PRESERVICE MATHEMATICS TEACHERS' CUE UTILIZATION AND ACCURACY IN JUDGING STUDENTS' UNDERSTANDING OF RATIONAL NUMBERS

by

Aybüke Karagözler

B.S., Primary Mathematics Education, Boğaziçi University, 2020

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics and Science Education
Boğaziçi University
2023

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my supporters who contributed their ideas to this research study. I could not have undertaken this journey without my dear thesis advisor, Assoc. Prof. Engin Ader. His creative thoughts and scrutiny always impressed me; his contributions to my ideas and endless patience always helped me to continue developing this study. I am delighted to have had a chance to be his student since my undergraduate years. Collaboration with him led to exploring my own strengths. Thanks to him, he will always be my role model as a scientist and teacher since I developed my scientific knowledge and skills to be both a qualified researcher and a teacher.

I am also thankful for my thesis committee members, Prof. Emine Erktin and Assoc. Prof. Seda Saraç for accepting to be in the dissertation committee. Their contribution to this research study was important for me to develop original ideas and come to impressive conclusions. Working with them was a pleasure for me.

I had the pleasure of working with preservice mathematics teachers who accepted to be a participant in this research study. I aimed to support enhancing their teaching and learning via the process of this study. Also, I give special thanks to my dear students who contributed to this study by presenting their different kinds of personalities and learning styles.

I am grateful to my parents, Yasemin and Gürcan. Their support and reliance on me have existed since my primary school years. They always believed me to achieve each success. Without their motivation and warm hearts, I could not take any risk to become a successful researcher.

I have never been alone on the road to becoming a qualified researcher. My dear friend, who had been my sister since we were seven, Esranur listened to my every complaint and tried to find solutions when I remained unsettled in this research process. My other sisters, Ayşe and Selen, made an effort with me to reach this amazing ending. Studying with my dear sisters, who are also future academicians like me, and sharing my thoughts, concerns, hopes, and happiness strengthened me. Facing the same difficulties strengthened our bonds. While working on this study, I understood that having friends with whom you laugh even in the most desperate times is the most important aspect of life.

Lastly, I would like to express my deepest thanks to the love of my life Furkan, for his endless support, motivation, and sensibility. He was always there when I felt upset or desperate for ten years. Knowing Furkan was always holding my hand to gather my strength was valuable and impressive while conducting this research study. This study brought me good fortune since I will engage with the love of my life after having my master's degree.

ABSTRACT

PRESERVICE MATHEMATICS TEACHERS' CUE UTILIZATION AND ACCURACY IN JUDGING STUDENTS' UNDERSTANDING OF RATIONAL NUMBERS

This study investigates preservice mathematics teachers' judgment accuracy (JA) under different cue-availability conditions. Whether there is a difference in JA or cue utilization (CU) under varying availability of cues and the relationship between preservice teachers' misconception knowledge about rational numbers and JA, together with CU during judgments, are explored. Twenty-five preservice teachers made judgments about middle-graders working with rational numbers. Firstly, they completed a test measuring their misconception knowledge. Then, online meetings were organized to present the materials for them to make judgments and collect their judgment and CU data simultaneously. Collected data were analyzed to make comparisons among JA and CU and investigate the relationships between constructs involved in the study. The results showed that more accurate judgments were made when both student and answer cues were available. Answer cues were used significantly more than student cues when both types of cues were available; student cues were used significantly more when only student cues were available. The CU did not significantly relate to JA except for overall grade in math cue—one of the available student cues. Misconception knowledge was not found to be related to JA but related to the utilization of the misconception cue—one of the available answer cues. The relationship between JA and CU requires further investigation with larger samples. Also, the misconception knowledge may be measured via different instruments to give insight into why no relationship was found.

ÖZET

MATEMATİK ÖĞRETMEN ADAYLARININ ÖĞRENCILERIN RASYONEL SAYILAR ANLAYIŞINI YARGILAMALARINDAKI İPUCU KULLANIMI VE YARGI DOĞRULUĞU

Bu çalışma, matematik öğretmen adaylarının farklı ipuçlarının mevcut olduğu koşullardaki yargı doğruluğunu araştırmaktadır. Bu koşullar değiştiğinde yargı doğruluğunda veya ipucu kullanımında farklılık olup olmadığı; öğretmen adaylarının öğrencilerin rasyonel sayılara ilişkin kavram yanılgılarına dair bilgileri ve yargı doğrulukları arasındaki ilişki ipucu kullanımları ile incelenmiştir. Yirmi beş öğretmen adayı, rasyonel sayılara ilişkin bir testi tamamlayan ortaokul öğrencilerinin performansı hakkında yargıda bulunmuştur. İlk olarak onların öğrencilerin rasyonel sayılara ilişkin kavram yanılgılarına dair bilgilerini ölçen bir test uygulanmıştır. Ardından, yargıda bulunmaları ve eş zamanlı olarak ipucu kullanımları hakkındaki verileri toplamak için çevrimiçi toplantılar düzenlenmiştir. Veriler, farklı ipucu koşullarında yargı doğruluğu ve ipucu kullanımı arasında karşılaştırmalar yapmak ve kavramların ilişkilerini incelemek için analiz edilmiştir. Sonuçlar, hem öğrenci hem de cevap ipuçlarının olduğu koşullarda daha doğru yargıda bulunduklarını; öğrenci ve cevap ipuçlarının olduğu koşullarda cevap ipuçlarının, öğrenci ipuçlarından önemli ölçüde fazla kullanıldığını; öğrenci ipuçlarının, yalnızca öğrenci ipuçlarının olduğu koşulda diğerlerine göre önemli ölçüde fazla kullanıldığını göstermiştir. Sadece öğrencilerin genel matematik puanı ipucunun kullanımı ile yargı doğruluğu arasında ve kavram yanılgısı bilgisi ile kavram yanılgısı ipucunun kullanımı arasında önemli ölçüde ilişki bulunmuştur. Sonuçlara göre yargı doğruluğu ile ipucu kullanımı arasındaki ilişkinin daha büyük örneklem boyutu ile araştırılması ve kavram yanılgısı bilgisinin farklı araçlarla ölçülmesi önerilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

CU	Cue Utilization
JA	Judgment Accuracy

1. INTRODUCTION

Teaching requires many abilities to provide effective learning opportunities. Determining students' level of achievement is one of these teaching abilities. By evaluating students' performances, teachers can judge their students' level of achievement in a related topic. Thus, they can subsequently generate activities to improve students' performances. Also, teachers can diagnose students' insufficient learning and make instructional decisions according to their students' needs. At this point, teachers' judgments of their students' performances play an important role. How teachers judge their students' performances has been investigated in many studies. Shavelson (1983) underlined the idea that teachers must incorporate a lot of information on pupils from several sources. To make a judgment, teachers must integrate this information with their own views and aims and the nature of the teaching work. Limits of the circumstances related to teaching and learning should be considered while doing this integration. Additionally, whether teachers' judgments are accurate or not has been discussed as a critical issue. Accurate judgments of teachers aid in their way of teaching by affecting their actions in the class, their types of measurements of achievement, and grading decisions (Artelt and Rausch, 2014). Thiede *et al.* (2019) contended that if teachers are able to make accurate judgments about their students' learning process, it would be possible for them to improve instructional quality, and this could lead to better student understanding. Similarly, Artelt and Rausch (2014) expressed that "judgment accuracy is regarded as a central element of teacher professionalism and is thought to affect students' learning outcomes" (p. 27). Preservice teachers are trained to become professionals who meet the needs of students, create activities aiding students' understanding, and assess and evaluate the performances of students in the right and accurate way. Investigating how preservice teachers make judgments, and their judgment accuracy (JA) can inform efforts to develop their skills towards reaching such professional goals.

Rather than categorizing students as successful or unsuccessful, ranking them by estimating only their performance order, or estimating the number of correct or incorrect answers of students in a test, judging students' possible answers to each question in a test relating to a specific domain can provide fine-grained information about teachers' judgments (Hoge and Coladarci, 1989). Investigating teachers' task-specific judgments in a specific domain is suggested in previous studies examining JA (Gabriele *et al.*, 2016; Oudman *et al.*, 2018). The rational number domain is an essential part of the content in the middle school mathematics curriculum Ministry of National Education (MONE), 2018. Students' confusion about the differences between natural numbers and rational numbers or generalizing rules of other number systems to rational numbers leads to various misconceptions in middle-grade students (Van de Walle, Karp, and Bay-Williams, 2012). Preservice teachers may lack the ability to detect students' misconceptions about rational numbers while evaluating their performances (Depeape *et al.*, 2015). Oudman *et al.* (2018) examined teachers' judgments about students' performances of decimal numbers. They concluded that the reason behind teachers' making inaccurate judgments may be their insufficient knowledge of possible misconceptions of students about decimal numbers. With these considerations, this study focuses on preservice mathematics teachers' judgments of students' performances in a specific domain of mathematics-rational numbers.The auxiliary aim of this study is to investigate the relationship between preservice teachers' knowledge of students' misconceptions about rational numbers and their JA.

One of the main reasons leading to inaccuracies in teachers' judgments has been considered as types of information that teachers use while they are making judgments of students' performances, referred to as cues in the related literature (Koriat, 1997; Van de Pol, 2021a; Van de Pol, 2021b). These cues in the literature are categorized according to their types of information or their relationship with students' actual performance. Performance or answer cues are information about students' performances; student cues are information about students' characteristics, gender or nationality, etc. When conceptualized in their relationship with teachers' JA, cues are also classified as diagnostic or non-diagnostic (Van de Pol, 2021a). Diagnostic cues are found to be in a strong relationship with students' actual performances, whereas non-diagnostic cues are not. Hence performance cues are accepted as more diagnostic than student cues (Van de Pol, 2021a). In this study, there are two categories of information available for preservice teachers' use during judgments. The student cue category includes information about students' interest in math, overall grade in math, extraversion, and engagement. The answer cues category includes students' answers and solution strategies or steps in a given assignment. Investigating what cues preservice teachers use while they are making judgments about students' performances of rational numbers is another objective of this study. To explore the cues used by preservice teachers, it is necessary to collect information about their judgment process. To learn how preservice teachers make their judgments, which cues they use, and whether detecting misconceptions is related to their JA, preservice teachers complete a survey about their judgment process at the same time. After preservice teachers completed the process of making judgments, the collected data about their cue utilization (CU), JA, and misconception knowledge were examined in terms of their relations with each other. In which cue availability conditions preservice teachers make more accurate judgments, which cues they use in different cue-availability conditions, and the relationships between their cue utilization, JA, and their misconception knowledge are the main focal points of this study.

1.1. Operational Definitions

Judgment Accuracy The correspondence between the predicted and actual level of performance (Thiede *et al.*, 2019; Artelt and Rausch, 2014).

Cue Utilization The extent to which teachers use specific cue(s) while they are making judgments about students' performances. (Van de Pol and van Loon, 2019).

Misconception knowledge Teachers' knowledge of students' misconceptions about rational numbers as measured by a Pedagogical Content Knowledge (PCK) test developed by Depaepe and her colleagues (2015).

Cue Availability A condition in which specific cue(s) are available for use while making judgments about students' performance.

1.2. Purpose of the Study

The purpose of this study is to investigate preservice mathematics teachers' CU and accuracy in judging students' understanding of rational numbers under different cue availability conditions. For this purpose, whether preservice mathematics teachers' JA under different cue-availability conditions differs or not is examined. The cues they use while they are making judgments are identified for each cue-availability condition. Besides, the relationship between the extent to which preservice teachers are knowledgeable about misconceptions of students and their CU and JA are also explored since this component of teacher knowledge requires investigation for its potential links with key constructs in teachers' judgment of students' performance.

1.3. Significance of the Study

This study is an investigation of the accuracy of preservice mathematics teachers' judgments of students' performance in a specific domain of mathematics – rational numbers. Artelt and Rausch (2014) emphasized that the importance of examining variables related to the JA of teachers has been highlighted. Thiede *et al.* (2019) contended that "it is especially important to study the accuracy of teacher judgments and the factors that influence judgment accuracy" (p. 690). This study incorporated an exploration of preservice teachers' knowledge of students' misconceptions about rational numbers and relates it to their JA and CU while they are making their judgments. The results are expected to give insights into preservice mathematics teachers' ability to judge their students' performances. Some studies check in-service mathematics teachers' JA, but they may lack why teachers made inaccurate judgments. In-service teachers make judgments about their own students' performances; thus, they may be under the influence of their personal thoughts about the students. Thus, these studies try to relate the reason behind their inaccurate judgments to teachers' judging their own students whom they have been familiar with. Since preservice mathematics teachers do not have their own classes and groups of students they are familiar with, the present study searches for the difference in their JA by manipulating the available cues in different conditions. Therefore, different cue-availability conditions are created for preservice mathematics teachers. In each condition, the preservice teacher judges students having different achievement levels. It is explained that studies about JA using task-specific judgments do not sufficiently exist (Artelt and Rausch, 2014). In this study, preservice teachers try to make judgments by indicating whether middle-grade students give correct or incorrect answers to each item in the given assignments about rational numbers. Rational numbers is a specific domain to be judged rather than the general mathematical ability of students. This is another contribution of this study since there is insufficient research investigating JA for a specific domain rather than general mathematical ability (Gabriele, 2016; Oudman *et al.*, 2018).

1.4. Research Questions

RQ1: Is there a significant difference in preservice teachers' judgment accuracy in different cue-availability conditions?

RQ2a: Is there a significant difference between the extent to which particular cues are used in the student/answer only condition and student+answer condition?

RQ2b: Is there a significant difference between preservice teachers' utilization of student cues and answer cues when both are available?

RQ3: Is there a significant relationship between preservice teachers' cue utilization and their judgment accuracy?

RQ4: Is there a significant relationship between preservice mathematics teachers' knowledge of students' misconceptions about rational numbers and their judgment accuracy of students' understanding of rational numbers?

RQ5: Is there a significant relationship between preservice mathematics teachers' knowledge of students' misconceptions about rational numbers and their cue utilization?

2. LITERATURE REVIEW

2.1. Judgment Accuracy

Shavelson (1983) described judgment, in general, as “the process of evaluating or categorizing a person or an object” (p. 395). This study investigates teachers’ judgments and various factors that can be related to teachers’ JA. Thus, these factors and the evaluation of the JA are discussed in the following subsection.

2.1.1. Teachers’ Judgment Accuracy

There are different operational definitions for JA. A common conceptualization is the correspondence between the predicted and actual level of performance of students (Thiede *et al.*, 2019). Similarly, Dunlosky and Thiede (2013) defined JA as “how well a person’s judgments are related to target performance” (p. 59). Artelt and Rausch (2014) stated that JA should be seen as “one facet of the outcome of diagnostic competence” (p. 28). JA is commonly addressed as the ability to judge students’ features and achievements in the right way. Absolute accuracy and relative JA are generally used in studies examining JA. Absolute accuracy indicates the degree of the relation between the performance of students on a task and the judgments of teachers. It is the difference between the predicted and actual performance of students. (Thiede *et al.*, 2018). “Relative accuracy describes the degree to which judgments discriminate between different levels of performance across students. This is typically reported as the intra-teacher correlation between predicted and actual performance computed across students” (Thiede *et al.*, 2018, p. 107). In this study, the correspondence between preservice teachers’ judgments and students’ performances is considered as their JA.

Teachers’ judgments about their students’ performances have a huge potential to affect both learning experiences and educational trajectories since teachers tend to make their instructional decisions with the help of these judgments. Making accurate judgments as a part of diagnostic competence is important because of its effect on deciding proper class activities, tasks, and evaluation types (Artelt and Rausch, 2014). Pit-ten Cate *et al.* (2014) stated that "a central aspect of teachers' professional competence is the ability to judge students' achievements adequately" (p. 45). Many studies about JA indicated that making accurate judgments about students' performances aid in observing the academic difficulties of students (Südkamp *et al.*, 2012). According to judgments about students' understanding level, teachers can decide how much time is necessary for students to review the topics, which students are to be given remedial work, and the pace of the instruction. Accurately judging students' performances has been considered as an essential characteristic of a qualified teacher. Teachers try to perceive students' skills to make an estimation about their participation in the lesson or achievement in the class via making judgments (Shavelson, 1983). Besides informing instructional decisions, accurate judgments help teachers to develop proper academic self-concept and to decide the placements of students fairly (Südkamp *et al.*, 2014).

A teacher can tailor instruction to each student's needs if they can accurately assess their learning and determine whether they are progressing toward the learning goal. In contrast, a teacher who is unable to differentiate between different levels of student learning may give each student less attention than necessary (Thiede *et al.*, 2019). Improving judgment can be achieved by concentrating on diagnostic competence, which involves acquiring the necessary skills for making more accurate judgments for people (Pit-ten Cate *et al.*, 2014). Diagnostic competence includes giving grades for students to place them among others or informal thoughts for the performance or attitudes of students (Pit-ten Cate *et al.*, 2014). Assessing the levels of preservice teachers' JA is important for developing their diagnostic competence, so this study investigates the accuracy of the judgments that are made.

