed by exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). As a result of EFA, three dimensions emerged. The first dimension obtained as a result of the analysis is the "Comprehension" consisting of 4 items. Some items in this dimension are "*S(he) rephrases the problem in his/her own words*", "*S(he) detect missing or surplus information.*", "*S(he) understands what is needed to find to solve the problem.*" The Cronbanch Alpha internal consistency coefficient for this dimension was 0.642. According to the results of EFA, the second dimension (Implementation) of the PSSS consists of 6 items related to researching the problem solution and carrying out the necessary actions, such as *"S(he) selects appropriate strategies to solve the problem."* and *"S(he) discusses with their peers the ways leading to the solution."* The Cronbach Alpha internal consistency coefficient was 0.780. The third dimension of the scale was "Evaluation", which includes expressions such as *"S(he) makes necessary arrangements by checking what (s)he do during the resolution process."* and *"S(he) makes generalizations for the solution of similar problems*." and consists of eight items. The Cronbach Alpha internal consistency coefficient was determined as 0.812 for this factor. Cronbach Alpha internal consistency coefficient for the whole scale is 0.875. Finally, a CFA was conducted in order to test the accuracy of the three-factor structure revealed by the EFA in the study. The chi-square value calculated for the model-data fit was found to be significant with CFA, which was conducted to examine to what extent the model of PSSS consisting of three factors complied with the collected data ($\chi2 = 321.45$, sd = 132, $p < .01$). Some of the goodness-of-fit indices calculated with the same analysis are as follows: ($\chi2$ / df) = 2.435, RMSEA = 0.064, RMR = 0.030, GFI = 0.91, AGFI = 0.88, NFI = 0.93, NNFI = 0.95, CFI = 0.96. With the first modification (between item no 18 and item no 17), the chi-square value calculated for model-data fit with CFA of the model consisting of three factors was found to be significant ($\chi2 = 303.30$, sd = 131, $p < .01$). Some of the goodness-of-fit indices in this respect were as follows: ($\chi2$ / df) = 2.315, RMSEA = 0.062, RMR = 0.028, GFI = 0.91, AGFI = 0.88, NFI = 0.94, NNFI = 0.96, CFI = 0.96. With the second modification (between item no 17 and item no 15), the chi-square value was significant for the model-data fit, which was identified through the CFA ($\chi2= 288.56$, sd=130, $p<.01$). Some of the calculated goodness-of-fit indices were: ($\chi2$ / df) = 2.219, RMSEA = 0.059, RMR = 0.027, GFI = 0.92, AGFI = 0.89, NFI = 0.94, NNFI = 0.96, CFI = 0.97.

The Personal Information section, which includes students' gender, grade, GIM, and GPA was added to the PSSS. In the determination of the academic achievement scores of the mathematics course (grades in mathematics), the report cards of the students were used. GIMs are obtained from the averages of written exams, in-class activity scores, and performance assignments.

Teachers responded on four-point scales ranging from 1 (weak) to 4 (very good). The raw scores ranged from 18 to 72. The PS skills of the students can be evaluated on the basis of either their raw scores or the mean of their raw scores. For instance, a student who gets a score of 56 on the scale has the following mean score: 56/18=3.1 (out of four). This study utilised a method to determine the width of the class interval by dividing the data range by the selected number of classes (Kan, 2009; Tekin, 1993). Using this approach, the range was calculated as the difference between the highest and lowest values (i.e 4-1=3), and the class interval is 3/4=0.75. Subsequently, the study employed these intervals to assess the findings obtained from data analysis: 3.26 to 4.00 'very good', 2.51 to 3.25 'sufficient', 1.76 to 2.50 'moderate' and 1.00 to 1.75 'weak'. A student's score (or mean score) can be calculated as described above. The greater a person's mean score is, the greater his/her mathematical PS skill is.

This research utilised both descriptive and inferential statistics. Descriptive statistics include arranging the data obtained from all units that make up the universe or sample of the research and finding the values to represent all the data. Frequency distribution, mean, standard deviation, correlation, etc. values are used in descriptive statistics. Inferential statistics, on the other hand, include the processes of estimating the characteristics of the universe through the data obtained from the samples assumed to represent the research universe. T-test, Analysis of Variance (ANOVA) etc. tests are commonly used as examples of the tests used in inferential statistics operations (Ekiz, 2009). The underlying assumption in multivariate data analysis is that the dependent variable follows a normal distribution and that the groups of data being analysed exhibit homogeneity. Therefore, before analysing, skewness and kurtosis values were calculated under the assumption of normality. As a result of the examination, when the Skewness-Kurtosis values of the variables were examined, it was found that the results were between -3 and +3, that is, the data showed a normal distribution (Tabachnick & Fidell, 2013). The data were also checked for outliers, but none were detected. In the study, Pearson Product Moments Correlation Coefficients were calculated in order to test the relationships between variables, and then a multi-directional hierarchical regression analysis was performed to determine to what extent GIM and GAP predicted PS skills. In accordance with the purpose of the study, the quantitative data were analysed through SPSS 25.0. They were interpreted on the basis of mean scores, standard deviation values, Independent Sample T-test, ANOVA, and a multiple comparison test (Post-Hoc) was used. T-test and ANOVA assumptions were checked. Besides the significance of the difference, the effect sizes were also examined. η2, also known as effect size, represents the proportion of total variance in the dependent variable that can be explained by the independent variable or factor. The effect size limit values are interpreted between .01 (small), .06 (medium), .14 (large). Independent samples T-test were calculated with the Cohen d statistic. When the Cohen's d values for the effect size are less than 0.2, the effect size is defined as small, when it is 0.5, it is medium, and when it is greater than 0.8, it is defined as large (Büyüköztürk, 2015; Cohen, 1988). As a result of the analysis, it was seen that the scores had a normal distribution, and homogeneity of the variance assumptions were provided. However, in order to explain the multiple linear relationships between independent variables and assess the appropriateness of the data for analysis, the tolerance and Variance Inflation Factors (VIF) values were examined (Field, 2013).