2.1.2. Factors Affecting Teachers' Judgment Accuracy

Shavelson and Stern (1981) prepared a decision model demonstrating the factors contributing to teachers' pedagogical judgments and decisions. They explained that information about students, differences among teachers' characteristics, and the nature of the instructional task are the main contributors to the judgments of teachers. When a task to be judged is about general ability, and teachers are expected to judge the specific ability of students, in that case, there is no accordance between judgment and test characteristics, and this negatively impacts JA (Südkamp *et al.*, 2012). Besides, Südkamp and colleagues (2012) explained the time gap between the judgments of teachers and tests done for measuring students' performances and congruence in domain specificity of teachers' judgments and students' performances as important factors affecting teachers' JA. They suggested that if the test that students complete and the rating scale that teachers are asked to fill are conducted simultaneously rather than at different time intervals, teachers' judgments and students' performances will show more correspondence. These factors are kept the same for each condition in the present study to make teachers' judgments more comparable. According to the Heuristic Model of Teacher Judgment Accuracy, the actual test performance of students can be affected by student characteristics or test characteristics (Südkamp *et al.*, 2012). The students' assignments in this study include items related to specific misconceptions of students about rational numbers. Preservice teachers were asked to make item-specific judgments about students' performances for answering questions on rational numbers topic. Which misconceptions students have about rational numbers were expected to reflect in their performances in the assignments they completed for this study since the items in the assignments were prepared in relation to specific misconceptions that students often have about rational numbers. Therefore, teachers' judgments about students' performances can be affected by both teacher characteristics and judgment characteristics. Südkamp *et al.* (2012) asserted that "a teacher's characteristics are thought to influence her or his judgment at various stages of the judgment process" (p. 8). It was expected that preservice teachers' knowledge of students' misconceptions about rational numbers could affect their judgment process. This could also influence the accuracy of their judgments. Südkamp *et al.* (2012) contended that the more teachers are familiar with their students in terms of spending more time with them or the more years of experience that teachers have, the more accurate judgments they can make. Since preservice teachers do not have their own students, their JA may not be as developed as in-service teachers'. Thus, other variables rather than teaching experience in this study may be related to a preservice teacher's JA.

Südkamp *et al.* (2012) assumed that teachers would judge their students' performances more accurately when teachers' judgments are domain-specific about students' particular performance in that domain. Artelt and Rausch (2014) supported this idea by arguing that in addition to variations among teachers in their JA levels, the task used for the judgment or features of both students and teachers themselves plays a role in those variations. While some studies investigate the judgments of teachers regarding their students' general academic ability, some of them focus on their judgments of the specific abilities of students. It can be expected that making judgments about the specific abilities of students gives more accurate results since, in that case, teachers are not concerned about other abilities, and their judgments can be more focused (Südkamp *et al.*, 2014).

To improve teacher JA at the teacher training level, the use of task-specific judgments enables more detailed information about judgments because tasks differ in their difficulty, demand, and type (Artelt and Rausch, 2014). Since preservice teachers made item-specific judgments for students' performances about the rational number domain for this study, the relationship between making more specific judgments and JA was investigated in detail.

2.1.3. Evaluation of Judgment Accuracy

The accuracy of judgment shows differences when it is evaluated with different types of measurement. The decision for the rank order of student performance is one of the indicators for the accuracy of teacher judgments. It correlates teacher judgments for the rank order of their students with the original rank order indicated as a result of standardized tests. However, the rank-order judgments done by teachers do not always show stability between results (Artelt and Rausch, 2014).

Südkamp *et al.*, (2012) examined the studies done to explore the relationship between the judgments made by teachers about their students' academic achievement and students' actual achievement in several conditions varying in terms of types of schools, grade levels of students, and subject areas. Thus, researchers reviewed studies to examine the correlation between the accuracy of teachers' judgments and students' actual performances. Their results showed that there is a moderate or more than moderate correlation between these two variables. It can be said that the correlation values did not reach the perfect value. They suggested that different measures rather than correlation can be used to check the association between teachers' JA and students' achievement. Insufficient definition of accuracy may be the limitation of the studies examining teachers' JA. Studies done to look for the association between teachers' JA and students' actual achievement on teacher JA create a discussible point, which is that they do not take into consideration whether teachers' judgments are influenced by the academic performance of the students in their class (Südkamp *et al.*, 2012).

Van de Pol (2021a) studied how information given to teachers about their students' understanding develops teachers' monitoring and regulation accuracy. According to their results, name-only (including information about students themselves) or diagram-only (including information about students' performances) conditions did not show any significant differences in terms of their relative or absolute monitoring accuracy. There were similar findings for the name-only and name+diagram conditions. When teachers examined students' answers in the name+diagram condition, they made more accurate judgments than in the diagram-only condition. Researchers contended that the reason why teachers' monitoring accuracy was higher when they had both performance and student cues accessible instead of only performance cues was that the monitoring accuracy of teachers was hampered in performance-only condition. This is because they still have a tendency to use information about students, even if they only receive information about the answers.

Oudman *et al.* (2018) examined whether teachers' JA about their students' conceptual mathematical understanding differed in various conditions or not. They hypothesized that if more diagnostic cues that better predict students' performances are used, teachers' JA would be more precise. Researchers made a comparison between teachers' JA when they used the performance cues, which were thought to be more diagnostic, and student cues which were considered as less diagnostic. The comparison was made between students' actual performance and teachers' item-specific predictions for their performances for decimal-magnitude problems. They selected a specific domain to be judged rather than a general one. More specifically, they hypothesized that when teachers knew only the name of their students, their JA would be lower than when they knew the answers only or knew both the names and the answers. Researchers found a difference between teachers' JA for what students do not understand in answer-only and name-only conditions. They concluded that the judgments of teachers were more accurate for what students did not understand in the answer only condition. Oudman *et al.* (2018).

Suggested investigating teachers' item-specific judgments for students' conceptual mathematics understandings by using the same item format to more precisely identify the differences in JA. The present study is designed to investigate teachers' JA regarding students' understanding of rational numbers measured by assignments including items in the same format rather than their general mathematical understanding.

2.2. Cue Utilization

Many studies benefited from the cue-utilization framework (Koriat, 1997) while they explained the necessity of using information that can be called "cues" in the judgment processes of teachers. In this study, types of cues are presented to preservice teachers for use while they are making judgments about students' performances. Thus, important aspects of teachers' CU and the types of cues and their relationship with teachers' JA are explained in this section.

2.2.1. Types of Cues

Van de Pol *et al.* (2021a) suggested that teachers make more accurate judgments when they use information about students' performances rather than information about students themselves in their judgments. The information about students' performances is called diagnostic cues in studies. Diagnostic cues are seen as the most predictive ones for estimating students' performances. Pit-ten Cate *et al.* (2014) asserted that judgments and decisions of teachers can be affected by non-academic factors besides academic ones, like social and migrant characteristics of students, which are often referred to as non-diagnostic cues.

Van de Pol *et al.* (2021b) looked for the diagnosticity of cues for the students' real performance by measuring actual cue values and correlating them with the test scores of students relating the text comprehension. Also, they asked teachers to judge the values of the cues that they use while they were making judgments about the text comprehension of students. The expected highly diagnostic cues were performance cues that correspond to, for example, correct relation in students' diagrams in their study investigating teachers' judgments about text comprehension. Van de Pol *et al.* (2021b) indicated that student and task cues are not as diagnostic as performance cues. They collected information from the students to access the actual values of cues. The cue diagnosticity was explained by the correlation between the actual cue values and the real test scores of students. They found that each student cue had low diagnosticity (all < 0.30), but certain student cues, i.e., general reading comprehension and IQ, were more diagnostic than other student cues. The results of the study conducted by Van de Pol *et al.* (2021b) showed that student and task cues were less diagnostic than performance cues. Some common student cues that are mostly used in studies are student effort, student extraversion, student interest, gender, nationality, self-efficacy, and general grades for the related subject; some common answer cues that are mostly used in studies are item performance of a judged test, general test performance, students' guessing and time taken while taking tests (Van de Pol *et al.*, 2021b).

2.2.2. Teachers' Cue Utilization

Research showed that teachers did not sufficiently use diagnostic cues, and when they were able to use them, it was difficult for teachers to reach the real meaning of those cues (Van de Pol *et al.*, 2021b). Interpreting a diagnostic cue in the wrong way to judge any skill or performance level of the student is a reason for inaccurate judgment (Van de Pol *et al.*, 2021b).

In previous studies, it has been highlighted that the types of cues used by teachers or other participants, according to the scope of the research, should be detected by using verbal reports concurrently or after the judgment process (Van de Pol *et al.*, 2019). Oudman *et al.* (2018) collected data about teachers' CU by conducting a think-aloud process. They identified the cues used by teachers in changing conditions. The unexpected finding was teachers' use of student cues even when they were in the answers-only condition that they were formally presented with information about students' performances in a mathematical task. Teachers were often found to make comments about students' characteristics as the underpinning reasons for their judgments in answers-only condition (Oudman *et al.*, 2018).

If the think-aloud process is not conducted, teachers' own reports for their CU may be a useful data collection tool. Van de Pol *et al.* (2021b) asserted that: "a scale that asks teachers to indicate to which extent each cue has been used could provide further information on the weighting of several cues" (p. 794). In this study, preservice teachers' CU is measured by using a rating scale that asks them to give scores for the utilization of presented cues at the moment that they make their judgments about students' performances.

2.2.3. The Relationship Between Teachers' Judgment Accuracy and Cue Utilization

In their research, Van de Pol *et al.*, (2021b) contended that the diagnostic cues could enhance teachers' monitoring accuracy while they were making judgments. Thiede *et al.* (2019) explained the importance of cue diagnosticity because of its effect on the JA of teachers by relating the use of diagnostic cues with more accurate judgments.

There are some behaviors of students affecting the judgment process of teachers, as summarized by Thiede *et al.* (2019). These are classroom engagement of students, interest in the domain, students' effort and conduct, verbal assertiveness, and self-control. Some of these students' behaviors are measured and presented to preservice teachers as student cues in this study to investigate their dependence on such cues and the relationship between the use of such cues and JA. Thiede *et al.* (2019) contended that these behaviors of students are unlikely related to students' performances, and teachers tend to make wrong interpretations of these behaviors. In the case of the use of such cues, inaccurate judgments are likely to occur by teachers about students' achievement. Van de Pol *et al.* (2021b) had expected that if teachers would use highly diagnostic cues and indicate the cue values accurately, they tend to make more accurate judgments. Oudman *et al.*'s (2018) study investigated the differences in teachers' CU under changing conditions. According to their review of the existing studies investigating teachers' cues that were used while evaluating their students, they compiled several commonly used cues under categories which are task content cues, answer cues, and student cues. So, it was hypothesized that if teachers used only answer cues, they would make more precise judgments about their students' performances. According to their findings, teachers' use of general knowledge about students did not significantly differ in the condition they knew only the names of the students, and they knew both the names and answers of the students.

Artelt and Rausch (2014) expressed that "teachers tend to overestimate the competence of students that are similar to themselves in terms of their personality profile" (p. 38). Overestimation of students' performances because of the similarity influence leads to generating more positive judgments for students' performances. This bias related to having similar students to their teachers affecting judgment situations only exists for global judgments, according to Artelt and Rausch (2014). Since teachers look for content domain or students' competence more when they do task-specific judgments, these judgments are not vulnerable to similarity bias. The preservice teachers make judgments about the performances of students in a specific task in this study.

The results of Oudman *et al.*'s (2018) study showed that teachers made more accurate judgments about what students did understand in the name-only condition than in the answer-only condition. This situation may be interpreted as knowing their students' names while making judgments about their performances, making teachers more positive about judging what students did understand. However, teachers made more accurate judgments about what students did not understand in the answers-only condition. Examining the errors in students' answers may make teachers more accurate in judging students' misunderstandings (Oudman *et al.*, 2018). Oudman *et al.* (2018) found no significant difference between teachers' use of answer cues when they are in answer-only or name+answers conditions. Also, they found no significant difference between the use of student cues when teachers had access to both students' names and answers or only the names.

The selection of the cues presented to participants in a study is an important issue. Thus, researchers investigate diagnosticity by checking the utilization of cues in other studies. While some cues are utilized mostly by teachers but have low diagnosticity, some are not utilized by teachers but have high diagnosticity. Thus, the selection of which cues are presented to teachers is made according to the aim of the studies. Rowan (2022) studied information that teachers use while making judgments of students' performances regarding text comprehension in her research. She examined how cue diagnosticity affects JA and how a cue-diagnosticity intervention affects cue use, JA, and judgment confidence. They asked secondary school teachers to judge anonymous students' text comprehension performance under different conditions. One of the conditions was the existence of only non-diagnostic cues, which were the student cues. Since the students were not teachers' own students, the information collected as student cues in another study was used. Rowan (2022) presented four cues in each condition to be used by teachers while they were judging. The presented cues were selected according to their utilization and diagnosticity values in the study of Van de Pol *et al.* (2021b). The most diagnostic performance cues were selected. However, the criteria for the selection of non-diagnostic cues from Van de Pol *et al.*'s (2021b) study was explained as having low diagnosticity ($< .30$), high reliability ($>.75$), and moderate to high CU. Rowan (2022) found that JA was lower when only non-diagnostic cues existed compared to the condition when only diagnostic cues existed, and after the intervention, teachers used non-diagnostic cues less when both diagnostic and non-diagnostic cues were available for them, which was another condition in her study.

In this study, the selection procedure of the cues presented to preservice teachers is similar to what has been reported in previous studies. However, this study investigates JA in a specific domain of mathematics; answer cues were selected according to this specific topic. Further, the extent to which preservice teachers use diagnostic and non-diagnostic cues while judging student performance for this specific topic is explored.

2.3. Cue Availability

When teachers make judgments about their students' performances, they need to benefit from various information. In studies measuring the JA of teachers, the information about students' characteristics and the information about students' performances in a similar topic judged at the same time, or both of these information is often presented to teachers.

Oudman *et al.* (2018) looked for studies investigating teachers' cues that they used while judging their students' performances, and they concluded that task content cues, answer cues, and students' cues were most widely used by teachers. To look for the differences in teachers' CU in different conditions, Oudman *et al.* (2018) described three conditions having different cue availability. These conditions were named name-only, answer-only, and name+answer. Oudman *et al.* (2018) asserted that answer cues were more predictive than student cues, and their use of them contributed to the teachers' JA in the answer-only condition and name+answer condition. To measure whether the existence of different cues led to differences in teachers' judgments, the availability of the cues was controlled by the researchers. They found that teachers could accurately judge what students have difficulties with and which errors students can make when answer cues were available to them and student cues were not available.

Since different cue-availability conditions allow teachers to make their judgments by benefiting from varied sources of information, this study investigated preservice mathematics teachers' JA in three cue-availability conditions which are student-only, answer-only, and student+answer. However, preservice mathematics teachers do not have their own classes, so they make judgments for anonymous students. Their only information source is the cues in the student-only conditions, which are presented above. Thus, it was not expected for preservice teachers to try using information about the general characteristics of students while they were making judgments in answer-only conditions since they did not know any of the students personally. In this study, teachers' JA and CU were examined in different cue availability conditions.

2.4. Misconceptions about Rational Numbers

Misconceptions can be thought of as misunderstandings or wrong interpretations attributed to wrong meanings (Ojose, 2015). There is not only one common misconception that all students may have for each topic in mathematics. They are expected to have misconceptions about any topic. Students may have misconceptions in different forms for a specific domain which collectively prevents their reasoning about a related topic. Since identifying and dealing with each misconception is not an easy task, how teachers can deal with the misconceptions of their students has been an important and discussed issue in the last years. If teachers are knowledgeable about the common misconceptions of their students, they have the potential to apply proper instructional strategies to reduce their negative effects (Ojose, 2015; Depeape *et al.*, 2015).

In this study, the focus is on students' misconceptions about the rational number concept and the extent to which preservice teachers' knowledge about these misconceptions is related to their judgments about students' performance. Students first meet with natural numbers in early grades, and then they start to learn rational number concept. It has been discussed by various researchers that perceived rules and properties of the natural number concept hinder students' learning of rational numbers. According to Van de Walle *et al.*, 2012, teachers are expected to support their students in seeing the similarities and differences between fractions and whole numbers. Fractions do not show all the properties of natural numbers, they are used to indicate the values between two natural numbers, and they cannot be counted like natural numbers. Besides, the operations with them require different strategies than those on natural numbers. Decimal numbers are different from natural numbers in terms of the relationship between the magnitude of the number and the number of the place values, and also they require different strategies while doing operations with them, like in fractions.

Van Hoof *et al.* (2015) mentioned that the dense structure of rational numbers, the numerical size of rational numbers, and the effect of operations are characteristics of rational numbers that are different from whole numbers. Van Hoof *et al.* (2015) stated this issue as "there is no such thing as a successor number of a given rational number, as there are always infinitely many numbers between any two rational numbers" (p. 3). Many students have challenges while noticing this difference between natural and rational numbers. While students make a comparison between two fractions, they can say that if a fraction's numerator and denominator increase, it has a greater numerical value. Besides, while they are making a comparison between two decimals, they make an error that if a decimal is longer, it has a greater numerical value. These kinds of mistakes can be attributed to the numerical size aspect of a rational number (Van Hoof *et al.*, 2015). Not being able to apply the rules of operations with whole numbers to rational numbers is another reason for students to have difficulties in making sense of rational numbers. For example, while addition and multiplication give greater results and subtraction and division give smaller results for natural numbers, this rule is not always valid if one wants to do these operations with rational numbers. All these differences between natural and rational numbers were stated as the elements of natural number bias in previous studies (e.g., Ni and Zhou, 2005).

Since rational numbers do not have one form, they are seen as more complicated than natural numbers by students. Thus, students tend to have misconceptions about any form of rational numbers. Behr (1981) worked on Kieren's (1976) analysis of the rational number construct and reorganized the meanings of a rational number. As a result, seven meanings have been emphasized: fractional measure, ratio, rate, quotient, linear coordinate, decimal, and operator (Behr, 1981). In this study, the fractional measure meaning and the decimal meaning are investigated in terms of their related misconceptions.