2.4. Procedure

The data collection process for this research was planned in detail for the purposes of the research. In practise, care has been taken to comply with this plan. Figure 1 and Figure 2 present the process followed during this study.

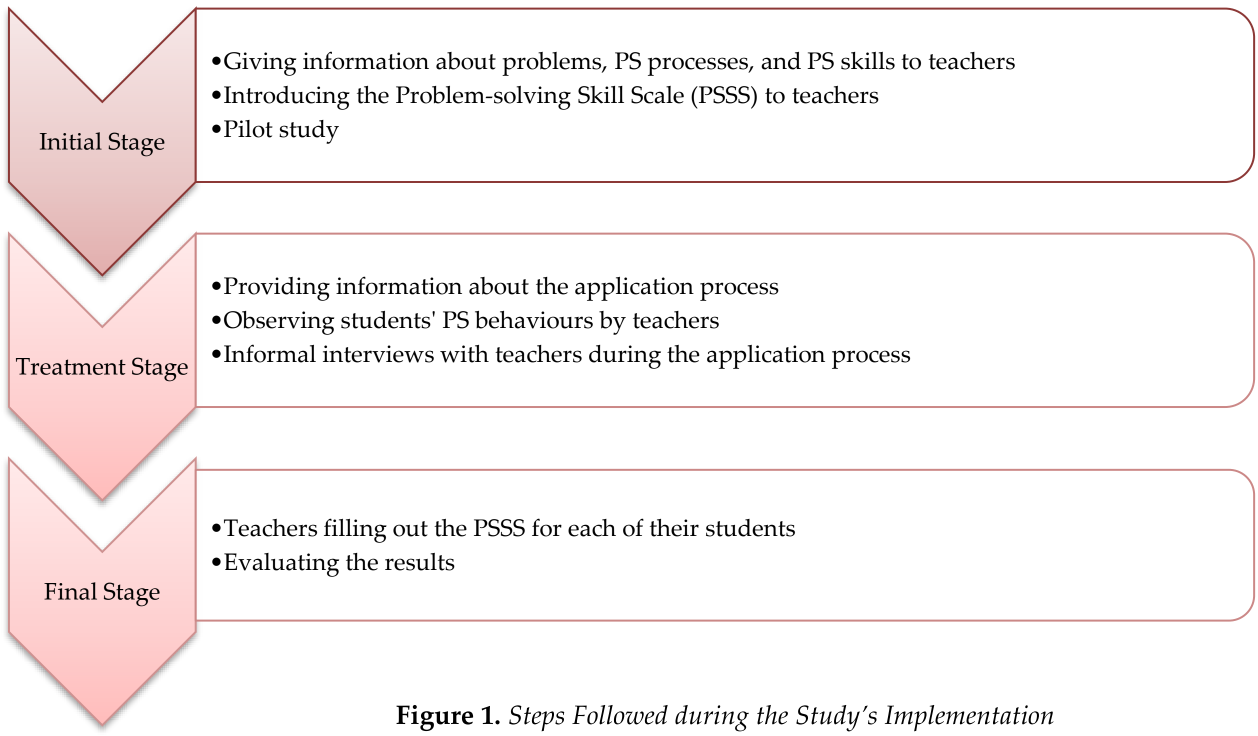

Figure 1. *Steps Followed during the Study's Implementation*

Prior to the study, the secondary school mathematics teachers working in a big city in Turkey were contacted, and 39 of them volunteered to participate in the study. They were given information about the purpose of the research, the application process, and the assessment scale. PSSS was examined together with the teachers, and questions related to the process were answered.

The study was implemented in both semesters of the academic year. In the first semester, the pilot application of the study was conducted. Teachers made their observations and practises in line with the items in PSSS, and they had experience with the process. At the end of this semester, a general evaluation was made of the applications with them and the grade levels and classes were determined for the next semester. Then, each teacher presented the number of classes and students s(he) was to evaluate.

Afterwards, they were provided with copies of the scale at the beginning of the spring term so that they could observe their students throughout the term and fill out the PSSS. Bi-weekly interviews were conducted with the teachers for the observation and evaluation of teachers. Also, the teachers administering the scale were always contactable during the data collection process. In this way, an attempt was made to guide them when necessary and to make sure that they were attentive enough. The researchers were allowed to conduct informal interviews with the teachers and the former was contactable whenever the latter needed any help, which suggested that the scale administration was meticulous.

Throughout the process, the teachers took personal notes concerning their students, which they took into account in addition to the individual photo files for the students while they were filling out the scale. It is deemed important to describe the evaluation process for the PS skills of students so that the results achieved can be meaningful. Before the implementation process, teachers created individual PS skill portfolios in order to require more reliable and dynamic data about students and reveal their skills effectively by observing the behavior of their students in detail. In these portfolios, besides the demographic characteristics of each student such as photograph, GPA, GIM; there are documents like students' activities (individual or group work), mathematical expressions which students used, interview notes, course notes, observation notes, written and oral tests, open questions, activity reports such as projects and presentations, problem solutions, exercises in and after class, questionnaire corresponding to each phase and criterion in PSSS. The teachers carefully and frequently entered the explanations regarding their observations in these files. While filling out the scale at the end of the semestre, evaluations were made based on student portfolios.

Figure 2. *Process of the Study*

2.5. Ethical

The data for the study were collected before 2020. Permission was obtained to use data collection tools applied within the scope of the research.

3. Findings

3.1. Examination of Problem-solving Skills According to Grade

Table 3 presents the descriptive statistics of students' PS skills according to the grade variable. ANOVA was performed in order to explore whether students' PS skills change according to their grades. Table 3 and Table 4 show the results of the analysis.

Table 3 *Students' PSSS Scores according to Grade*

Grade Level	n	$\overline{X}$ ±SS
6	238	2.55±0.83
7	419	2.58±0.85
8	353	2.42±0.86
Total	1010	2.52±0.84

The students' total mean score in PS skills was 2.52 (out of 4) whereas their mean scores by grade were 2.55, 2.58, and 2.42 respectively. This suggested that the students had 'sufficient' PS skills. However, when the PSSS mean scores are examined considering the grade, it is remarkable that while the PS levels of the sixth and seventh grades are 'sufficient', the eighth grade students are in the 'moderate' level. An overview of the participants' mean scores by their grade indicated significant discrepancies at the .05 level.

Table 4. *ANOVA Results of Students' PSSS Scores according to Their Grades*

Dimensions	Source of Varıance	Total of Squares	df	Mean of Squares	F	p	Post-hoc
Comprehension	Between Groups	4.301	2	2.150	2.890	0.056	Not
	Within Groups	749.255	1007	0.744			
	Total	753.556	1009				
Implementation	Between Groups	5.326	2	2.663	3.490	0.031*	7-8
	Within Groups	768.273	1007	0.763			
	Total	773.599	1009				
Evaluation	Between Groups	6.203	2	3.102	4.243	0.015*	7-6, 8-7
	Within Groups	736.152	1007	0.731			
	Total	742.355	1009				
Total Score	Between Groups	5.225	2	2.612	3.608	0.027*	7-8
	Within Groups	729.022	1007	0.724			
	Total	734,.47	1009				

*p< 0.05

Table 4 suggests that there was no significant difference in students' PS skills in the dimension of comprehension [$F_{(2, 1007)}$ = 2.890, p>.05]. However, a significant difference was seen in the dimensions of implementation [$F_{(2, 1007)}$ = 3.490, p<.05], evaluation [$F_{(2, 1007)}$ = 4.243, p<.05] and total score [$F_{(2, 1007)}$ = 3.608, p<.05].

Tukey's HSD test was conducted so as to determine the groups that led to the discrepancies. The test results indicated that the seventh grade students had higher PS skills when compared to the eighth grade students. A similar assessment was made for the dimensions of the scale as well (Table 5). For all three dimensions, the ranking was as follows in descending order: seventh grade students, sixth grade students, and eighth grade students.

Table 5. *The Results of Descriptive Statistics of Students' PS Skills and PS Dimensions*

Dimensions	Grade	n	$\overline{X}$ ±SS	F	p	Post-hoc	Effect size
Comprehension	6	238	2.57±0.86	2.890	0.056	Not	
	7	419	2.55±0.86				
	8	353	2.42±0.87				
Implementation	6	238	2.48±0.85	3.490	0.031*	7-8	0.07
	7	419	2.51±0.87				
	8	353	2.35±0.89				
Evaluation	6	238	2.58±0.83	4.243	0.015*	7-6, 8-7	0.08
	7	419	2.66±0.86				
	8	353	2.48±0.87				
Total	6	238	2.51±0.86	3.608	0.027*	7-8	0.07
	7	419	2.44±0.87				
	8	353	2.58±0.86				

After conducting a post-hoc analysis to test for significant differences in scores among groups, the Tukey multiple comparison test revealed that, for the implementation dimension, a significant difference existed between seventh and eighth grade students, with an effect size of .07. In the evaluation dimension, a significant difference was observed between seventh and sixth grade students, as well as between eighth and seventh grade students, with an effect size of .08. The mean score for sixth grade students was 2.51 ± 0.86, 2.44 ± 0.87 for seventh grade, and 2.58 ± 0.86 for eighth grade, according to the PSSS. The effect size obtained from the test was .07, indicating a medium effect size for this difference (Cohen, 1988).

It is seen in Table 5 that the students mostly have deficiencies in the implementation dimension, and they are partially better in the comprehension and evaluation dimensions. When the table was examined, it could be said that, according to the dimensions, although the students were sufficient in the comprehension and evaluation dimensions, they were at a moderate level in the implementation dimension. When the dimensions were considered, it was found that the lowest PSSS mean score was eighth grade in all three dimensions.

3.2. Examination of Problem-solving Skills According to Gender

This section of the study focused on the relationship between students' PS skills and gender by using an independent samples T-test. Table 6 presents the findings on whether the students' PS skills differed depending on gender.

Table 6. *T-test Results for the PSSS Scores according to Gender of Students*

Gender	n	$\overline{X}$ ±SS	Df	t	p	Effect size
Female	536	2.60±0.81	1008	3.08	0.002*	0.20
Male	474	2.43±0.89				

*p< 0.05

The students' mean scores in the PSSS differed significantly depending on gender [$t_{(1008)}$ = 3.08, p<.05]. The female students had higher mean scores (2.60±0.81) than the male students (2.43±0.89). The size of the effect shows that this difference has a small effect. This finding was also confirmed by the relationship between gender and the dimensions of the scale, which is presented in Table 7.

Table 7. *T-test Results of the Comparison of Means of Students' PSSS Dimension Scores according to Gender*

Dimensions	Female (n=536)	Male (n=474)	T	p	Effect size
	$\overline{X}$ ±SS	$\overline{X}$ ±SS			
Comprehension	2.60±0.83	2.42±0.89	3.38	0.001*	0.21
Implementation	2.51±0.84	2.37±0.91	2.57	0.010*	0.16
Evaluation	2.66±0.81	2.49±.0.90	2.23	0.001*	0.20

*p< 0.05

The results of the test indicated that the female students got higher mean scores than the male students in all three dimensions of the PSSS. The mean scores of the female and male students for the comprehension dimension was $\overline{X}$ = 2.60, $\overline{X}$ = 2.42, the implementation dimension was $\overline{X}$ = 2.51, $\overline{X}$ = 2.37 and the evaluation dimension was $\overline{X}$ = 2.66, $\overline{X}$ = 2.49 respectively. The analysis showed that there were significant differences in PS skills in the comprehension dimension [$t_{(1008)}$ = 3.38, p<.05], implementation [$t_{(1008)}$ = 2.57, p<.05] and evaluation [$t_{(1008)}$ = 2.23, p<.05] according to gender. As seen in Table 7, there were significant differences between the two groups in all three dimensions, suggesting gender is a significant factor in PS skill.

3.3. Examination of PSSS Mean Scores,Their Correlation and Regression with GPA and GIM

The correlation between the students' mean scores in the PSSS and their GPA and GIM was analysed through the Pearson product-moment correlation coefficient. The findings are presented in Table 8.