In addition, while Van de Walle *et al.* (2012, p. 291) were interpreting the causes of the challenges of students with fractions, they presented the existence of multiple meanings of fractions as the main cause. When Van de Walle *et al.* (2012) focused on fractions, they explained the constructs of it as part-whole, measurement, operator, and ratio. In the scope of this study, fractions will be mentioned as a part-whole construct.

Also, Behr (1981) related the base ten numeration system with the decimal meaning of a rational number. Van de Walle *et al.* (2012) asserted that to show that rational numbers can be presented in different forms, the relationship between decimals and fractions can be used by converting them to each other.

Van de Walle *et al.* (2012) explained the reason behind the students' difficulties with fractions as their tendency to apply strategies they have learned for whole numbers while they are dealing with fractions. Also, the different representation of fractions than whole numbers and the conceptual understanding of fractions are not emphasized by many teachers, which leads to difficulties while learning them (Van de Walle *et al.*, 2012). One of the misconceptions about the concept of a fraction is students' interpretation of the numerator and denominator as independent values from each other. Another misconception is that if there are two fractions with the same numerator, students might think that a fraction having a greater denominator is greater than the other fraction (Van de Walle *et al.*, 2012).

Van de Walle *et al.* (2012) stated that students tend to overgeneralize properties of operations with whole numbers while they are doing operations with fractions. He listed the most common misconceptions about addition and subtraction with fractions as: adding/subtracting both numerators and denominators, failing to find common denominators, difficulty in finding common multiples, and difficulty with mixed numbers. Siegler *et al.*, (2010) explained that misconceptions related to adding and subtracting fractions can be attributed to students' not thinking of fractions as numbers with numerical values. Learning the rules about adding and subtracting with fractions can cause some mistakes while students are doing multiplication and division with fractions. Also, the most common misconceptions about multiplication with fractions were listed as: treating the denominator the same as in addition/subtraction problems, inability to estimate the approximate size of the answer, and matching multiplication situations with multiplication (dividing by 1/3 instead of multiplying with 1/3). Some important misconceptions for the division were listed as: thinking that the answer should be smaller, connecting the illustration with the answer, and writing remainders (Van de Walle *et al.*, 2012).

There are also common misconceptions regarding decimal numbers explained in previous studies. Irwin (2001) expressed and used the four common misconceptions about decimal numbers in his study, which are:

(i) Students' thinking that the whole and the decimal part are distinct parts that are separated from each other with a "decorative dot",

(ii) Students' making a connection between the length of the decimal with its numerical value, misconceptions about the representation of the quantities,

(iii) The effect of the zero according to its place in the decimals.

Durkin and Rittle Johnson (2015) summarized these misconceptions as the whole number misconception, the role of zero misconception, and the fraction misconception. The whole number misconception can be expressed as students' thinking that decimals have all the characteristics of whole numbers by applying the strategies used for whole numbers to decimals, which can be stated as whole number bias (Ni and Zhou, 2005). Also, the role of zero misconception can be related to the whole number bias because if there are two decimal numbers with the same whole part, students tend to think that the existence of the zero in the tenth place does not affect the value of the decimal. Students' applying strategies for fractions to decimals can be stated as a fraction misconception in decimals. For example, relating the denominator with the decimal by saying that if there is a longer decimal, it should have a smaller value since fractions with longer denominators have smaller values (Durkin and Rittle Johnson, 2015).

Oudman *et al.* (2018) conducted research on the judgments of teachers about students' performances for decimals by altering the available information that teachers can access. Oudman *et al.* (2018) prepared assessments about decimals for students to be completed. The questions in the assessment are related to specific misconceptions about decimals. The used misconceptions in this study are listed by Oudman *et al.* (2018) as:

> (1) thinking of decimals as if they are whole numbers (e.g., 0.35 is greater than 0.8 because 35 is greater than 8); (2) ignoring a zero that is in the tenths place (e.g., 0.08 is the same at 0.8); (3) assuming that adding a zero at the end of the decimal increases its magnitude (e.g., 0.30 is greater than 0.3); (4) viewing decimals less than one as being less than zero or more than one (e.g., 0.2 is less than 0); and (5) treating decimals as fractions thus thinking that numbers with more decimals are smaller. (p. 217).

2.4.1. Teachers' Knowledge of Misconceptions

Teachers' PCK plays an important role in teachers' detection and correction of misconceptions. PCK is not only having the necessary information about the subject taught but also knowing how learning can be easy or difficult for students. Therefore, teachers should understand the misconceptions of students that make their learning process difficult and develop strategies to decrease the negative effects on students' performances as a part of their PCK (Shulman, 1986).

Newton (2008) examined prospective teachers' content knowledge (CK) about rational numbers. His results showed that prospective teachers tend to make mistakes while doing operations with fractions as students. It is presented that their mistakes can also be attributed to their prior understanding of fractions instead of their knowledge of natural numbers. She designed a course to develop prospective teachers' understanding of rational numbers, and they analyzed the changes occurring in teachers' understanding at the end of the course. As a result, they observed that prospective teachers' difficulties due to their misconceptions about fractions were reduced. If it is thought that they teach fractions to students in their future careers, the decrease in their difficulties can be thought to be a favorable development for their instructional decisions to deal with misconceptions of students. Also, Turnuklu and Yeşildere (2007) worked with preservice elementary teachers to examine their mathematical knowledge and PCK by looking at the relations between these two. They designed a task with four problems related to integers, fractions, and decimals and answers given by students to these problems. Preservice teachers were asked to investigate the students' answers, and they explained the students' understanding of a question by making interpretations about the reasoning of students. They tried to detect misconceptions of students and developed proper strategies to correct them, and they decided on a criterion about how they could assess the performance of the students in this task. The results of the study showed that preservice teachers had a lack of ability to recognize the possible causes of students' misconceptions, and their insufficient CK led them to behave without considering the topic they were teaching (Turnuklu and Yeşildere, 2007).

In another study, Stacey *et al.* (2001) worked with preservice elementary school teachers to look for their CK and PCK about decimals. They asked teachers to choose items from a given decimal comparison test, in which students could have made a mistake while solving. Their results showed that teachers' CK is not sufficient to see the connection between decimals, whole numbers, and fractions. Also, teachers were found to lack PCK since they could not detect that students might have trouble with the items of the type that they themselves got wrong.

Tirosh (2000) investigated preservice teachers' PCK about reasons for students' mistakes in division with fractions. They designed a course to improve preservice teachers' reasoning for the students' answers about division with fractions. According to their results, before taking the course, most preservice teachers had tended to attribute the causes of students' erroneous answers to their difficulty in reading or algorithmically based errors.

However, at the end of the course, preservice teachers started to see the causes, like generalizing rules of dividing natural numbers to dividing fractions, and they considered the outcomes of these kinds of misconceptions for students' conceptions in the future. Tirosh (2000) recommended that teacher education courses should give some information to preservice teachers about students' way of thinking, their kinds of cognitive processes, and their common mistakes.

Besides, İsiksal and Cakiroglu (2011) studied the PCK of prospective elementary teachers about the students' difficulties with fraction multiplication. They looked for whether prospective teachers were able to give reasons for these difficulties by noticing the misconceptions of students and also giving recommendations for dealing with these misconceptions. İsiksal and Cakiroglu (2011) found that:

> The prospective teachers' perceptions of children's mistakes about multiplication of fractions could be grouped under five headings: namely, algorithmically based mistakes, intuitively based mistakes, mistakes based on formal knowledge of fraction operations, misunderstanding of the symbolism of a fraction, and misunderstanding of the problem. (p. 119).

The mistakes of misunderstanding the symbolism of a fraction and misunderstanding the problem are different from the causes that Tirosh (2000) had presented before. According to the prospective teachers, algorithmically based mistakes could primarily be caused by rote memorization, and conceptions related to primitive models lead to intuitively based mistakes. Besides, not having sufficient formal knowledge of fractions is related to formal knowledge mistakes. According to prospective teachers, feeling anxious while solving a problem and lacking mathematical knowledge prevented students from showing high performances when solving a problem (İsiksal and Cakiroglu, 2011). It was recommended that prospective teachers' PCK should be investigated in further studies to understand how their knowledge is improved in the teacher education programs and how PCK can be connected to others. Besides, the examination of knowledge of prospective teachers about how they teach rational numbers was suggested by (İsiksal and Cakiroglu, 2011).

2.4.2. Links Between Teachers' Knowledge of Misconceptions and Other Instructional Constructs

If student misconceptions about rational numbers can be detected by teachers, they can be easily corrected. The detection of misconceptions helps teachers to understand what and how their students have learned or why their students have made such mistakes. If the lack of understanding of students is due to misconceptions, teachers should be able to notice the difficulties of students and evaluate their performance of the students by considering the effect of having a misconception. Therefore, teachers should be knowledgeable about the misconceptions of students to correctly detect them and evaluate their performances by considering their misconceptions.

Depaepe *et al.* (2015) conducted research about the teachers' CK and PCK on rational numbers by stating that rational numbers are one of the difficult topics in mathematics, so there is a need for teachers to be aware of challenges that students have about this topic and solve them. They looked for the relationship between content knowledge (CK) and PCK of prospective elementary teachers on rational numbers by applying a test measuring both. Since Shulman (1986) divided PCK into two parts as teachers' knowledge of students' misconceptions and of instructional strategies to overcome them, Depaepe *et al.* (2015) have created a test that includes items related to both of these parts. They hypothesized that prospective teachers' CK and PCK are inadequate for rational number topic. Also, they had thought that they would find the performance of prospective teachers on the PCK part worse than CK. Depaepe and colleagues (2015) stated that CK is a required but insufficient prerequisite for the acquisition of PCK for a certain mathematical topic. Their results supported this idea because prospective teachers had performed better on CK items than PCK items, so their CK was not sufficient to perform well on PCK items. Prospective teachers' lack of PCK was attributed to their inadequate classroom experience, but their practice of instruction about rational numbers in their teacher education program showed that they were not unaware of how to teach this topic (Depeape *et al.*, 2015).

Investigating teachers' CK and PCK gives meaningful ideas about teachers' proficiency, but there is a need to examine the reasons for the gap between these two types of knowledge and how teacher education programs can be designed to emphasize the knowledge about rational number topics (Depeape *et al.*, 2015).

3. METHOD

3.1. Participants

Twenty-five preservice mathematics teachers (M_{age} = 23.6, SD = 1.19) from a public university in Istanbul constituted the participants in the study. They were senior students doing their internship in middle schools in Istanbul, and most of them were in their final semesters at the university. The convenience sampling method was used for the selection of the participants since participants were chosen from the university where this research was conducted. Participation in the study was voluntary, and a small amount of extra credit in their seminar on practice teaching was offered in return for their participation.

3.2. Design

This study had two main aims. These were investigations of JA and CU of preservice teachers in different cue availability conditions and the investigation of the relationship between JA and CU of preservice teachers and their knowledge about students' misconceptions. A quantitative research design was adopted to investigate the targeted phenomena.

A within-subject design was used for making comparisons of preservice teachers' JA about students' performance in rational number tests under three conditions. These conditions were student-only, answer-only, and student+answer. In the student-only condition, preservice teachers could access only information about students' effort and extraversion, interest in math, and general success in math. In the answer-only condition, preservice teachers could access only the answers of students for a test related to rational numbers similar to the one they were making judgments for. In the student+answer condition, preservice teachers could access both information presented in the student-only and answer-only conditions. For cues that were available in two conditions, i.e., either in both student-only and student+answer conditions or answer-only and student+answer conditions, comparisons of CU for different conditions were carried out.

A correlational design was also adopted for investigating the third, fourth, and fifth research questions. As a part of a complex correlational design, multiple regression was used to explore the relationship between preservice teachers' CU and their JA. Three multiple regression analyses were conducted among a mean score of JA and the mean scores of CU for 3 conditions.

As part of the correlational design, correlation values were calculated to interpret the relation between preservice teachers' knowledge about misconceptions and JA. Correlation values were also calculated for interpreting the relation between preservice teachers' knowledge about students' misconceptions and their CU.

3.3. Instruments

In this study, together with the instruments and tools used for data collection, various instruments were used for collecting auxiliary data. These auxiliary data collection instruments were used to collect data about middle school students so that the collected data could be used as student cues during the study. The Google Forms and PCK tests were the main data collection tools that measured preservice teachers' JA, CU, and misconception knowledge in different cue-availability conditions.

3.3.1. Tests for Collecting Data to be used as Student Cues

For this study, four student cues were chosen to be used in the student-only condition. These student cues were students' overall grade in math, extraversion of a student, effort of a student in a class, and interest of a student in mathematics. While data for three of the cues were collected via corresponding scales, students' overall grades in mathematics were obtained from their subject teacher in their school.

3.3.1.1. The Big Five Extraversion Scale. For the students' extraversion cue, data were collected via The Big Five Extraversion Scale, which is a part of the Big-Five Factor Structure (Goldberg, 1992). In this scale, 14 human traits related to being introvert or extravert are presented to participants. The participants are asked to describe themselves in general by indicating the points. The points range from 1, which means the trait is extremely inaccurate for the participant, to 9, which means extremely accurate. The maximum point that a participant can get from this scale is 126. Higher scores correspond to greater levels of extraversion, while lower scores correspond to greater levels of introversion.

3.3.1.2. The Ongoing Engagement Subdomain Scale. For the students' effort cue, data were collected via the Ongoing Engagement Subdomain Scale (IRRE, 1998). In the scale, students' attention to the lessons and how they present their efforts to the schoolwork are measured with five items. The items are rated by students from A (very true) to D (not at all true). These responses are turned into points to constitute a score for student efforts. The maximum point that a participant can get from this scale is 20. Higher scores correspond to greater levels of effort.

3.3.1.3. The Situational Interest Scale. For interest in math cue, data were collected via The Situational Interest Scale developed by Linnenbrink-Garcia *et al.*, (2010). This scale includes 12 items related to maintained situational interest and triggered situational interest. The maintained situational interest was described by Linnenbrink-Garcia *et al.*, (2010) as "the extent to which the material itself was enjoyable and engaging" (p. 747) and triggered situational interest was described as "the presentation of course material that grabbed students' attention" (p. 747). Students are expected to rate each item on a scale from 1 to 5, from not at all true to very true. The maximum point that a student can get from this scale is 50. Higher points are related to higher mathematical interest.

3.3.2. Student Assignments for Collecting Information about Answer Cues for Preservice Teachers' Judgments

The assignments were constructed for assessing students on rational numbers, a topic they learned previously at their school. Both assignments consisted of 10 questions, and students were expected to work individually without requesting extra explanations for the questions. Each question in an assignment had a corresponding question in the other, and these pairs of questions were based on the same learning objective from the Turkish middle school curriculum. The Pearson Product Moment Correlation coefficient between students' scores on these two assignments was calculated as r= 0.80, indicating parallel forms reliability for the assignments.

3.3.2.1. The First Student Assignment. The answers students gave on this assignment were used as cues for the second assignment. The items in this assignment were related to common students' misconceptions about rational numbers. All questions in this assignment were open-ended. Students were asked to show all their operations and solution strategies on paper while finding the answers to each question. The misconceptions related to each question in this assignment were chosen among those reported by Depaepe *et al.* (2015).

3.3.2.2. The Second Student Assignment. In this assignment, 10 questions addressed the same misconceptions. All questions were in multiple-choice form, with four options per question. In some questions, each multiple-answer choice was related to different misconceptions; in others, each multiple-answer choice was related to the same misconception. For each correct answer, students received 1 point, and for each incorrect answer, they did not receive any points. The questions in the second assignment were presented to preservice teachers to collect data related to their JA and CU.

3.3.2.3. Tool for Collecting Information About Judgments and Cue Utilization. Preservice teachers made judgments about each item on the second assignment for nine students' performances by using the cues that were available for them. The collected data from nine middle school students about student cues were summarized and presented with graphs to preservice teachers. The place of the student on the graph whom the preservice teacher judged was demonstrated to her (see Appendix A). Their judgment decisions and CU ratings were collected using a form prepared on Google Forms.

Preservice teachers judged items in the second assignment by indicating whether students had given answers to these items correctly or incorrectly in student-only condition. Also, in answer-only and student+answer conditions, they tried to decide on the answer each student would select since preservice teachers had the opportunity to examine students' answers for the first assignment (see Appendix B). This response by preservice teachers was important for the present study in analyzing their ability to detect and accurately judge students' misconceptions.

On the Google Form, after preservice teachers judged students' performances, they were asked to give scores regarding the extent to which they used each cue, which was the final step of their judgments about a student's performance. In the student-only and answer-only conditions, there were four types of cues that preservice teachers could have given scores from 0, indicating never use of this information, to 3, complete use of this information while making judgments. However, in the student+answer condition, there were six types of cues to be scored by preservice teachers (see Appendix C). Four cues in the student+answer condition, which were performance for each question, overall test performance, misconception, and students' guessing, were the same as the cues that were presented in the answer-only condition. Since preservice teachers had seen students' answers for the first assignment as answer cues in the student+answer condition also, they accessed 4 of the answer cues without any restriction. Thus, all the answer cues were scored by each preservice teacher in the student+answer condition.

However, to decrease the allocated time for the judgment process because of the possibility that preservice teachers could lose their concentration while making judgments if a judgment process took a long time (Oudman *et al.*, 2018), only two of the cues in the student+answer condition were chosen to be presented to preservice teachers among the four cues in the student-only condition. These two cues were interest in math and overall math grade. These cues were chosen according to their diagnosticity, as indicated by the correlation between actual cue values and students' scores for the second assignment. The most diagnostic student cue for this study was found overall grade in math with a correlation value of 0.65, and one of the least diagnostic cues was interest in math with a correlation value of 0.14. Interest in math cue was highly utilized in studies. One high diagnostic and one low diagnostic cues were chosen by taking into consideration how much they were used in studies investigating CU, as suggested by Van de Pol *et al.* (2021b) and Rowan (2022).

3.3.3. Test for Measuring Preservice Teachers' PCK in the Rational Number Domain

Teachers' knowledge about misconceptions of students and the instructional decisions of teachers are two main parts that create their PCK (Depeape *et al.*, 2015). Depeape *et al.* (2015) developed an instrument measuring teachers' CK and PCK about rational numbers. CK and PCK parts have 24 items separately. 12 of the 24 items were related to the fraction domain, while the remaining 12 items were related to the decimal number domain. Both conceptual and operational understanding of students were checked via these items. The test constructed by Depaepe *et al.*, (2015) about teachers' PCK for rational numbers was used as an instrument in this study to measure preservice mathematics teachers' knowledge about students' misconceptions about rational numbers and to look for their decisions for dealing with these misconceptions. However, the scope of this study is to understand teachers' PCK to observe their knowledge about misconceptions. The 24 items related only to PCK were selected and presented to the preservice teachers. Depaepe and colleagues (2015) documented acceptable internal consistency values for CK and PCK items respectively as 0.70 and 0.68.

The scoring of the test was made according to coding guidelines in the original study Depaepe *et al.* (2015). Some of the items in the test required multiple answers since they had sub-questions; one correct answer was sufficient for the remaining items in the test. If a preservice teacher gave a completely correct answer for the item, she had 1 point for this item; but if she gave an incorrect answer, she had 0 points. If a preservice teacher gave a partially correct answer for the items requiring multiple answers, she received 0.5 points. Some middle school students' answers for rational number problems were presented to preservice teachers in the PCK test. These questions in the PCK test asked preservice teachers to explain the reasoning of students' thinking and to decide whether a student's answer was correct and acceptable. If a preservice teacher decided whether the answer was correct or not but did not explain the underlying reasoning of a student's answer, 0 point was given for this item. Indicating students' answers as correct or incorrect without explaining the reasoning was not sufficient to take partial or whole credit for these kinds of items (Depeape *et al.*, 2015).

The internal consistency of the PCK test was calculated for this study after the preservice teachers had completed the test. The Cronbach alpha coefficient was calculated as 0.64, which is slightly below the acceptable value of 0.7 (Pallant, 2016). However, DeVellis (2017) explained that the alpha values between 0.6 and 0.65 are undesirable but close to the acceptable value.

3.4. Procedure

The data collection process and how the data were prepared to make the analyses of this study are explained in detail in this section. The conducted analyses for answering each research question are also elaborated (see Appendix D for the diagram of the procedure of the data collection).

3.4.1. Data Collection

This study consisted of two stages. The first one was conducted with middle school students for the purpose of collecting auxiliary data to be used in the second stage with preservice teachers. Before the onset of data collection, ethics approval was taken from the Ethics Approval Committee of the university where this study was conducted (see Appendix E for the Ethics Committee Approval).

Thirty six students were invited to complete the given scales for collecting data to be used as student cues, and fifteen of them, who were seventh or eighth-graders, were invited to work on mathematical assignments. They were informed that the given assignments would not constitute a part of their instruction or assessment at their school, and their answers would be kept confidential. Students completed the scales and assignments in approximately 40 minutes. This process was conducted by the researcher. Data from 9 students were chosen for data collection with preservice teachers. These nine students were selected according to their mathematics performance in the first assignment: 3 low-performing, 3 medium-performing, and 3 high-performing. This enabled preservice teachers to judge the performance of students from varying performance levels, controlling for any potential confounding effect of this variable.

Data collection with preservice teachers was completed in two stages, one relating to the measurement of preservice teachers' knowledge of students' misconceptions about rational numbers and the other for their JA and CU in different conditions. In the first stage, preservice mathematics teachers were administered the PCK test on paper. This lasted approximately 50-70 minutes.

In the second stage, preservice teachers were invited to make their judgments about students' performances for rational numbers by meeting the researcher. The judgment process with preservice teachers was conducted by the researcher on Zoom with online meetings. They made judgments for students' performances on the second assignment only. While preservice teachers were making judgments about students' performances with rational numbers in the second assignment, the researcher requested them to think aloud so that their progress could be supervised. While preservice teachers were working on the student performance, they marked their judgments on the scales provided to them on Google Forms. They also scored how much they made use of each cue. In each condition, a preservice teacher started the judgment process by examining the given information about related cues. In the student-only condition, the information about 3 students' extraversion, effort, interest in math, and overall grade in math were presented to them. They were asked to make their judgments by considering this information about the cues. When a preservice teacher completed her judgments about students' performances on ten questions in the second assignment, the cue list was shown to them to be scored on how much the cues were used. The same procedure was conducted for each of the three conditions.

3.4.2. Data Preparation for Analyses

The preservice teachers made judgments for 3 students in each condition. Thus, their JA and CU scores were summarized using proper statistics to prepare them for the analyses.

3.4.2.1. Preparation of Preservice Teachers' Judgment Accuracy Scores. Preservice teachers made judgments for students' answers for the second assignment. For each of the ten questions in the second assignment, preservice teachers were asked to make a judgment about whether a student would give the correct answers or not. Teachers judged students' answers as correct or incorrect by deciding which option they thought a student would select. It was expected that teachers should have made an interpretation about their reasoning for the judgment without guessing students' answers in the second assignment by looking at the first assignment. A preservice teacher should have looked at students' operations and solution strategies in the first assignment and tried to estimate students' thinking for the questions for the second assignment. If a preservice teacher indicated that a student would have given a correct answer for a question and if a student had really given the correct answer, she got 1 point for

JA. Alternatively, if she indicated that a student would have given a wrong answer and if a student had given the wrong answer, she got 1 point again. If a preservice teacher indicated that a student would have given the correct answer but a student had given the wrong answer or if she indicated that a student would have given the wrong answer and a student had given the correct answer, she could not get any points. JA scores preservice teachers obtained for a student ranged between a minimum score of 0 and a maximum score of 10. In each condition, there were 3 middle-school students' performance to be judged, so preservice teachers obtained three JA scores, which was followed by the calculation of the mean value for these three JA scores. As a result, three mean values of JA scores were obtained for each preservice teacher for the three conditions.

<u>3.4.2.2. Preparation of JA Scores for Students' Answers Involving Misconceptions.</u> In order to study whether preservice teachers made accurate inferences for the questions for which students' misconceptions were observed, the questions in the first and the second assignments where students displayed the same misconceptions were identified in the answer-only and student+answer condition. Six students in these conditions were found to display the same misconceptions in both assignments for a total of 9 question pairs. Thus, another JA score was calculated for the preservice teachers according to their judgments on these nine questions in the second assignment. In order to obtain this score, preservice teachers were given 1 point if they correctly decided which answer the student might have given to questions that displayed the same misconceptions as the questions in the first assignment. This yielded a maximum score of 9 if a preservice teacher judged the answers accurately on all questions.

<u>3.4.2.3. Preparation of Preservice Teachers' Cue Utilization Scores.</u> For each preservice teacher, CU scores were calculated according to their responses on how much they made use of particular cues for each student. A mean value using the scores obtained for three students was calculated for preservice teachers' CU for each cue in each condition. The scores ranged between 0, indicating no use of this information in their

judgment and 3, complete use of this information in their judgment. Four CU scores for the student-only condition, four CU scores for the answer-only condition, and six CU scores for the student+answer condition were calculated.

3.4.3. Data Analysis

All the analyses were carried out with IBM SPSS Statistics 29.0. Repeated measures Analysis of Variance, t-test, multiple regression, and correlation analyses were performed on SPSS to give answers to the research questions.

3.4.3.1. Analysis of Accuracy Differences Between Conditions. To answer the first research question about the differences in preservice mathematics teachers' JA under different cue-availability conditions, repeated measures analysis of variance (ANOVA) was conducted. Since the same preservice teachers had scores in each of the three cue availability conditions being compared, repeated measures ANOVA was the appropriate inferential statistic (Morgan *et al.*, 2011).

Before conducting the analysis, the assumptions for ANOVA were checked. The dependent variable (i.e., JA) was at the ratio level so the first assumption was satisfied. Other assumptions are normality and sphericity. The normality assumption requires the population of a sample to be normally distributed (Pallant, 2016). To check whether scores were normally distributed or not, the Shapiro-Wilk test can be used since the sample size of the present study was less than 50. If the results of the test turn out to be statistically significant at the level of $p < 0.05$, it can be concluded that the distribution of the scores is not normally distributed. However, if the test is not statistically significant, the scores do not show any difference from a normal distribution. Also, checking kurtosis and skewness values is another way of controlling normality since if their values are between -2 and +2, a normal distribution of scores can be accepted (Field, 2018).

Sphericity, whether there are equal variances and covariances for each group of the within-subject variable or not, is another assumption for ANOVA. Maucly's Test of Sphericity is done to check this assumption. If the sphericity assumption is violated, correction can be done by adjusting the degrees of freedom. If the significance value of Mauchly's test is less than 0.05, the sphericity assumption is violated (Field, 2018). If epsilons meaning the estimates of the degree of sphericity are more than 0.75, the Huynh-Feldt correction is used, but if it is less than 0.75, the Greenhouse-Geisser correction is used (Morgan *et al.*, 2011).

3.4.3.2. Analysis of Cue-Utilization Differences Between Conditions. To answer the second research question about how preservice teachers' CU differs under different cue-availability conditions, paired sample t-tests were performed on SPSS. Six t-tests were conducted since six different types of cues were scored twice by each preservice mathematics teacher in two different conditions. Two of the student cues (interest in math and overall grade in math) were presented to preservice teachers in students-only and student+answer conditions. Four answer cues (performance for each question, overall test performance, misconception, and students' guessing) were presented to them in answer-only and students+answer conditions. Thus, for each of these cues, preservice teachers had two mean scores for two different conditions. Since the scores belonged to the same group of preservice teachers, a paired sample t-test was chosen as the appropriate inferential analysis.

The sub-question of the second research question investigated whether there was a difference between preservice teachers' CU of student and answer cues when both were available in the student+answer condition. Another paired sample t-test was performed to answer this question as well since the scores being compared belonged to the same group.

The assumptions for the paired sample t-test were checked before conducting the tests. The scores to be compared need to be measured at interval or ratio levels using a continuous scale, which was the case for the CU variable. So, this assumption was met. Besides, data needs to be normally distributed for the paired-sample t-test (Field, 2018). This assumption was checked with the same strategies explained in section 3.4.3.1. Since the same participants scored utilization of these 6 cues for this research question, there is an expectancy of the scores being correlated to each other to some degree (Field, 2018). Also, the bootstrapping method is suggested when the mean scores, which are compared to each other by using paired sample t-test, do not show normality because the size of the sample is small. If normally distributed data is not available to show how a sampling distribution looks like, bootstrapping helps to estimate the characteristics of the sampling distribution (Field, 2018). The paired-samples t-tests done to check differences between CU under different conditions in this study were conducted using bootstrapping options in SPSS if the assumptions were violated.

<u>3.4.3.3. Analyses of the Relationship Between JA and CU.</u> The third research question investigates whether there is a relationship between preservice teachers' JA and CU in different cue-availability conditions. For each condition, a preservice teacher has one JA score and 4 or 6 CU scores. Since there is one dependent and more than one independent variable in each condition, multiple regression analysis was appropriate for examining these relationships. For each condition, one multiple regression was performed on SPSS. A standard multiple regression was conducted since the dependent variables were entered into the model simultaneously. The aim was to explore how much variance in the dependent variable, JA scores of preservice teachers, could be explained by the independent variables, i.e., the CU scores of each cue.

Analyses of multicollinearity, singularity, outliers, normality, linearity, homoscedasticity, and independence of errors were carried out in order to check the assumptions. Multicollinearity is about the relationship between the independent variables. If this relationship is explained with a correlation value higher than0.9, there is a threat of having multicollinearity between independent variables. Singularity exists if one independent variable is the combination of others (Pallant, 2016). The multiple regression analysis in SPSS also includes collinearity diagnostics. The tolerance and variance inflation factor (VIF) values are important for checking multicollinearity. The tolerance value shows the degree of variability of the selected independent variable not explained by the other independent variables. The VIF value is found by dividing 1 by the Tolerance value, so it is the inverse of the Tolerance value (Pallant, 2016). VIF value demonstrates whether there is a strong relationship between the selected independent variable and the other independent variables. While the Tolerance value should be above0.1, the VIF value should be below 10 in order not to threaten the multicollinearity assumption (Field, 2018). Removal of one of the independent variables that intercorrelate with the other independent variable from the regression model is suggested for cases where Tolerance and VIF values are not in the recommended intervals (Pallant, 2016).

Outliers can also affect the result of multiple regression, so checking them before performing the regression analysis is necessary. Using the residuals plot is one of the methods for checking the outliers. Outliers are more likely to occur in large samples. If the sample size is smaller than one thousand, the standardized residual values should be between -3.3 and +3.3, showing no outliers (Tabachnick and Fidell, 2013). Mahalanobis distance is another way of checking outliers in data. The critical chi-square value should be decided. These values use the number of independent variables as the degrees of freedom (Pallant. 2016). The maximum value of the Mahalanobis Distance should not exceed the critical value selected according to the number of independent variables in the analysis (Pallant, 2016).

Residuals scatterplots provide an opportunity to check the normality, linearity, and homoscedasticity of the data. The difference between the obtained and the predicted dependent variable is called as residuals. The normally distributed residuals about the predicted dependent variable scores show normality (Pallant, 2016). A skewed distribution of residuals showing no pileup of residuals in the center of the plot at each value in the scatterplot of the residuals indicates a violation of the normality assumption (Tabachnick and Fidell, 2013). Also, the histogram of the data should be symmetrical and closer to the bell shape to indicate the normality. The dots on the normal P-P plot lying approximately on the diagonal line is another sign of normal distribution (Field, 2018).

The straight-line relationship between residuals and predicted dependent variable scores shows linearity (Pallant, 2016). If the scatter plot is curved rather than rectangular, nonlinearity exists. If there is a curvilinear relationship between independent and dependent variables, it can be concluded that there is a relationship between them that is not explained by a linear correlation coefficient. Thus, if the linearity assumption is violated because of this reason, the regression analysis is weakened but not invalid (Tabachnick and Fidell, 2013).

When the variance of residuals about predicted dependent variable scores is the same for all predicted scores, the homoscedasticity assumption is met (Tabachnick and Fidell, 2013, p. 127). If the band enclosing the residuals is close to being equal in width at all predicted dependent variable values, homoscedasticity exists. However, when the band is wider at larger predicted dependent variable values, heteroscedasticity exists. Also, the existence of both skewed and not skewed variables leads to heteroscedasticity. In case of an interaction between independent variables, heteroscedasticity may occur (Tabachnick and Fidell, 2013). The bootstrapping method can be used to generate confidence intervals and p-values again if there is a threat to accepting the homoscedasticity assumption (Field, 2018).

Another assumption of regression, the independence of errors of predictions, can be evaluated through the use of The Durbin-Watson statistic (Tabachnick and Fidell, 2013). The test examines the correlations between errors by checking whether the adjacent residuals are correlated or not. The test statistic can be between 0 and 4. If the value is 2, the residuals are not correlated with each other. However, values under 2 mean there is a positive correlation, while values above 2 mean there is a negative correlation. The number of predictors and observations influence the Durbin-Watson test. To sum up, when the test statistic is under the value of 1 or above the value of 3, this assumption is threatened (Field, 2018).

3.4.3.4. Analysis of the Relationship Between JA and Misconception Knowledge. The fourth research question examined the relationship between preservice teachers' misconception knowledge and their JA. Since there were two variables, correlation analysis was appropriate to define the direction and strength of the relationship between them (Pallant, 2016). Pearson product-moment correlation coefficient was used since both variables were continuous.

Normality and linearity were checked as the assumptions of correlation. The existence of the outliers needed to be checked before conducting the analysis (Field, 2018). The scatterplot was generated to check whether the linearity assumption was violated or not. Also, outliers that have extreme values different from the main cluster of points were studied from the scatterplot. The normal P-P plots were also used to check the normality assumption.

The Pearson correlation coefficients (r) take values between -1 and +1. The negative sign indicates that when one of the variables increases, the other one decreases. The positive sign indicates that both variables increase or decrease. If the r value is closer to -1 or +1, the correlation gets high. If the correlation is 0, this means that there is no correlation between variables (Pallant, 2016).

3.4.3.5. Analysis of the Relationship Between CU and Misconception Knowledge. Preservice teachers gave scores for cues to indicate how much they used those cues while they were judging students' performances. There were 8 cues in total, and six of them were used in two conditions. The fifth research question examined whether there was a relationship between preservice teachers' CU and their misconception knowledge. If a cue was used in two conditions, the mean score of the utilization of these cues was calculated. As a result, there were 8 CU scores to be related to preservice teachers' misconception knowledge.

Correlation analysis was performed to check 8 relationships. The assumptions and properties of the analysis are the same as those explained in Section 3.4.3.4. for the previous research question. Spearman rho correlation coefficient instead of Pearson Product Moment Correlation coefficient was used if the level of measurement for the variables was ordinal. Also, if the data does not meet the normality criteria for using the Pearson correlation, the Spearman rho is chosen to interpret the relationship.