Table 8. *Correlation Between the Students' Mean Scores in the PSSS, Their GPA and GIM*

	GPA	GIM	PSSS
GPA	1.00		
GIM	0.838**	1.00	
PSSS	0.798**	0.866**	1.00

** p< 0.01

The results indicate a highly significant relationship between students' PS skills and their GPA, with a positive correlation coefficient of .798 (p<.01). This suggests that as students' PS skills increase, their GPA increases as well. A similar positive and highly significant relationship was observed between students' PS skills and their GIM, with a correlation coefficient of .866 (p<.01). This suggests that as students' PS skills improve, their GIM increases. To further explore the relationship between PS skills and academic achievement, a hierarchical regression analysis was conducted (See Table 9).

Table 9. *Regression Analysis Results on Predicting PS Skills*

		Beta	t	p	VIF	Adj R2	F	p
Model 1	Constant	0.705	19.779	0.000		0.750	3021.823	0.000
	GIM	0,548	54.971	0.000	1.000			
Model 2	Constant	0.426	9.117	0.000		0.767	1663.741	0.000
	GIM	0.419	23.788	0.000	3.350			
	GPA	0.189	8.787	0.000	3.350			

The GIM variable was taken first in the analysis. This variable explains 75% of the variance in PS, and GIM is a significant predictor. The second GAP variable was taken into the analysis. When the GIM variable is controlled, it is seen that the GAP variable contributes 17% to the variance explained previously in the PS skill variable. Thus, the total variance explained increased to 77%.

4. Discussion and Conclusion

PS generally means the examination and evaluation of problems and arriving at a solution (Heppner & Petersen, 1982). Due to the growing emphasis on PS in mathematics classrooms (Lester, 2013; Wilson, Fernandez, & Hadaway, 1993), the need for evaluating progress and instruction in PS has become increasingly urgent. With this in mind, the present study sought to investigate the PS skills of lower secondary school students and analyse these skills according to various variables. According to the mean scores of the students in the PSSS, they were deemed to have achieved a satisfactory level of proficiency in PS. Different from this result of the study, Bunar (2011) concluded that sixth grade students were not successful in mathematical PS. Similarly, according to the results of the study by Peranginangina and Surya's (2017) the majority of seventh grade students were unable to complete all stages of the PS indicators required to solve the given problem. Gökkurt et al. (2015) found that the students were not sufficient in the three stages (understanding the problem, preparing and evaluating the plan for the solution) and problem posing stage. When the dimensions were considered, it was found that although the students were sufficient in the comprehension and evaluation dimensions, they were at a moderate level in the implementation dimension. Some studies suggest that PS skills are learnable. When a suitable learning environment is provided, that is, when students are subjected to a continuous and systematic PS process, this acquired skill will ensure success in solving other problems. (Incebacak & Ersoy, 2016; Verschaffel, De Corte, & Lasure, 1999). Other studies in the literature also support the findings of the present study regarding the dimensions of PSSS (Baraké, El-Rouadi, & Musharrafieh, 2015; Nurkaeti, 2018; Peranginangina & Surya, 2017). According to Baraké, El-Rouadi, and Musharrafieh (2015) findings, middle school students struggle with selecting a strategy and persisting with it until they arrive at a solution. Moreover, the researchers noted that these students tend to overlook the importance of verifying whether their answers are correct or even make sense. Another study was conducted by Nurkaeti (2018) aimed to examine the PS obstacles encountered by primary school pupils using the Polya strategy. The results showed that students faced difficulties, particularly in understanding the problem, determining an effective PS plan so that the settlement is also wrong or difficult, making connections between mathematical concepts, and reviewing the accuracy of the answers with questions. Similarly, Incebacak and Ersoy (2016) carried out a study to discover secondary school students' PS abilities and PS strategies levels, based on the Polya strategy. In their study on PS skills, Peranginangina and Surya (2017) revealed that the stage of performing the plan and verifying the answer was difficult for students. This study also showed that students were able to select appropriate strategies, but encountered challenges when attempting to implement the chosen strategy. In summary, the literature made known that students had difficulties in the implementation dimension while solving problems. As a matter of fact, in this study, the students' average scores in the dimension of implementation were found to be lower than the other dimensions. Therefore, it can be said that this result obtained from the study coincides with the literature. Dissimilarly from this result obtained from the study, according to Rahmawati, Sulisworo, and Prasetyo's (2020) study, students demonstrated good performance in understanding the problem, planning for a solution, and implementing a recommended method for PS activities. However, they tended to perform poorly in rechecking the suitability of the results.

A significant discrepancy was observed between the students' PS abilities and their academic grades. In fact, the seventh grade students got higher mean scores in PS when compared to the eighth grade students, which might be attributed to the fact that eighth grade students in Turkey prepare for an exam, namely the High School Entrance Exam, and they are used to attempting to solve problems quickly. İskenderoğlu, Akbaba and Olkun (2004) also stated that the students made the wrong choice of procedures because of the quick solution. This can be seen as the reason that the eighth grade students are not peroming at the expected level in the items related to the strategy selection process and solution. Whereas the important thing is that the students can decide how to solve the problem by understanding the situational meaning of the problem rather than quickly solving the problem. On the other hand, this finding is consistent with the results of Çelik and Taşkın (2015) from seventh grade students showed the best performance in PS when grade was considered. However, the finding is contradicted by Işık and Kar (2011), and Bozkurt and Karslıgil-Ergin (2018). Işık and Kar (2011) reported that students experience an improvement in their perceptions of numbers and in solving non-routine problems in mathematics as they move to higher grade levels. Similarly, Bozkurt and Karslıgil-Ergin (2018) conducted a study with students in the fourth, the fifth and the sixth grade and concluded that most students lacked the ability to correctly identify appropriate solution strategies and effectively solve problems. However, as the grade level increased, students' competence in PS increased. Similarly, Cai (2003) concluded that a higher percentage of students in that grade level reached the correct answers as the grade level advanced in the problems given to the students. On the other hand, Yeap (2002) found that there was no statistically significant interaction effect between grade level and PS ability.