4. RESULTS

The results of the analyses regarding the accuracy of teachers' judgments in different cue availability conditions, the difference between utilization of the types of cues, and the relationship between JA, CU, and preservice teachers' misconception knowledge are presented in corresponding subsections in this section.

4.1. Teachers' Judgment Accuracy

For the first research question, an investigation focused on whether preservice teachers' JA under different cue-availability conditions differed or not. In Table 4.1, descriptive statistics for mean scores of 25 preservice mathematics teachers' JA are presented.

Table 4.1. Descriptive statistics for JA scores of preservice teachers.

Cue Availability Condition	N	Mean	Std. Dev.	Min Score	Max Score
Student-only	25	7.01	0.77	5.67	8.67
Answer-only	25	7.71	0.54	6.33	7.67
Student+answer	25	8.79	0.59	7.67	9.67

Preservice teachers' judgments accuracy scores were measured at the ratio level, so this met one of the assumptions of repeated measures ANOVA. To check normality, the Shapiro-Wilk test was conducted for the three sets of JA scores of preservice teachers. All the significance values for the test were found higher than 0.05, so JA scores for each condition were normally distributed. Also, skewness and kurtosis values for each set of JA scores were between -2 and +2. The sphericity assumption was checked by using Mauchly's Test of Sphericity. This assumption was not violated since the significance of the test was0.06, which is higher than the reference value of 0.05.

The result of the repeated measures ANOVA showed that there was a significant difference between JA scores under different conditions, $F(2,23)= 54.25$, $p= 0.00$, $\eta^2= 0.83$. This difference corresponded to a large effect size as shown by the eta squared value (Pallant, 2016). Post hoc analysis showed that the scores of JA in the student-only condition ($M= 7.01$ $SD= 0.77$) were significantly lower than the scores of JA in the answer-only condition ($M= 7.71$ $SD= 0.54$), JA scores in the answer-only condition ($M= 7.71$ $SD= 0.54$) were significantly lower than JA scores in the student+answer condition ($M= 8.79$ $SD= 0.59$). It can be concluded that judgments in the student+answer condition were the most accurate among all the judgments made under different cue availability conditions.

4.2. Teachers' Cue Utilization

For the second research question, differences between CU in varying cue availability conditions for the six cues were investigated by comparing the means. Each paired sample t-test was conducted with bootstrapping in order to compensate for the small sample size.

Table 4.2 shows the descriptive statistics for 25 preservice mathematics teachers' CU scores for six cues. The CU scores could vary between a minimum score of 0 and a maximum score of 3. In what follows, assumption checks and results of the comparisons of preservice teachers' CU under different conditions will be presented. All CU variables were continuous, so the first assumption for the paired sample t-test was satisfied.

Table 4.2. Descriptive statistics for CU scores of preservice teachers.

Cue Types	Cue Availability Condition	N	Mean	Std. Dev.
Interest in math	Student-only	25	2.07	0.75
	Student+answer	25	1.33	0.85
Overall Grade in math	Student-only	25	2.82	0.31
	Student+answer	25	2.15	0.67
Performance for each question	Answer-only	25	2.91	0.20
	Student+answer	25	2.93	0.24
Overall test performance	Answer-only	25	2.05	0.76
	Student+answer	25	2.12	0.80
Misconception	Answer-only	25	2.43	0.61
	Student+answer	25	2.39	0.76
Students' guessing	Answer-only	25	1.40	0.71
	Student+answer	25	1.36	0.92
Extraversion*	Student-only	25	0.85	0.86
Students' effort*	Student-only	25	1.83	0.85
*These cues were represented to preservice teachers just in student-only condition, so they were not included in the analysis of the second research question.				

4.2.1. Interest in Math Cue

Interest in math cue was available for preservice teachers while they were making judgments in student-only and student+answer conditions. According to Shapiro-Wilk test results, the mean scores of CU for interest in math cue in student-only and student+answer conditions were normally distributed since the significance values were higher than0.05 which was0.06 in student-only cındition and0.24 in student+answer condition, so the normality assumption was satisfied.

The main paired sample t-test result, $t(24)= 6.11$, $p < 0.001$, and represented an effect of $d= 1.22$, indicated that in the student-only condition, the interest in math cue was used significantly more than in student+answer condition with a large effect size (Field, 2018).

4.2.2. Overall Grade in Math Cue

The overall grade in math cue was presented to preservice teachers in student-only and student+answer conditions. Before conducting a paired samples t-test, the normality of the scores for utilization of overall grade in math cue for each condition was checked. The Shapiro-Wilk test results indicated that the scores of CU for overall grade in math cue did not display normality in the student-only condition ($p= 0.00$). However, the mean scores of CU for overall grade in math cue in the student+answer condition did not significantly differ from the normal distribution since $p= 0.06$. Also, while the skewness value of the mean score of CU in the student-only condition was -1.88, the kurtosis value was 2.79 which did not lie between the acceptable threshold values -2 and +2. The skewness and kurtosis values for mean scores of CU in the student+answer condition lay between -2 and +2 without causing a threat to normality. The reason behind the non-normal distributed data in student-only condition for utilization of overall grade in math cue can be considered the small sample size. This indicated the necessity of using bootstrapping while conducting the paired sample t-test for checking CU difference for overall grade in math cue. Figure 4.1 and Figure 4.2 show the histograms of the mean scores of CU for overall grade in math cue in student-only and student+answer conditions, respectively.

Results of bootstrapping for the paired sample t-test demonstrated a statistically significant difference between mean scores of CU for the overall grade in math cue in student-only and student+answer conditions. This difference between the CU scores for the two conditions,0.68, BCa 95% CI [0.44, 0.96], was significant since the confidence intervals did not include zero. Consequently, the overall grade in math cue was significantly used more in the student-only condition ($M= 2.82$, $SE= 0.60$) than in the student+answer condition (M= 2.15, SE= 0.13). Paired sample t-test results and bootstrapping results showed there is a statistically significant difference between CU, so the non-normal distributed data belonging to mean score of CU in student-only condition did not create a considerable threat.

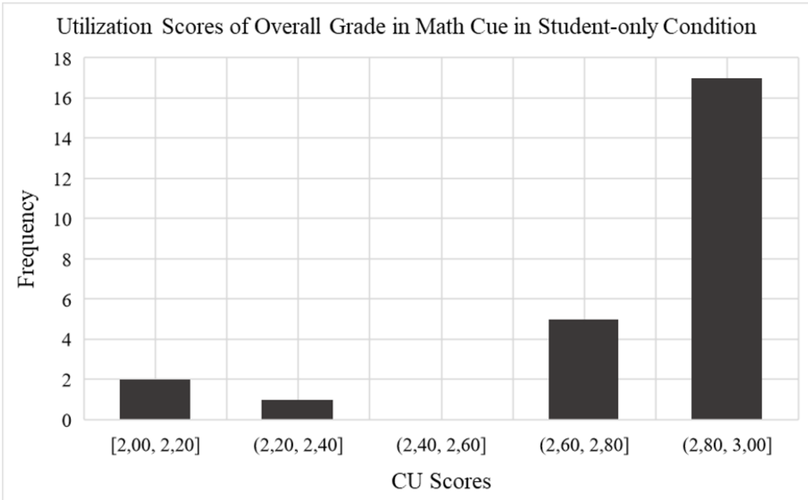


Figure 4.1. Histogram of utilization scores of overall grade in math cue in student-only condition.

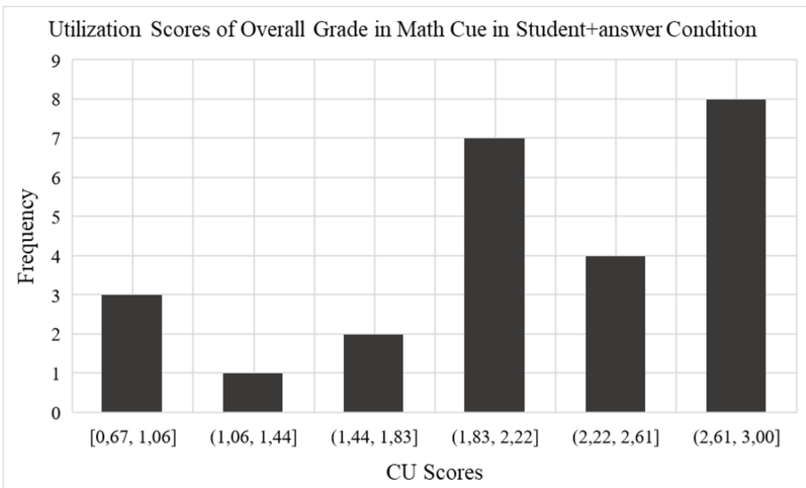


Figure 4.2. Histogram of utilization scores of overall grade in math cue in student+answer condition.

4.2.3. Performance for Each Question Cue

Each preservice teacher observed middle-grade students' answers to ten questions about rational numbers in the first assignment to make judgments about their answers for the second assignment. Since they examined each of the question's answers, performance for each question of middle-grade students was accepted as a cue presented in both answer-only and student+answer conditions. Normality was checked for the mean scores of CU for performance for each question cue in both of the conditions. Figure 4.3 and Figure 4.4 show the histograms of the mean scores of CU for performance for each question cue in answer-only and student+answer conditions, respectively.

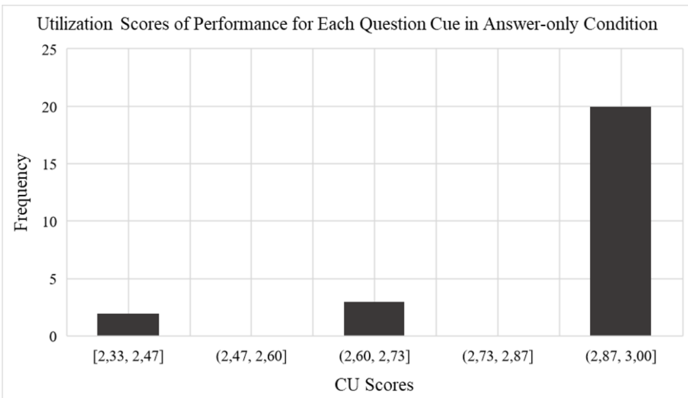


Figure 4.3. Histogram of utilization scores of performance for each question cue in answer-only condition.

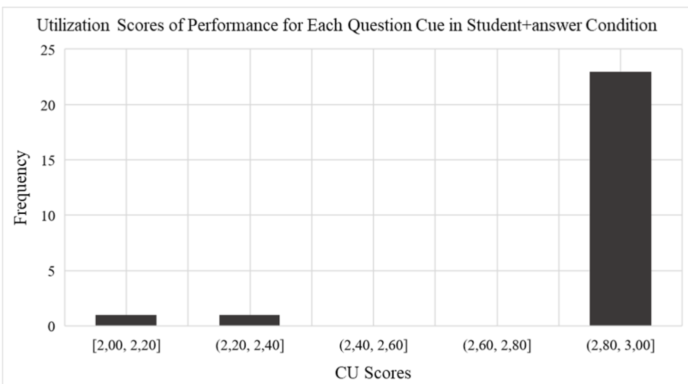


Figure 4.4. Histogram of utilization scores of performance for each question cue in student+answer condition.

According to the Shapiro-Wilk test results, the distribution of the mean scores of CU for this cue did not show normality for these conditions. Significance values for the test were below0.001, which meant that the scores differed significantly from a normal distribution. Additionally, skewness and kurtosis values did not lie between -2 and +2. Hence, bootstrapping was used while conducting the paired samples t-test.

With bootstrapping, the confidence intervals lay between -0.133 and 0.07 containing zero. Thus, the difference between the mean values of CU under the two conditions, -0.03, BCa 95% CI [-0.133, 0.068], was not significant. Hence the results showed that the CU scores for performance for each question cue in the answer-only condition (M= 2.91, SE= 0.04) did not significantly differ from the CU scores in the student+answer condition (M= 2.93, SE= 0.05).

4.2.4. Overall Test Performance Cue

The normality assumption was checked for CU scores for the overall test performance cue like the others. It was observed that the mean scores of CU did not show normality in both of the conditions. The Shapiro-Wilk test was found significant since the p values are lower than 0.05. In the answer-only condition, p= 0.02 and in the student + answer condition p= 0.002. However, skewness and kurtosis values were between -2 and +2. It can be concluded that CU scores for overall test performance cue were not normally distributed. In Figure 4.5 and Figure 4.6, the histograms of the CU scores for overall test performance cue in answer-only and student+answer conditions can be found. Due to the violation of the normality assumption, bootstrapping was used for the paired samples t-test. However, the bootstrapped paired samples t-test did not demonstrate any statistically significant differences between the CU in the two conditions. The difference, -0.06, BCa 95% CI [-.266,0.121] was not significant, which meant that the scores of CU for overall test performance cue in the answer-only condition (M= 2.05, SE= 0.15) were not significantly different from the scores of CU in the student+answer condition (M= 2.11, SE= 0.16).

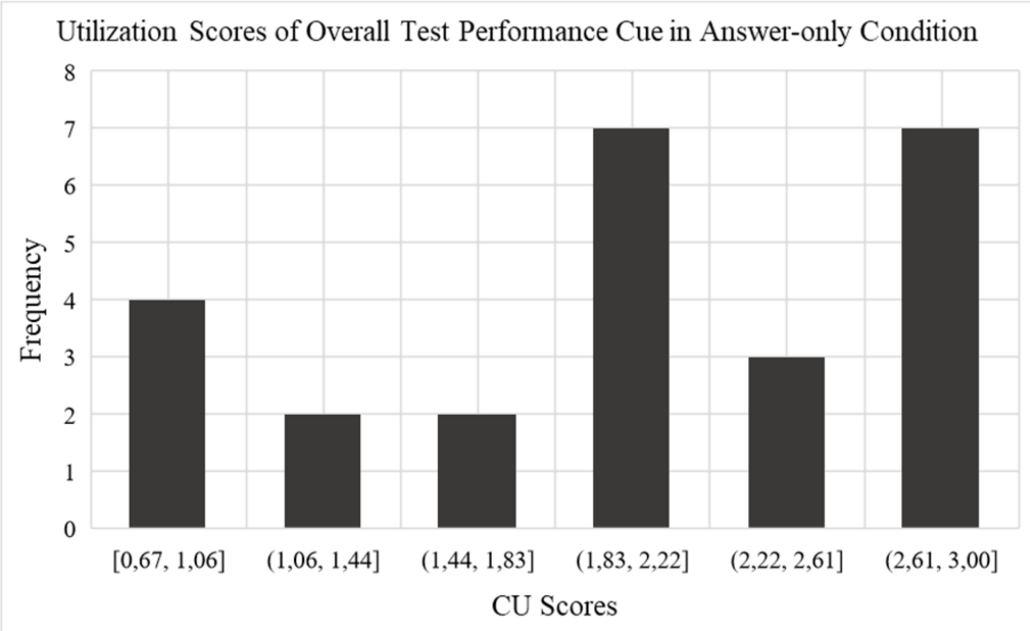


Figure 4.5. Histogram of utilization scores of overall test performance cue in answer-only condition.

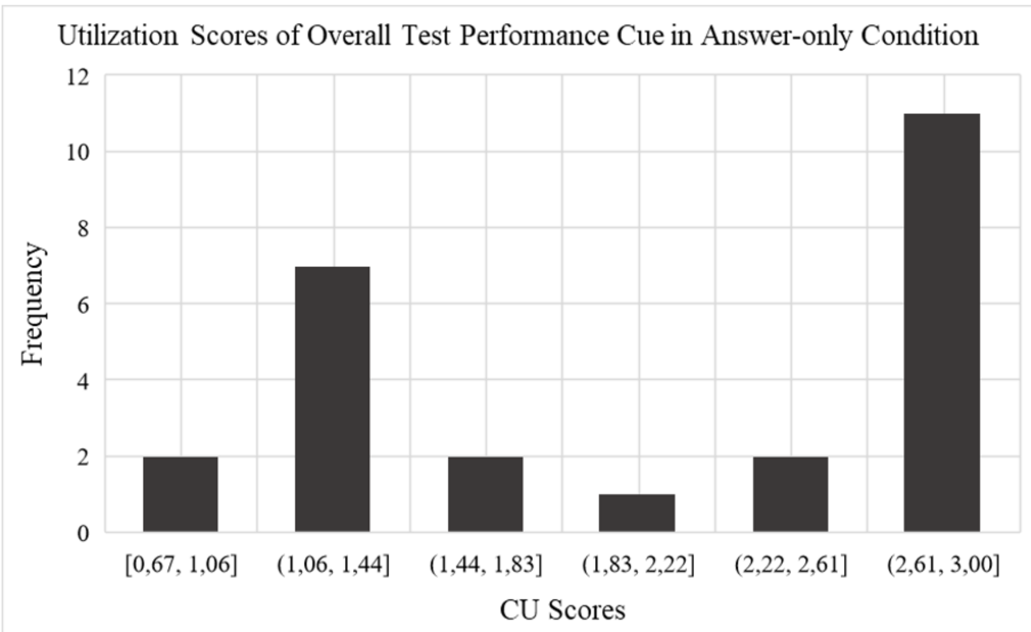


Figure 4.6. Histogram of utilization scores of overall test performance cue in student+answer condition.

4.2.5. Misconception Cue

The misconception cue was available to preservice teachers in answer-only and student+answer conditions. The normality assumption was violated since scores of CU for misconception cue in these two conditions were not normally distributed according to the Shapiro-Wilk test. In answer-only condition $p= 0.002$ and in student+answer condition $p= 0.00$. However, the skewness and kurtosis values were between -2 and +2 in both of the conditions indicating there is a normal distribution. Thus, the histograms (see Figure 4.7 and Figure 4.9) and normal plots were analyzed to see whether mean scores of CU for misconception cue can be accepted as normally distributed data. Since, according to the graphs, the values were not far away from the diagonal lines and the kurtosis and skewness values were within acceptable ranges for the assumptions, the data were considered fit for analysis with the assumptions for normal distribution.

The paired sample t-test yielded a nonsignificant result: $t\ (24) = 0.51$, $p= 0.62$, $d= 0.10$. The CU scores for the misconception cue in the answer-only condition ($M=$ 2.43, $SE= 0.12$) did not differ from the CUscores in the student+answer condition ($M= 2.39$, $SE= 0.09$).

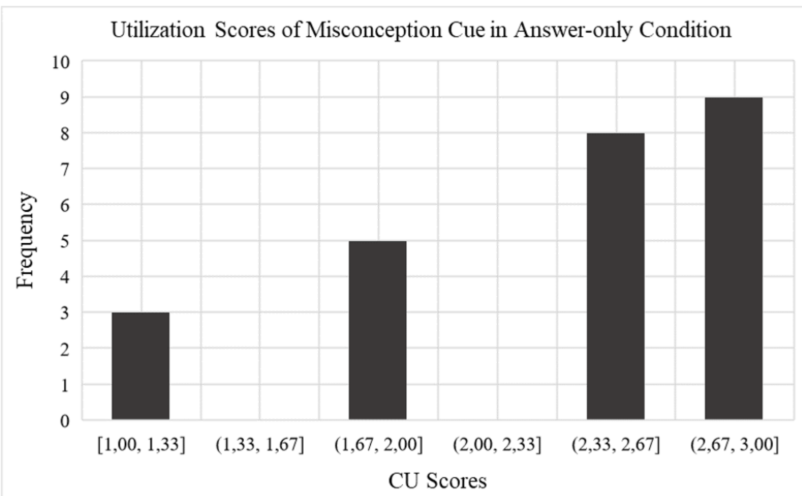


Figure 4.7. Histogram of utilization scores of misconception cue in answer-only condition.

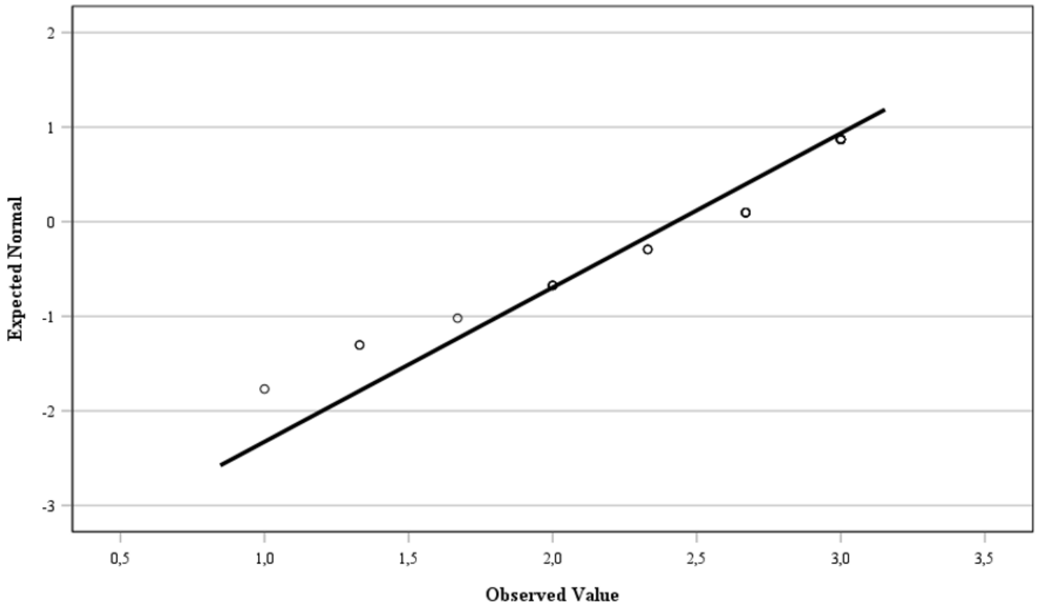


Figure 4.8. Normal q-q plot of utilization scores for misconception cue in answer-only condition.

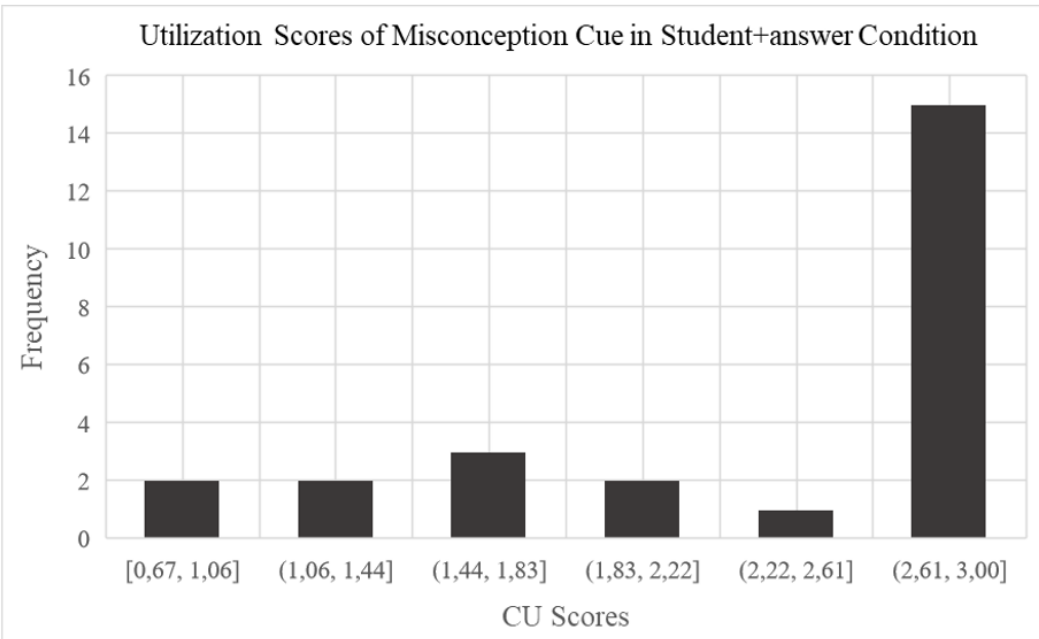


Figure 4.9. Histogram of utilization scores of misconception cue in student+answer condition.

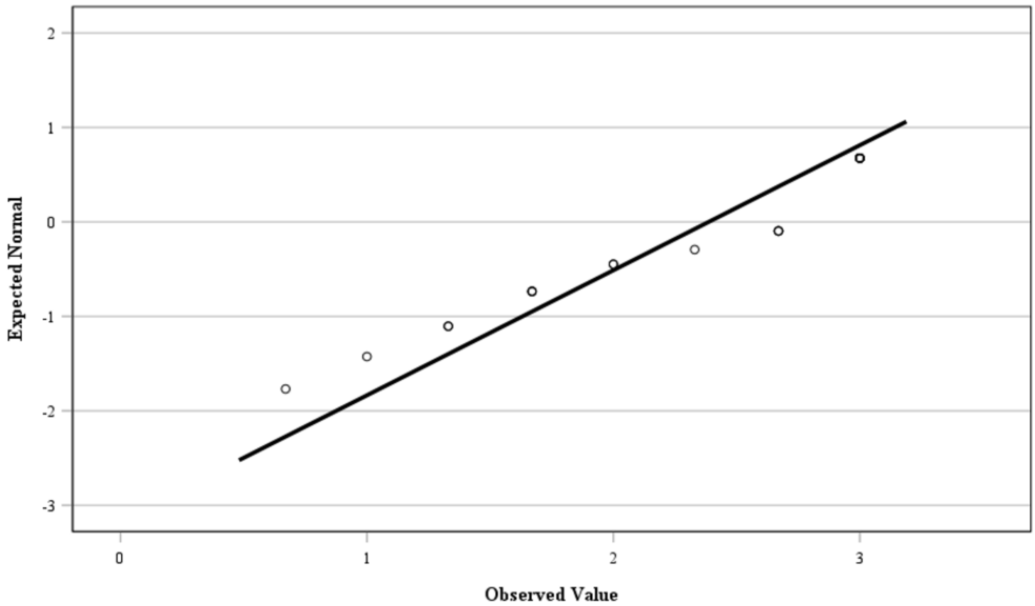


Figure 4.10. Normal q-q plot of utilization scores for misconception cue in student+answer condition.

4.2.6. Students' Guessing Cue

Students' guessing cue was provided in the answer-only and student+answer conditions. The CU scores for students' guessing cue in the answer-only condition showed normality with a p-value of0.44, whereas CU scores for the student+answer condition did not show normality with a p-value of0.003 according to Shapiro-Wilk test results. However, the skewness and kurtosis values of the scores of CU for students' guessing cue in the student+answer condition were 0.80 and -0.51, respectively. Thus, the normal distribution of the data can be accepted. The histogram and the normal plot for the scores in the student+answer condition can be seen in Figure 4.11 and Figure 4.12.

According to the paired-sample t-test, the result was nonsignificant: $t(24)= 0.26$, $p= 0.80$, $d= 0.05$. The CU scores for students' guessing cue in the answer-only condition ($M= 1.40$, $SE= 0.14$) were not significantly different from the CU scores for the student+answer condition ($M= 1.36$, $SE= 0.18$).

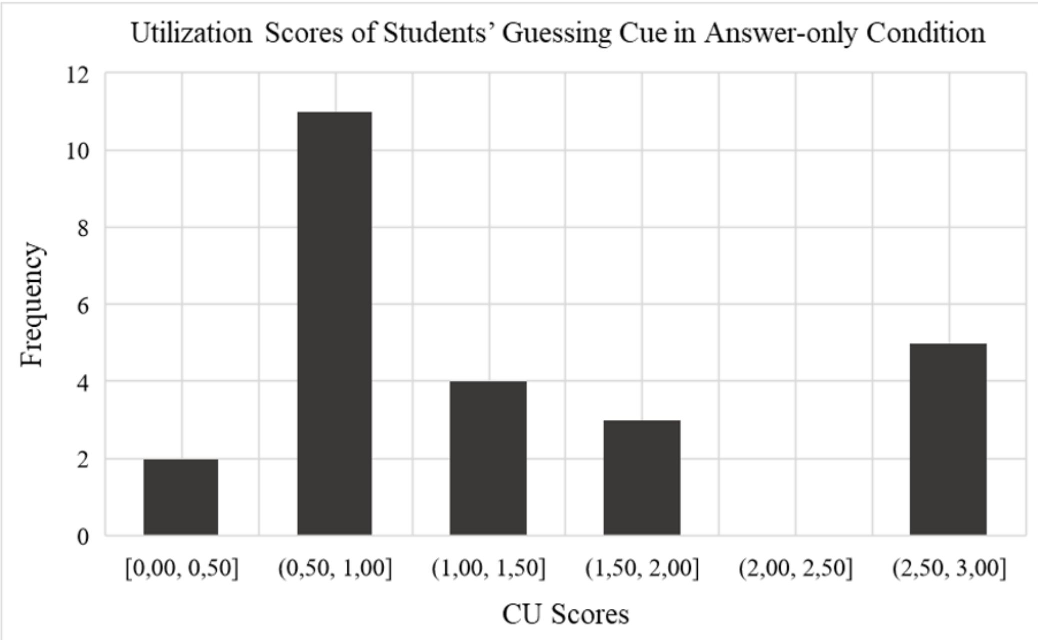


Figure 4.11. Histogram of utilization scores of students' guessing cue in answer-only condition.

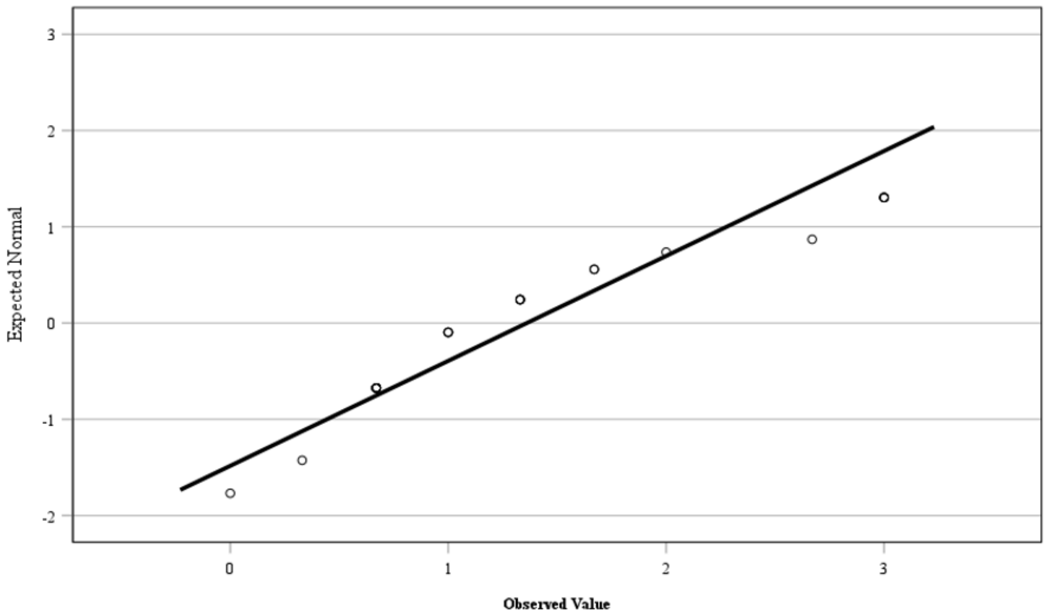


Figure 4.12. Normal q-q plot of utilization scores for students' guessing cue in student+answer condition.

4.2.7. CU Differences Between Student and Answer Cues

Two student cues and four answer cues were presented to preservice teachers in the student+answer condition, as explained above. The mean scores of CU per cue for the two groups of cues (i.e., one for student cues and one for answer cues) were calculated in order to obtain comparable measures. After that, the assumptions for normality were checked. The Shapiro-Wilk test results showed that both mean scores of CU were normally distributed since p= 0.58 for mean scores of student cues and p= 0.29 for mean scores of answer cues.

Paired sample t-test yielded a significant result $t(24)$= -3.16, p= 0.004, with a medium effect size d= -0.63 (Field, 2018). This result showed that student cues (M= 1.73, SE= 0.13) were used significantly less than the answer cues (M= 2.2, SE= 0.05) in the student+answer condition.

4.3. Teachers' Judgment Accuracy and Cue Utilization

Three multiple regression analyses were conducted to check how CU scores were predictive of JA scores of preservice teachers under different cue-availability conditions.

4.3.1. Teachers' JA and CU in Student-only Condition

In the student-only condition, four student cues were presented to preservice teachers for use while they were making judgments about students' performances. These four cues were students' extraversion, students' effort, interest in math, and overall grade in math. In Table 4.3, the descriptive statistics for the variables were presented.

Table 4.3. Descriptive statistics for JA and cu scores of preservice teachers in student-only condition.

Student-only Condition		N	Mean	Std. Dev.	Min. Score	Max. Score
Preservice Teachers' JA		25	7.01	0.77	5.67	8.67
Preservice Teachers' CU	Students' extraversion	25	0.85	0.86	0.00	3.00
	Students' effort	25	1.83	0.85	0.00	3.00
	Interest in math	25	2.07	0.75	0.00	3.00
	Overall grade in math	25	2.83	0.31	2.00	3.00

Before conducting the multiple regression analysis, the assumptions were checked. To check the multicollinearity assumption, correlations between independent variables were investigated. The maximum value for the Pearson Product Moment Correlation coefficient was found as 0.56. This correlation was between the interest in math cue and students' effort cue. Since this value was under0.9, it was not a signal of multicollinearity. The Tolerance and VIF values were also checked. The Tolerance values were not below0.1, and VIF values did not exceed 10, so there was no multicollinearity.

According to the Scatterplot of Residuals (see Figure 13), the standardized residual values were between -3.3 and +3.3, which indicated that no outliers existed in the data. Since the number of independent variables was four for this regression analysis, the critical value of the Mahal. The distance was calculated as 18.67, according to Pallant (2016). The maximum Mahal. Distance in this analysis was 10.04. This result also demonstrated that there were no outliers in the student-only condition. The Normal P-P plot did not show any major deviations from normality since the dots on the plot lie approximately on a straight diagonal line. The histogram was approximately bell-shaped. The scatterplot was not curved shape, so there was no threat of nonlinearity. As it was explained in the Data Analysis section, the band enclosing the residuals on the standardized residual plot got wider at larger predicted dependent values. Thus, the homoscedasticity assumption was under threat. Use of a robust regression was suggested for such cases so bootstrapping was also performed for this analysis (Field, 2018). The Durbin-Watson test was applied to check errors of prediction assumption and the result of a test statistic was equal to 2.08 which was very close to the desirable value 2. This assumption was not violated.