The students got relatively higher grades in the evaluation dimension of the PSSS. On the other hand, Karataş (2008) concluded from his study that students are incompetent in evaluation and do not tend to solve an evaluation-level problem in different ways. Similarly, Saygı (1990) observed that the participants in her study failed to evaluate their conclusions during PS. The study by Gökkurt et al.'s (2015) revealed that the students showed the lowest performance in the evaluation stage. Apparently, the results of the present study are not consistent with the findings of previous studies by Karataş (2008), Saygı (1990) or Gökkurt et.al. (2015). Undoubtedly, in school, teachers greatly affect their students' PS process of. Therefore, the higher scores of students in the evaluation dimension compared to other dimensions may be due to the fact that teachers give them the opportunity to think about the solutions they have made in order to use the evaluation stage proficiently in the PS process. Teachers may also have questioned what their values mean in solving the problem and the logical rationale behind their actions (Gökkurt et al., 2015).

Gender differences in mathematical PS, an important factor believed to affect mathematics performance, have attracted more attention from researchers in recent years (Zhu, 2007). Nurhayanti, Riyadi and Usodo (2020) stated that while solving mathematical PS problems, each individual's ability is definitely different when viewed in terms of gender differences. In recent years, there has been a significant amount of research examining gender differences in mathematical PS, with many studies indicating a tendency for males to perform better than females in this area (e.g. Gallagher et al., 2000; Zhu, 2007). However, in the present study, mean scores in the PSSS differed significantly depending on gender, which contradicts some previous studies in the literature (Uysal, 2007). The female participants in this study got significantly higher mean scores than the male participants in all three dimensions of the PSSS. Similarly, in Çakır's (2017) study, female participants had higher PS abilities compared to male participants. Kaytancı (1998) reported a significant discrepancy between male and female students in exhibiting behaviours critical to PS, with the difference being in favour of the latter, a finding that supports that of the present study. On the other hand, in some studies, it was found that PS scores did not show significant differences according to gender (Güven & Çabakçor, 2013; Nurhayanti, Riyadi & Usodo, 2020). Additionally, Caplan and Caplan (2005) argued that there is little to no relationship between gender and performance in mathematics. The finding is also in line with some previous studies that used the Problem-Solving Inventory (developed by Heppner and Peterson (1982) and adapted by Sahin, Sahin and Heppner (1993)). In these studies, no significant correlation was observed between the scores in the Problem-Solving Inventory and gender; however, a significant discrepancy was revealed regarding the mean scores in the dimensions of the inventory (Çakır, 2017; Kaya, Izgıol & Kesan, 2014). Also, Vermeer, Boekaerts and Seegers (2001) explored sixth grade students' mathematical PS behaviours, finding gender not to be a significant predictor. In another study on mathematical PS, Gallagher et al. (2000) reported that male students are more successful in choosing appropriate strategies for a given problem and better at taking shortcuts and finding multiple solutions. This finding is not supported by that of the present finding, for the choice of strategy is included in the implementation dimension of the PSSS, in which the female students were observed to have higher mean scores. Indeed, Zhu (2007) stated that educationalists may consider that males and females exhibit distinct approaches to solving mathematical problems. It is crucial for teachers to devise ways in which both male and female students can help them improve their PS skills using appropriate instructions. In addition, taking the positive and negative effects of classroom variables critically into consideration, efforts should be made to promote gender equality in mathematics learning (Zhu, 2007).

The literature places significant emphasis on the importance of achievement in mathematics. Academically successful students are those who acquire self-understanding using effective PS strategies that are associated with higher conceptual understanding (Garrett, Mazzocco, & Baker, 2006; Ramirez et. al, 2016). Students are successful when they solve problems related to PS. The acquisition of such skills not only leads to success in mathematics but also equips students with the necessary skills for the age we live in (Incebacak & Ersoy, 2016). It was also found in this study that there is a strong positive and significant correlation between the students' PS skills and their GPA and GIM. Similarly, in his study, Bhat (2014) sought to examine the impact of PS skills on the mathematics performance of 10th-grade students. The results of the study demonstrated that PS ability was the most reliable predictor of achievement in mathematics. Ceylan (2008) explored the connection between mathematical PS abilities in school and PS skills in the daily lives of sixth grade primary school students. For this purpose, the students completed both a mathematics test and a PS inventory. The results indicated a strong and meaningful relationship between the mathematics test scores and the PS inventory results. Similarly, Özsoy (2005) explored the interplay between fifth grade students' PS skills and their achievement in mathematics and reported a positive and significant correlation between the two. The findings support those of the present study. Kaytancı (1998) reported a significant correlation between students' exhibiting behaviours critical to mathematical PS and variances in their grades in mathematics. Thus, the results of the present study are in line with those of Kaytancı (1998). Bunar (2011), in her study conducted with sixth grade students, determined that GIM variable was an important factor in students' problem-posing and solving skills. Students with high GIM are more successful in posing and solving problems. Güven and Çabakçor (2013) found in their study with seventh grade students, a highly significant relationship between academic achievements and PS skills. A similar result was obtained at the end of the study in which Özgen et al. (2017) examined the problem posing skills of the students. The research revealed that there was a notable variation in the problem-posing scores of students depending on their overall academic achievement and their mathematics performance. These results are parallel to the results of the study. All these suggest that PS skills have a significant influence on GPA and GIM. For this reason, it is reasonable to anticipate that students' PS skills will be significantly and positively affected by their high GPA and GIM. Also, it was determined that PS skills significantly predicted GIM and GAP. In line with these discussions, we can conclude that PS is an important factor contributing to success in mathematics, highlighting the crucial role teachers play in cultivating students' PS skills. They are expected to design a reassuring environment for students to explore, take risks, and question each other (Bhat, 2014). On the basis of the conclusions, it can be argued that it is essential to improve the PS skills of students regardless of their grade levels. In addition, teachers should arrange activities that are likely to foster their students' PS skills while they are planning the instructional process. This may involve thoughtfully selecting problems that align with instructional objectives, encouraging collaborative PS among students, and creating opportunities for students to share their PS strategies with one another, all of which can enhance overall PS success (Grouws, & Cebulla, 2000). Thanks to its ability to identify students' PS skills, the PSSS will hopefully help students be guided in accordance with their abilities. It is also recommended that future studies should consider other factors, including metacognition and the affective domain, in the PS process.

5. References

Altun, M. (2004). *Matematik öğretimi* [Mathematics teaching]. Alfa.

Altun, H. (2019). Examining the problem solving skills of primary education mathematics teacher candidates according to their learning styles. *International Education Studies*, 12(4), 60-73. https://doi.org/10.5539/ies.v12n4p60

Anderson, J. R., Lee, H. S., & Fincham, J. M. (2014). Discovering the structure of mathematical problem solving. *NeuroImage*, 97, 163-177. https://doi.org/10.1016/j.neuroimage.2014.04.031

Arslan, P. Ç. (2007). *Ortaöğretim kurumları sınavına hazırlanan öğrencilerin problem çözme aşamasında karşılaştıkları güçlüklerin belirlenmesi* (The determination of the difficulties which students meet at the stage of problem solving in maths while preparing for the entrance exam for secondary school) [Master's Thesis]. Balıkesir University, Balıkesir, Turkey. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Baraké, F., El-Rouadi, N., & Musharrafieh, J. (2015). Problem solving at the middle school level: A comparison of different strategies. *Journal of Education and Learning*, 4(3), 62-70. http://dx.doi.org/10.5539/jel.v4n3p62

Bhat, M. A. (2014). Effect of problem solving ability on the achievement in mathematics of high school students. *Indian Journal of Applied Research*, 4(8), 685-688. https://doi.org/10.15373/2249555X/August2014/195

Bozkurt, A. & Karsligil-Ergin, G. (2018). Students' achievement and mathematical thinking in process of problem solving and problem posing. *E-International Journal of Educational Research*, 9(3), 1-33. DOI: 10.19160/ijer.393529

Brennan, R. L. (1996). Generalizability of performance assessments. In *Technical issues in large-scale performance assessment* (NCES 96–802), Edited by G. W. Phillips, 19–58. Washington, DC: National Center for Education Statistics.

Bunar, N. (2011). *Altıncı sınıf öğrencilerinin kümeler, kesirler ve dört işlem konularında problem kurma ve çözme becerileri* (The sixth graders' problem posing and solving abilities in sets, fractions and four operations) [Master's Thesis]. Afyon Kocatepe University, Afyon, Turkey. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Büyüköztürk, Ş. (2015). *Sosyal bilimler için veri analizi el kitabı* [Handbook of data analysis for social sciences]. Pegem Akademi.

Cai, J. (2003). Singaporean students' mathematical thinking in problem solving and problem posing: An exploratory study, *International Journal of Mathematical Education in Science and Technology*, *34*(5), 719-737. https://doi.org/10.1080/00207390310001595401

Cakir, M. (2017). Investigating prospective teachers' perceived problem-solving abilities in relation to gender, major, place lived, and locus of control. *Universal Journal of Educational Research*, *5*(6), 1030-1038. DOI: 10.13189/ujer.2017.050616

Caplan, J. B., & Caplan, P. J. (2005). The perseverative search for sex differences in mathematics abilities. In A. M. Gallagher & J. C. Kaufman (Eds.) *Gender differences in mathematics: An integrative psychological approach*. Cambridge University Press.

Ceylan, F. (2008). *İlköğretim 6. sınıf öğrencilerinin günlük hayat problemlerini çözme envanteri puanları ile matematik problemlerini çözme başarıları arasındaki ilişki* (The relationship between the scores of the six-grade primary school students at problem solving inventory and their performance at mathematical problem solving). [Master's Thesis]. Gazi University, Ankara, Turkey. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Chiu, C. W. T. (2001). *Scoring performance assessments based on judgements: generalizability theory*. Kluwer.

Chui, M. S. (2016). Effects of teacher assessment and cognitive ability on self-concepts: Longitudinal mechanisms for children from diverse backgrounds. *Saudi Journal of Engineering and Technology*, 1(4), 180-189. DOI:10.21276/sjeat.2016.1.4.9

Cohen, J. (1988). The T test for means in *Statistical power analysis for the behavioural sciences.* Lawrence Erlbaum Associates, 19–74.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th Edition). Routledge Falmer.

Çanakçı, O., & Özdemir, A. Ş. (2011). Matematik problemi çözme tutum ölçeğinin geliştirilmesi [The constructıon of mathematıcs problem solvıng attıtude scale]. *AİBÜ Eğitim Fakültesi Dergisi*, 11(1), 119-136. Retrieved from https://www.acarindex.com/pdfler/841-4673.pdf

Çelik, D., & Taşkın, D. (2015). Investigation of 5th, 6th, and 7th grade students' solving processes in arithmetic word problems. *Elementary Education Online*, 14(4), 1439-1449. http://dx.doi.org/10.17051/io.2015.27608

D'Zurilla, T. J., Maydeu-Olivares, A., & Kant, G. L. (1998). Age and gender differences in social problem-solving ability. *Personality and Individual Differences*, 25(2), 241-252. https://doi.org/10.1016/S0191-8869(98)00029-4

Ekiz, D. (2009). *Bilimsel araştırma yöntemleri* [Scientific research methods] (2nd Edition). Anı.

English, L., & Gainsburg, J. (2016). Problem solving in a 21st-century mathematics curriculum. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd Ed., pp. 313–335). Routledge.

Erden, M. (1986). Primary school 1st, 2nd, 3rd, 4th, and 5th graders' behaviours when solving problems based on four operations. *Hacettepe University, The Journal of the Education Faculty*, 1, 105–113.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Sage.