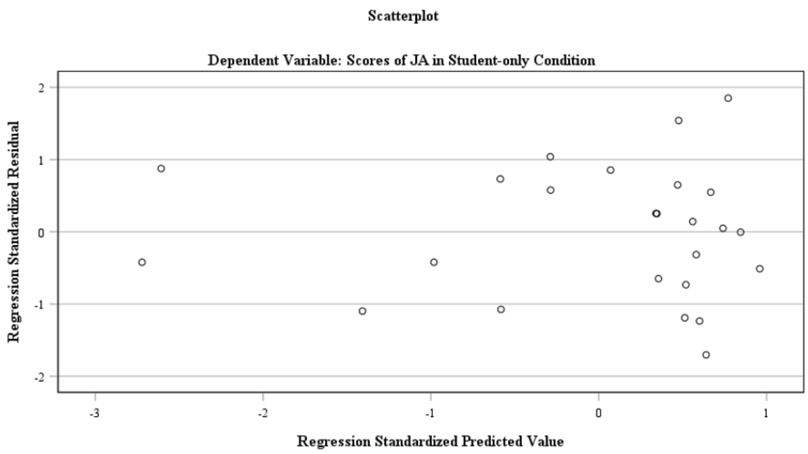


Figure 4.13. Scatterplot of residuals for scores of JA in student-only condition.

Multiple regression was conducted to predict JA by using the CU scores for the four cues in student-only condition. The model including all student cues did not significantly predict preservice teachers' JA scores, $F(4,20)= 1.59$, $p= 0.22$, $R^2= 0.09$. It was suggested that the Adjusted R Squared should be considered when a small sample size exists (Pallant, 2016). The adjusted R Squared which is 0.09 showed the corrected value for R Squared to generalize it better for a population (Pallant, 2016). The Adjusted R Squared was found to be0.09 meaning 9% of the variance of JA scores of preservice teachers in student-only condition can be predicted from their CU.

It can be explained that the Adjusted R Squared is lower than the unadjusted version because other independent variables have little predictive value for the dependent variable. The coefficients showed that just the overall grade in math cue significantly contributed to the regression equation with the largest beta value, which is 1.23 (t= 2.49, p=0.02). Thus, it can be concluded that the utilization of the overall grade in math cue significantly predicted the dependent variable which is JA in the student-only condition in that case. Beta values for all student cues are shown in Table 4.4.

Table 4.4. B values for student cues in student-only condition.

Cues	B values
Students' extraversion	-0.08
Students' effort	0.05
Interest in math	-0.05
Overall grade in math	1.23

Bootstrap results did not rely on normality or homoscedasticity assumptions; the b values were calculated for each predictor to give an accurate estimate of a population. The coefficients did not show any changes from the previous values, which indicates that the assumptions were not violated. The overall grade in math cue was again a significant predictor of the model.

4.3.2. Teachers' JA and CU in Answer-only Condition

In the answer-only condition, four answer cues were scored by preservice teachers for their CU, which were performance for each question, overall test performance, misconception, and students' guessing. The multiple regression analysis was conducted to observe whether use of these cues could predict JA scores in the answer-only condition. Descriptive statistics for the variables in the answer-only condition are shown in Table 4.5.

Table 4.5. Descriptive statistics for JA and CU scores of preservice teachers in answer-only condition.

Answer-only Condition		N	Mean	Std. Dev.	Min. Score	Max. Score
Preservice Teachers' JA		25	7.71	0.54	6.33	8.67
Preservice Teachers' CU	Performance for each question	25	2.91	0.20	2.33	3.00
	Overall test performance	25	2.05	0.76	0.67	3.00
	Misconception	25	2.42	0.61	1.00	3.00
	Students' Guessing	25	1.40	0.71	0.00	3.00

The multicollinearity assumption was not violated since correlations between independent variables did not exceed the critical value 0.9. The Tolerance values were not below 0.1, and VIF values did not exceed 10, indicating there was no threat for multicollinearity.

According to the Scatterplot of Residuals (see Figure 4.14), it can be observed that the standardized residual values were between -3.3 and +3.3. The Mahal. Distance should be 18.67 like the previous multiple regression analysis, and it was equal to 10.46 in the answer-only condition. Therefore, there were no outliers for this condition. The dots on the normal P-P plot lay approximately on a straight diagonal line, and the histogram was close to a bell shape. The distribution of the values did not show major deviations from the normal distribution. The residuals on the scatterplot were nearly rectangularly distributed. The residuals were close to the zero line generally. Errors of prediction assumption was not violated since the Durbin-Watson test was between the critical values.

Multiple regression analysis gave the following result: $F(4,20) = 0.61$, $p = 0.6$, $R^2 = -0.07$. It can be explained that the independent variables did not constitute a combination that significantly predicted the dependent variable. The Adjusted R Square, which was -0.07, was taken into consideration since the sample was small. When the coefficients table was investigated, the largest beta value was 0.27, belonging to the misconception cue ($t = 1.28$, $p = 0.21$). However, the values showed that the misconception cue did not significantly contribute to the regression equation in the answer-only condition. There was no independent variable that significantly predicted the JA scores of preservice teachers according to coefficient values. Beta values for all answer cues are shown in Table 4.6.

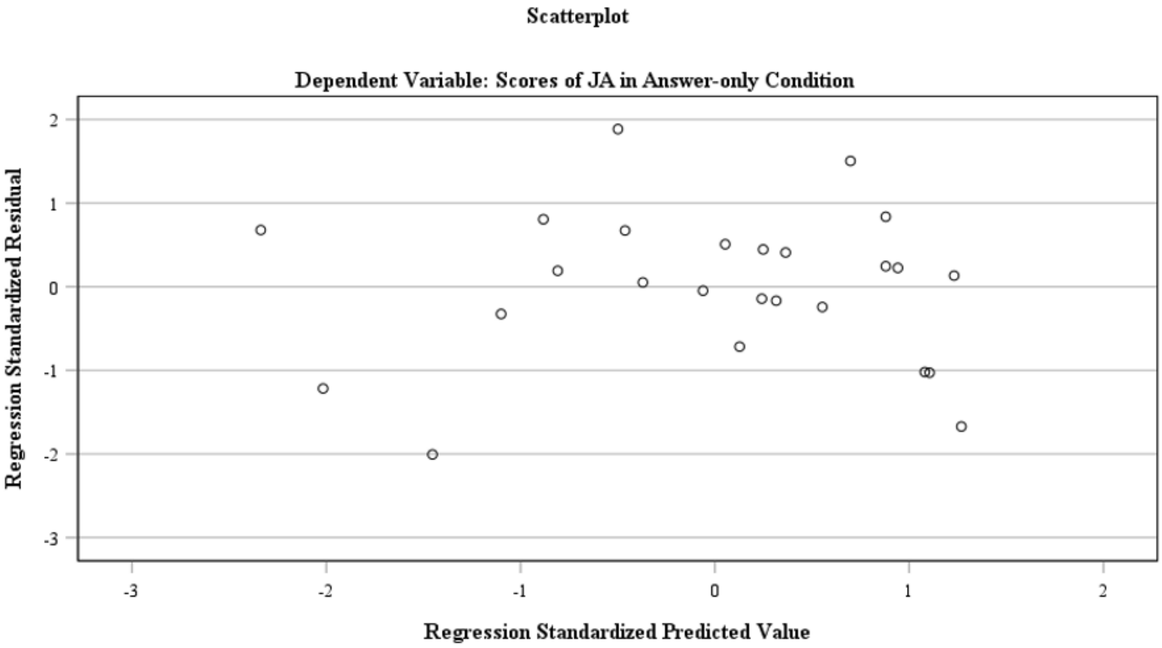


Figure 4.14. Scatterplot of residuals for scores of JA in answer-only condition.

Table 4.6. B values for answer cues in answer-only condition.

Cues	B values
Performance for each question	0.03
Overall test performance	-0.10
Misconeption	0.27
Students' guessing	0.10

4.3.3. Teachers' JA and CU in Student+Answer Condition

In the student+answer condition, two of the student cues and all answer cues were able for preservice teachers' utilization. These cues were interest in math, overall grade in math, performance for each question, overall test performance, misconception, and students' guessing. The multiple regression was conducted to investigate how much variance in JA scores of preservice teachers can be explained by CU of them in student+answer conditions. The descriptive statistics for this condition are presented in Table 4.7.

Table 4.7. Descriptive statistics for JA and CU scores of preservice teachers in student+answer condition.

Student+Answer Condition		N	Mean	Std. Dev.	Min. Score	Max. Score
Preservice Teachers' JA		25	8.79	0.59	7.67	9.67
Preservice Teachers' CU	Interest in math	25	1.33	0.85	0.00	3.00
	Overall grade in math	25	2.15	0.67	0.67	3.00
	Performance for each question	25	2.93	0.24	2.00	3.00
	Overall test performance	25	2.12	0.80	0.67	3.00
	Misconception	25	2.39	0.76	0.67	3.00
	Students' Guessing	25	1.36	0.92	0.00	3.00

Correlations between the independent variables were not above 0.9, so the multicollinearity assumption was not violated. Besides, Tolerance values were above0.1, and VIF values were under 10, indicating no violation.

The Scatterplot of Residuals (see Figure 4.15) showed that there were no outliers. All values were between -3.3 and +3.3. In the student+answer condition, there were six independent variables, so the maximum Mahal. Distance should be 22.46 (Pallant, 2016). The maximum Mahal. Distance in this analysis was 16.33, indicating no existence of outliers. The normal P-P plot did not show any major deviations from normality, and the histogram of the values was approximately bell-shaped. Also, the linearity assumption was met. The Residuals on the scatterplot were nearly rectangularly distributed, which was a desirable situation for meeting the homoscedasticity assumption. However, the dots did not concentrate along the center, indicating a threat to the homoscedasticity assumption. Bootstrap was performed to deal with this kind of violation. Errors of prediction assumption was met since the result of the test was desirable.

It can be observed that the model did not predict the JA scores of preservice teachers in the student+answer condition, $F(4,20)= 0.43$, $p= 0.85$, $R^2= -0.17$, which was not significant. The Adjusted R square was -0.17. The independent variables did not significantly predict the dependent variable.

The largest beta value belonged to the interest in math cue, which is 0.25, ($p= 0.26$, $t= 1.16$) not significant again. Thus, the utilization of any cue did not significantly create a variance in the JA of preservice teachers. Beta values for all student and answer cues are shown in Table 4.8.

Because of a small sample size, the Bootstrapping results were also investigated. The beta values or the significance values did not show any differences from the main regression analysis. Thus, the independent variables did not significantly contribute to the regression equation to predict preservice teachers' JA scores in student+answer condition.

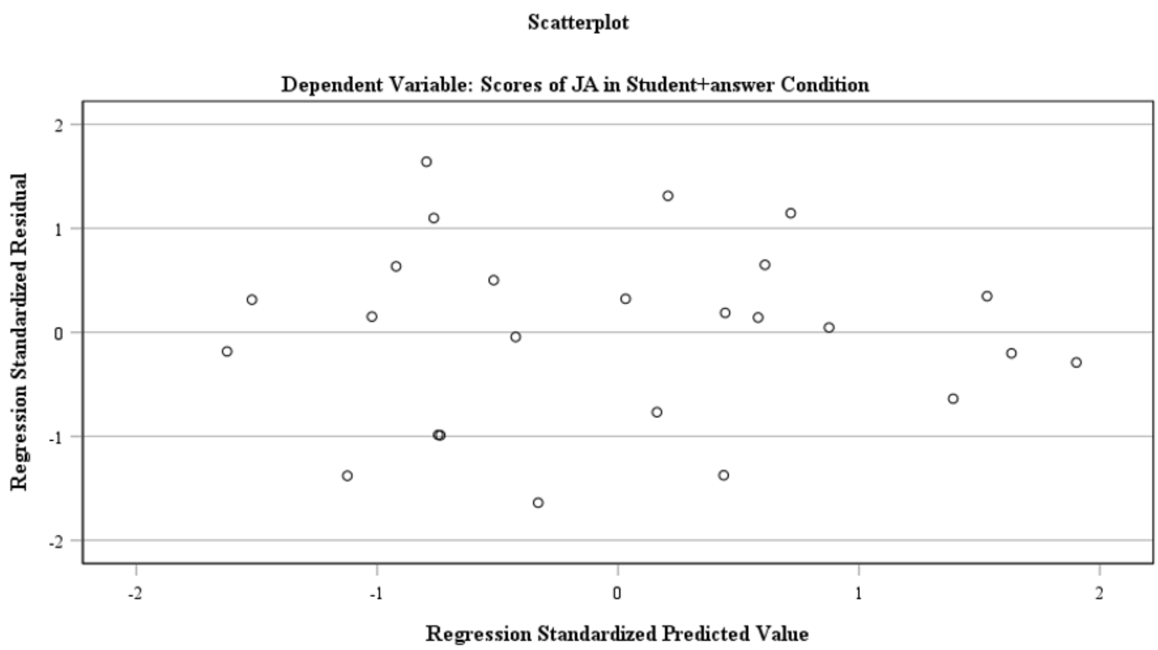


Figure 4.15. Scatterplot of residuals for scores of JA in student+answer condition.

Table 4.8. B values for student and answer cues in student+answer condition.

Cues	B values
Interest in math	0.25
Overall grade in math	-0.18
Performance for each question	0.15
Overall test performance	-0.04
Misconeption	0.10
Students' guessing	-0.22

4.4. The Relationship Between JA and Misconception Knowledge

Correlation analysis was performed to check the relationship between preservice teachers' JA for students' performances involving misconceptions. The normality and linearity assumptions were checked and results showed that the assumptions were met for conducting the analysis. Therefore, the Pearson Product Moment Correlation coefficient was used.

Preservice teachers' accuracy in judging students' answers involving misconceptions and their knowledge of students' misconceptions about rational numbers were not correlated significantly, $r(23)= 0.15$, $p= 0.47$.

4.5. The Relationship Between CU and Misconception Knowledge

Correlation analysis was used to check the relationship between preservice teachers' CU and their knowledge of students' misconceptions about rational numbers. The normality assumptions were checked to decide whether the Pearson Product Moment Correlation coefficient or Spearman Rank Order Correlation coefficient would be used. CU scores for students' effort, interest in math, overall math grade, and students' guessing cues were normally distributed, so the Pearson Product Moment Correlation was calculated for the correlation between these CU scores and misconception knowledge (see Table 4.6). CU scores for students' extraversion, performance for each question, overall test performance, and misconception were not normally distributed. For the correlation between these four CU scores and misconception knowledge, the Spearman Rank Order Correlation coefficients were calculated (see Table 4.7).

The only significant correlation was found between the utilization of misconception cue and preservice teachers' misconception knowledge. The Spearman correlation coefficient r =0.42, p =0.04 which is a small to medium correlation (Field, 2018). The relationship between the utilization of the remaining cues and preservice teachers' misconception knowledge was not found significant.

Table 4.9. Pearson product moment correlation coefficients for correlations between preservice teachers' misconception knowledge and cu (students' effort, interest in math, overall math grade, and students' guessing cues).

Cues	Pearson Product Moment Correlation Coefficient	Sig.
Students' effort	0.17	0.42
Interest in math	0.03	0.90
Overall math grade	-0.14	0.50
Students' guessing	0.28	0.18

Table 4.10. Spearman rank order correlation coefficients for correlations between preservice teachers' misconception knowledge and cu (students' extraversion, performance for each question, overall test performance, misconception).

Cues	Spearman Rho	Sig.
Students' extraversion	0.04	0.86
Performance for each question	0.14	0.49
Overall test performance	0.31	0.13
Misconception	0.42*	0.04*
*Correlation is significant at the 0.05 level.		

5. DISCUSSION AND CONCLUSIONS

This study involved an investigation of preservice teachers' JA-a phenomenon critical for teaching and learning processes-with a particular focus on teachers' CU. A novel contribution of the study is the specific emphasis on a particular topic in mathematics and taking into consideration a key aspect of teacher knowledge about this topic, knowledge of students' misconceptions about rational numbers. How preservice teachers' JA scores vary when different type of cues are available for them, whether CU differed under varying cue availability conditions and whether scores for the utilization of the available cues are related to their JA scores were analyzed. Besides, the relationship between preservice teachers' knowledge of students' misconceptions about rational numbers and both JA and CU variables were investigated. In what follows these results are discussed.

The results showed that preservice teachers' JA showed statistically significant differences under varying cue-availability conditions. It has been observed that preservice teachers made more accurate judgments when both student and answer cues were available for them in comparison with when a single type of cue (i.e., student cues or answer cues) was available. In most of the previous studies (Thiede *et al.*, 2015; Oudman *et al.*, 2018; Van de Pol *et al.*, 2019; Van de Pol *et al.*, 2021b), JA was investigated with in-service teachers. It has been observed in the results of such studies that in-service teachers' existing views about their own students may have impacted their JA negatively. This may be because in-service teachers make judgments by combining their personal thoughts about a student with the available information about her (Artelt and Rausch, 2014). The present study examined preservice teachers' JA differently from the studies investigating in-service teachers' JA. Since preservice teachers did not know the students whose performances they were asked to judge, they could only use the provided cues without integrating any personal thoughts into their judgments. This enabled investigation of use of student cues without potential interference from any previous thoughts about the students. In this way, using both types of cues may have resulted in more accurate decisions about student performances in the present study.

Oudman *et al.* (2018) indicated that teachers made more accurate judgments of what students did not understand without knowing who the student was when a student's answers to problems were presented. They asserted that when teachers made judgments about students' performances in a condition where both student and answer cues were available for them, knowing a student's name interfered with their sufficient use of answer cues. Since teachers made judgments for their own students, knowing the name of a student in the student+answer condition may have affected their JA negatively in contrast to the results of the present study where preservice teachers made judgments for anonymous students. However, Van de Pol *et al.* (2021b) reported similar results to those of the present study. They explained that teachers' monitoring accuracy was higher when performance and student cues were presented to them rather than when only performance cues were presented. In this study, preservice teachers' judgments were less accurate in student-only condition than answer only-condition, but having access to both student and answer cues made their judgments more accurate than the cases in which only one of them were available.