Fülöp, É. (2021). Developing problem-solving abilities by learning problem-solving strategies: An exploration of teaching intervention in authentic mathematics classes. *Scandinavian Journal of Educational Research*, *65*(7), 1309-1326, DOI: 10.1080/00313831.2020.1869070

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, 75, 165–190. https://doi.org/10.1006/jecp.1999.2532

Garrett, A. J., Mazzocco, M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research & Practice*, 21(2), 77-88. https://doi.org/10.1111/j.1540-5826.2006.00208.x

Gonzales, N. A. (1994). Problem posing: A neglected component in mathematics courses for prospective elementary and middle school teachers, *School Science and Mathematics*, 94(2), 78-84. https://doi.org/10.1111/j.1949-8594.1994.tb12295.x

Gökkurt, B., Ornek, T., Hayat, F., & Soylu, Y. (2015). Assessing students' problem-solving and problem-posing skills. *Bartın University Journal of Faculty of Education*, Volume 4(2), 751-774.

Grouws, D.A., & Cebulla, K.J. (2000). *Improving student achievement in mathematics* (Educational Practices Series—4). Brussels, Belgium: International Academy of Education & Geneva, Switzerland: International Bureau of Education.

Güven, B., & Çabakçor, B. O. (2013). Factors influencing mathematical problem-solving achievement of seventh grade Turkish students. *Learning and Individual Differences*, 23, 131-137. https://doi.org/10.1016/j.lindif.2012.10.003

Han, S., & Kim, H. M. (2020). Components of mathematical problem solving competence and mediation effects of instructional strategies for mathematical modeling. *Education and Science*, *45*(202), 93-111. DOI: 10.15390/EB.2020.7386

Hay, P. J., & Macdonald, D. (2008). (Mis) appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy & Practice*, *15*(2), 153-168. https://doi.org/10.1080/09695940802164184

Heppner, P. P., & Petersen, C. H. (1982). The development and implications of a personal problem-solving inventory. *Journal of Counseling Psychology*, *29*(1), 66-75. https://doi.org/10.1037/0022-0167.29.1.66

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*(3), 297-313. https://doi.org/10.3102/0034654305900

Incebacak, B. B., & Ersoy, E. (2016). Problem solving skills of secondary school students. *China-USA Business Review*, 15(6), 275-285. doi: 10.17265/1537-1514/2016.06.002

Işık, C., & Kar, T. (2011). Investigating of the number perception and nonroutine problem solving skills of 6th, 7th and 8th grade students. *Ahi Evran University Journal of Education*, 12(1), 57-72.

İskenderoğlu, T., Akbaba, S. A., & Olkun, S. (2004). Elementary school students' decisions in choosing an arithmetic operation for standard word problems. *Journal of Hacettepe University Faculty of Education* 27, 126-134.

Kan, A. (2009). Ölçme sonuçları üzerine istatistiksel işlemler [Statistical operations on measurement results]. In H. Atılgan (Ed.), *Eğitimde ölçme ve değerlendirme* [Measurement and evaluation in education] (ss.397–456). Anı.

Karasar, N. (2005). *Bilimsel araştırma yöntemleri* [Scientific research methods] (15th Edition). Nobel.

Karataş, İ. (2002). *8. sınıf öğrencilerinin problem çözme sürecinde kullanılan bilgi türlerini kullanma düzeyleri* [The levels of 8th grade students use of types of knowledge used problem solving process]. [Master's Thesis].

Karadeniz Technical University, Trabzon,. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Karataş, İ. (2008). *Problem çözmeye dayalı öğrenme ortamının bilişsel ve duyuşsal öğrenmeye etkisi* (Effects of problem solving-based learning environment on students' cognitive and affective learning) (Doctoral dissertation]. Karadeniz Technical University, Trabzon. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Kaya, D., İzgiol, D., & Kesan, C. (2014). The investigation of elementary mathematics teacher candidates' problem solving skills according to various variables. *International Electronic Journal of Elementary Education*, 6(2), 295-314. Retrieved from http://hdl.handle.net/20.500.11787/6843

Kaytancı, N. (1998). *İlköğretim dördüncü sınıf matematik öğretiminde öğrencilere problem çözme ile ilgili kritik davranışları kazandırılmasında öğrenme düzeyinin belirlenmesi* (Determine of learning capacity in acquired of critic behaviours related to problem solution in teaching of mathematics of primary education fourth class student) [Master's Thesis]. Pamukkale University, Denizli. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Kızılkaya, G., & Aşkar, P. (2009). The development of a reflective thinking skill scale towards problem solving, *Education and Science*, 34(154): 82-92. Retrieved from https://www.proquest.com/docview/1009841725?pq-origsite=gscholar&fromopenview=true

Kilpatrick, J. (1985). A retrospective account of the past twenty-five years of research on teaching mathematical problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 1–16). Lawrence Erlbaum.

Kyriakides, L., & Gagatsis, A. (2003). Assessing student problem-solving skills. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(4), 609-621. https://doi.org/10.1207/S15328007SEM1004_7

Lesh, R., & Zawojewski, J. S. (2007). Problem solving and modeling. In F. K. Lester (Ed.), *Handbook of research on mathematics teaching and* learning (2nd Edition, pp. 763–804). Information Age.

Lester, F. K. (1994). Musings about mathematical problem-solving research: 1970-1994. *Journal for Research in Mathematics Education*, 25(6), 660-675. https://doi.org/10.5951/jresematheduc.25.6.0660

Lester, F. (2013). Thoughts about research on mathematical problem-solving instruction. *The Mathematics Enthusiast*, 10(1-2), 245–278. https://doi.org/10.54870/1551-3440.1267

Lester, F. K., & Cai, J. (2016). Can mathematical problem solving be taught? Preliminary answers from 30 years of research. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems: Advances and new perspectives* (pp. 117–136). Springer. https://doi.org/10.1007/978-3-319-28023-3

Marcenaro-Gutierrez, O., & Vignoles, A. (2015). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, *57*(1), 1-21. https://doi.org/10.1080/00131881.2014.983720

Maulyda, M. A., Hidayati, V. R., Rosyidah, A. N. K., & Nurmawanti, I. (2019). Problem-solving ability of primary school teachers based on Polya's method in Mataram City. *PYTHAGORAS*, 14(2), 139-149. https://doi.org/10.21831/pg.v14i2.28686

Muir, T., Beswick, K., & Williamson, J. (2008). "I'm not very good at solving problems": An exploration of students' problem solving behaviours. *The Journal of Mathematical Behavior*, 27(3), 228-241. https://doi.org/10.1016/j.jmathb.2008.04.003

National Council of Teachers of Mathematics [NCTM] (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: NCTM.

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

Nurhayanti, H., & Usodo, B. (2020). Analysis of mathematical problem-solving skills viewed from initial ability and gender differences in an elementary school. *Ilkogretim Online*, 19(3), 1127-1141. doi:10.17051/ilkonline.2020.716848

Nurkaeti, N. (2018). Polya's strategy: an analysis of mathematical problem solving difficulty in 5th grade elementary school. Edu Humanities. *Journal of Basic Education Cibiru Campus*, 10(2), 140. https://doi.org/10.17509/eh.v10i2.10868

Özgen, K., Aydın, M., Geçici, M. E., & Bayram, B. (2017). Sekizinci sınıf öğrencilerinin problem kurma becerilerinin bazı değişkenler açısından incelenmesi [Examining the problem posing skills of eighth grade students in terms of some variables]. *Turkish Journal of Computer and Mathematics Education*, *8*(2), 323-351. DOI: 10.16949/turkbilmat.322660

Özpınar, İ. (2012). *6-8. sınıflar matematik öğretim programında yer alan becerileri ölçmeye yönelik ölçek geliştirme çalışması* (Study of developing scales to measure the skills in 6th - 8th grade mathematics curriculum). [Doctoral dissertation]. Karadeniz Technical University, Trabzon, Turkey). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Özpınar, I., & Arslan, S. (2016). Development, reliability, and validity of a scale for measuring secondary school students' problem-solving skills. *PONTE Academic Journal*, Volume 72(3), 206-215. doi: 10.21506/j.ponte.2016.3.25

Özsoy, G. (2005). Problem çözme becerisi ile matematik başarısı arasındaki ilişki [The relationship between problem solving skills and mathematical achievement]. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 25(3), 179–190. Retrieved from https://dergipark.org.tr/en/download/article-file/77235

Peranginangin, S. A., & Surya, E. (2017). An analysis of students' mathematics problem solving ability in VII grade at smp negeri 4 pancurbatu. *International Journal of Sciences: Basic and Applied Research* (IJSBAR), 33(2), 57-67. Retrieved from https://core.ac.uk/download/pdf/249335807.pdf

Pimta, S., Tayruakham, S., & Nuangchalerm, P. (2009). Factors influencing mathematic problem-solving ability of sixth grade students. *Journal of School Sciences*, *5*(4), 381-385. Retrieved from https://files.eric.ed.gov/fulltext/ED506983.pdf

Polya, G. (1957). *How to solve it: A New aspect of mathematical method*. Princeton University Press.

Rahmawati, T. D., Sulisworo, D., & Prasetyo, E. (2020). Enhancing students' motivation and problem solving skills in mathematics using guided discovery learning. *Universal Journal of Educational Research*, 8(12), 6783-6789. doi:10.13189/ujer.2020.081244

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, 141, 83-100. https://doi.org/10.1016/j.jecp.2015.07.014

Reeves, D. J., Boyle, W. F., & Christie, T. (2001). The relationship between teacher assessments and pupil attainments in standard test tasks at Key Stage 2, 1996–98. *British Educational Research Journal*, *27*(2), 141-160. https://doi.org/10.1080/01411920120037108

Republic of Turkey Ministry of National Education [RTMNE], (2013). *Secondary education mathematics curriculum (Grades 5, 6, 7, and 8) of Turkey*, TTKB.

Reys, R., Lindquist, M., Lambdin, D.V., & Smith, N.L. (1998). *Helping children learn mathematic* (5th Edition). Allyn & Bacon.

Rothstein, B. (2005). *Social traps and the problem of trust*. Cambridge University Press.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*(4), 215. DOI: 10.1037/0003-066X.63.4.215

Sahin, N., Sahin, N. H., & Heppner, P. P. (1993). Psychometric properties of the problem solving inventory in a group of Turkish university students. *Cognitive Therapy and Research*, *17*, 379-396. Retrieved from https://link.springer.com/content/pdf/10.1007/BF01177661.pdf

Saygı, M. (1990). *Assessment and analysis of prospective mathematics teachers' mathematical problem-solving skills for selected variables of math-ability. Reading comprehension and attitudes toward mathematics* [Doctoral dissertation], Middle East Technical University, Ankara,. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Schoenfeld, A. H. (1985). *Mathematical problem solving*, Academic.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics*, *92*(3), 109-115. https://doi.org/10.1111/j.1949-8594.1992.tb12154.x

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

Tekin, H. (1993). *Eğitimde ölçme ve değerlendirme* [Measurement and evaluation in education]. Yargı Yayınevi

Uysal, O. (2007). *İlköğretim II. kademe öğrencilerinin matematik dersine yönelik problem çözme becerileri, kaygıları ve tutumları arasındaki ilişkilerin değerlendirilmesi* (Evaluation of relation between problem solving skills, anxiety and attitudes of secondary education grade ii students directed towards mathematics course). [Master's Thesis]. Dokuz Eylul University, Izmir. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/

Vendlinksi, T., & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *The Journal of Technology, Learning and Assessment*, *1*(3), 1-20. Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1669/1511

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2001). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, *92*(2), 308-315. https://doi.org/10.1037/0022-0663.92.2.308

Verschaffel, L., De Corte, E. & Lasure, S. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, *1*(3), 195-229. https://doi.org/10.1207/s15327833mtl0103_2

Wilson, J., Fernandez, M., & Hadaway, N. (1993). Mathematical problem solving. In P. S. Wilson (Ed.), *Research ideas for the classroom: High school mathematics* (pp. 57–77). MacMillan.

Yeap, B. H. (2002). *Relationship between children's mathematical word problem posing and grade level, problem-solving ability and task type* [Doctoral dissertation]. Nanyang Technological University, Singapore.

Zhao, X., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2018). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science and Mathematics Education,* 16(7), 1387-1407. DOI 10.1007/s10763-017-9841-3

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, 8(2), 187-203. Retrieved from https://files.eric.ed.gov/fulltext/EJ834219.pdf