When the scores of JA in each condition are examined, it can be observed that all mean scores of JA were higher than 7. The minimum JA score of a preservice teacher could be 0, and the maximum JA score of a preservice teacher could be 10. Even in student-only condition pre-service teachers could estimate 7 of the answer of students correctly on average. A significant difference between JA scores of preservice teachers was found, but it can be said that preservice teachers who participated in this study were good at judging students' performances under each condition since their mean scores were not under 7. Certain characteristics of preservice teachers, such as their success in their courses or their teaching experience with middle-grade students, can be related to their higher level of JA. Besides, the characteristics of the questions in the judged assignment, prepared by the author, may have affected to judgment process of preservice teachers.

The relationship between CU and JA of preservice teachers was also investigated. Whether the use of diagnostic or non-diagnostic cues leads to differences in preservice teachers' JA is a key issue. According to previous findings, answer cues were considered more diagnostic than student cues (Van de Pol *et al.*, 2021b). Thiede *et al.* (2019) contended that "one of the best predictors of future performance is past performance, that is, past performance is a highly diagnostic cue" (p. 686). The past performance of students in a related topic is considered as a diagnostic cue for teachers to attribute their judgments to students' achievement. Besides, some of the student cues are more predictive of a student's performance than others. Van de Pol *et al.* (2021a) stated that there is a possibility that some of the student cues may be diagnostic to some degree. Thus, they suggested that examining the diagnosticity of student cues before starting a study may be beneficial. The overall grade in math cue was the most diagnostic student cue for this study when the cue values and students' performances were correlated. The students' overall grade in math were accessed by asking their subject teacher at school. The average exam scores of students taken in a school were presented as overall performance of a student. Thus, this cue is the result of the evaluation of a student in general mathematical performance rather than an evaluation of a performance in a specific topic. The overall grade in math cue was chosen to be presented to preservice teachers in student-only and student+answer conditions as a student cue. This cue was not accepted as an answer cue since it represents the level of general mathematical performance rather than the rational number domain, which is a specific domain used to be investigated in this study. However, overall grade in math is the cue that is closest to the diagnostic answer cues because it is somehow related to students' mathematical achievement in general. The CU scores showed that overall grade in math cue was used more than others in student-only condition. Preservice teachers may have thought that overall grade in math cue was more predictive of students' performances without knowing it was the most diagnostic student cue. The student-grade subject cue in the study of Van de Pol *et al.* (2021b), which corresponds to the overall grade in math cue in this study, was not as diagnostic as it is in this study. However, in their study the subject was different; Van de Pol *et al.* (2021b) explored teachers' CU for judging students' text comprehension. It may be concluded that the diagnosticity of the same cue relating to students' general performance for a subject can change according to the subject area. Further investigation may be needed to understand how preservice teachers make decisions about which cue to use and when, since using a cue relating to students' previous subject grades can vary from field to field and according to the specifications of the judged tasks. The overall grade of students in mathematics may be more diagnostic for judgments than in other subject areas if teachers judge students' general understanding in these areas. Van de Pol *et al.* (2021b) claimed that the diagnosticity of the cues can vary from student to student or from task to task and the findings of this study corroborated this claim.

Accessing four answer cues which were thought to be diagnostic cues, and two student cues, one more diagnostic than the other, may have helped the preservice teachers to make more accurate judgments in the student+answer condition. Still, answer cues were used more than student cues in the student+answer condition. The performance for each question cue was used the most in the student+answer condition, whereas the interest in math cue was the least used one. Misconception cue is the second most used cue. It was observed that overall grade in math cue was used more than other answer cues, which are overall test performance and students' guessing cues in the student+answer condition. Preservice teachers may have thought that overall grade in math cue may be as beneficial as other answer cues so they made judgments by considering this cue.

The performance for each question cue and overall test performance cue were used more than the misconception cue and students' guessing cue in both answer-only and student+answer conditions. The availability of the overall grade in math cue, which was highly utilized in student-only condition, may have reduced the utilization of two answer cues that are misconception and students' guessing in student+answer condition. Pit-ten Cate *et al.* (2014) claimed that factors about students' characteristics could influence teachers' judgments, like the factors about academic performance, even though they are explained as non-diagnostic cues in studies. However, their interpretation was made for teachers' JA rather than preservice teachers. Since preservice teachers did not personally know the students they judged, their judgment behaviors in the existence of the student cues and the relation between their use together with the use of answer cues may need further investigation.

The performance for each question cue was used more than other answer cues in the answer-only and student+answer conditions in line with previous findings (e.g. Oudman *et al.*, 2018). This cue was one of the cues that were scored high by preservice teachers for its utilization. The mean scores of CU for this cue in answer only and student+answer condition were 2.91 and 2.93 out of 3 respectively. Performance for each question cue can be thought of as the most informative source for estimating students' answers. Each preservice teacher used this cue in their judgments to a high degree, so the variability for the CU scores was low. The scores varied between 2.33 and 3.00 and 2.00 and 3.00 in answer-only and students+answer conditions, respectively. Preservice teachers in this study may have decided that the examination of students' answers one by one may improve their judgment decisions, so they may have preferred to use performance for each question cue. This high utilization of the performance for each question cue is one of this study's key findings, which shows preservice teachers tried to be as specific as possible in their judgments by using this cue as a reference whenever it was available. The exploration of the utilization of performance for each question cue in more detail to observe its connection with preservice teachers' JA should be researched further. Besides, how the utilization of this cue can be supported for developing preservice teachers' JA when they start actual teaching can be examined.

The result of Oudman *et al.*'s (2018) study indicated that teachers did not prefer to use fewer student cues when they could access both answer and student cues. However, in the present study, it was observed that the use of student cues in the student-only condition was significantly more than their use when answer cues were also available. According to the second research question results, interest in math cue and overall grade in math cue were used significantly more in the student-only condition than student+answer condition. The fact that this study was conducted with preservice teachers receiving teacher training according to contemporary teacher training trends, a focus on students' mathematical thinking rather than correct answers while evaluating student performance might be preferred by preservice teachers. Oudman *et al.* (2018) indicated that teachers' utilization of answer cues did not significantly differ in answer-only and student+answer conditions. A similar finding was obtained in this study that utilization of answer cues in answer-only and student+answer conditions showed nonsignificant differences.

To generate a more general interpretation of the CU of preservice teachers under student+answer condition, CU scores for student and answer cues were compared in addition to examining CU scores of each cue. The use of student cues was significantly lower than the use of answer cues in the student+answer condition. This shows that preservice teachers preferred to attribute their judgment decisions more to answer cues when both student and answer cues were available. However, when student cues were available for them in student+answer condition they still decided to benefit from these cues. The student cues presented to the preservice teacher in the student+answer condition were interest in math and overall grade in math cues. Since both of the student cues were somehow related to students' approaches towards mathematics or abilities for mathematics, preservice teachers may have thought that the use of these cues could make their judgments more accurate.

Besides, test characteristics are among the factors influencing teachers' JA (Südkamp *et al.*, 2014). Using task-specific judgments to improve the JA of teachers was considered as an important issue (Artelt and Rausch, 2014). The present study used assignments measuring students' achievement in a specific mathematical domain rather than their general mathematical ability. Thus, making judgments for each item about a specific topic may have led preservice teachers to examine and use cues by considering their roles and importance in each condition in detail. Judging students' performances in a specific domain rather than judging their general mathematical ability may have supported preservice teachers to select the more diagnostic student cues (even though they did not know the actual diagnosticity of cues) to be used in addition to answer cues. Their considerations underpinning their selection process requires further investigation.

How CU scores predict the JA scores of preservice teachers in each condition was also investigated in this research. It was hypothesized that the JA of preservice teachers might be related to their CU. In the student-only condition, only the overall grade in math cue significantly predicted the JA scores of preservice teachers.

In the answer-only condition, none of the answer cues significantly predicted the JA scores. This situation was unexpected since answer cues were accepted as diagnostic cues since they include information about students' performances. The common method for measuring the utilisation of a cue is using a scale that asks teachers to indicate to what extent they use each cue. (Van de Pol *et al.*, 2021a). The scales from 0 to 3 were presented to each preservice teacher to give scores for utilization of each cue, and the CU scores of answer cues were not lower than 1.4 (the lowest score was obtained for students' guessing cue), showing that preservice teachers perceived that they made use of all cues at least to a certain extent. However, high utilization did not mean that JA scores could be predicted from the combination of cues that preservice teachers utilized. Van de Pol *et al.* (2021b) asserted that teachers sometimes have difficulty interpreting the presented cues. In their study, one of their cues (number of correct relations in students' diagrams) was assumed as diagnostic for the performance of students and was misinterpreted by teachers while making judgments. In the present study, preservice teachers may have incorrectly interpreted students' answers, so they might not have used the presented cue as they thought. In that case, it could be asserted that even if preservice teachers stated that they used the presented cues to a certain degree, some of their misinterpretations about a cue might have led to a non-significant relationship between their JA in the answer-only condition. This study did not focus on how preservice teachers interpreted each cue or the accurateness of interpretation of available cues, but further investigation is required in order to better understand the relationship between utilization of cues and JA.

Similar to the case in answer-only condition, in the student+answer condition, the presented cues did not significantly predict the JA scores of preservice teachers. The existence of no significant relationship between CU and JA of preservice teachers may be attributed to the small sample size. Once again this points to the need for interpretation of cues used by preservice teachers. They may have overestimated or underestimated what extent they used the available cues. In addition to preservice teachers' own scoring for their CU, a methodological enrichment can be introduced by using another measure such as collecting think-aloud data so that experts could observe and interpret the cues being used during judgments in further studies. This would also support the validity of the claims made for CU.

Some of the students' answers in the first assignment involved misconceptions. Preservice teachers' judgments for these questions were separately examined to create another JA score for them. The scores taken from Depeape *et al.*'s (2015) PCK test were an indicator of the misconception knowledge of preservice teachers about rational numbers. A small correlation was found between preservice teachers' JA scores for students' performances involving misconceptions and their misconception knowledge. Preservice teachers may not have observed students' misconceptions in detail since their main aim was not to detect misconceptions of students for this study like in Depeape's (2015) PCK test. Thus, a weak correlation between their JA scores and misconception knowledge may have been found. Depeape *et al.*'s (2015) test was used to measure preservice teachers' misconception knowledge, but the test had two parts measuring both CK and PCK of teachers. They aimed to check whether teachers performed better on the CK part than the PCK part, whether the CK and PCK of teachers were correlated, and check differences between the CK and PCK of teachers at different statuses. The PCK part of the test was applied to preservice teachers in this study for different purposes than theirs (Depeape *et al.*, 2015). This PCK test was developed by Depeape *et al.* (2015) may not have sufficiently measured preservice teachers' misconception knowledge of rational numbers. Other methods may be used to check their knowledge rather than a paper-pencil test, such as assessing their knowledge more qualitatively or presenting assignments including real-classroom situations more (Depeape *et al.*, 2015). Oudman *et al.* (2018) also suggested that measuring teachers' knowledge of students' misconceptions may be important for studies exploring teachers' JA. This is because they assumed that teachers who are knowledgeable about students' misconceptions may be more likely to make accurate judgments. However, the present study showed there is a small correlation between JA and misconception knowledge. The rational number topic is difficult for students to understand, and it includes many types of misconceptions. Preservice teachers may not have been capable of detecting each type of misconception when they see answers from real students. On the other hand, preservice teachers may have detected misconceptions of students while making judgments, but they might not have explained the possible reasons behind students' misconceptions on the PCK test. Turnuklu and Yeşildere (2007) explained that preservice teachers sometimes could not identify why students' misconceptions occurred. While preservice teachers were making judgments, they were just asked to judge students' answers without making any explanation about potential misconceptions. In Depeape *et al.*'s (2015) PCK test, preservice teachers were asked to explain why students had a particular misconception. Even if preservice teachers correctly detected a misconception in the given question on the PCK test, they could not gain any points for a question if they did not provide a satisfactory explanation. Their lack of explanations might have led them to gain low scores from the PCK even if they detected the existence of misconception in the question of the PCK test. However, preservice teachers may have gained higher JA scores for students' answers involving misconceptions since this did not require them to explain any reason. The complex relationships among knowledge about student misconceptions (both diagnosing and explaining any potential student misconception) and their JA needs further investigation.

Preservice teachers' CU and their misconception knowledge were also hypothesized to be related. However, the results showed that only the misconception cue was significantly related to their misconception knowledge. Finding a relationship between students' misconception knowledge and use of misconception cue was not surprising since it meant that preservice teachers who had more knowledge about students' misconception used misconception cue more in their judgments. However, other answer cues like performance for each question and overall test performance may be expected to be related to teachers' misconception knowledge. This is because students' answers involved some misconceptions, and this affected firstly their performance for each question and then their overall test performance. Also, not having a significant relationship between student cues and misconception knowledge of preservice teachers may be thought of as an expected situation. Preservice teachers used student cues without thinking of any possible misconceptions of students in student-only condition. Besides, the negative correlation between student cues and misconception knowledge may have been expected since preservice teachers who were more knowledgeable about misconceptions of students may have used fewer student cues, especially in student+answer conditions

6. LIMITATIONS

The small sample size may be considered as the one of the limitation of this study. If further studies are carried out with larger samples, their results may give more accurate information about the JA and CU of preservice teachers. Since preservice teachers were accurate while judging students in each condition by estimating at least 7 of the students' answers correctly on average, more information about preservice teachers' characteristics can be collected to relate them with their level of judgment accuracy in each condition. Also, preservice teachers' knowledge of misconceptions about rational numbers may need to be measured via different methods or an instrument which is created for only measuring the misconception knowledge of teachers. Since the part of a test prepared by Depeape *et al.* (2015) to measure both CK and PCK of teachers (2015) was used, the reliability of it did not reach a desirable value. Another limitation of this research may be the method for examining preservice teachers' CU. Even preservice teachers report to what extent they used the presented cues, their rating on a scale for utilization of a cue may not have reflected the real utilization degree. Collecting and analyzing think-aloud data for examining CU in addition to preservice teachers' own scoring may be done in further studies.

6.1. Implications

This study selected preservice teachers who will work with their own students in a real classroom environment as participants. They made judgments about students' performances on a difficult and important mathematics topic: rational numbers. They had a chance to use different types of information while making judgments, so they observed that using answer cues while making judgments in a specific domain of mathematics was beneficial.

They may give emphasis to the use of information related to students' performances rather than general information about them in their further professional lives.

Investigation of types of information about students used in the judgment process and their relation with JA and CU of teachers may result in more extensive and purposeful integration of them in teacher education.

Preservice teachers tried to interpret students' answers to judge their performances accurately in answer-only and student-answer conditions. Thus, this study also includes an understanding of students' thinking styles while solving questions about the rational number domain. Furthermore, this study aligns with current research focused on noticing of students' mathematical thinking. This alignment is also important for the more comprehensive improvement of research in mathematics education.

REFERENCES

Artelt, C., and T. Rausch, 2014, "Accuracy of Teacher Judgments", *Teachers' Professional Development*, Vol. 1, pp. 27-43.

Behr, M. J., 1981, *Rational Number Ideas and the Role of Representational Systems*, National Science Foundation, Washington, D.C.

Depaepe, F., J. Torbeyns, N. Vermeersch, D. Janssens, R. Janssen, G. O. Kelchtermans and W. Van Dooren, 2015, "Teachers' Content and Pedagogical Content Knowledge on Rational Numbers: A Comparison of Prospective Elementary and Lower Secondary School Teachers", *Teaching and Teacher Education*, Vol. 47, pp. 82-92.

DeVellis, R. F., 2017, *Scale Development: Theory and Applications*, Fourth Edition, Sage, Thousand Oaks.

Dunlosky, J. and K. W. Thiede, 2013, "Four Cornerstones of Calibration Research: Why Understanding Students' Judgments Can Improve Their Achievement", *Learning and Instruction*, Vol. 24, pp. 58-61.

Durkin, K. and B. Rittle-Johnson, 2015, "Diagnosing Misconceptions: Revealing Changing Decimal Fraction Knowledge", *Learning and Instruction*, Vol. 37, pp. 21-29.

Field, A., 2018, *Discovering Statistics Using IBM SPSS Statistics: North American Edition*, London, Sage.

Gabriele, A. J., E. Joram and K. H. Park, 2016b, "Elementary Mathematics Teachers' Judgment Accuracy and Calibration Accuracy: Do they Predict Students' Mathematics Achievement Outcomes?", *Learning and Instruction*, Vol. 45, pp. 49-60.

Goldberg, L. R., 1992, "The Development of Markers for the Big-Five Factor Structure", *Psychological Assessment*, Vol. 4, No. 1, pp. 26-43.

Institute for Research and Reform in Education (IRRE), "Student Effort: Institute for Research and Reform in Education", 1998, Research Assessment Package for Schools (RAPS) Manual for Elementary and Middle School Assessments, Retrieved From http://www.irre.org/publications/research-assessment-package-schools-raps-manual, accessed on May 21, 2023

Irwin, K. C., 2001, "Using Everyday Knowledge of Decimals to Enhance Understanding", *Journal for Research in Mathematics Education*, Vol. 32, No. 4, pp. 399-434.

Isiksal, M., and E. Çakıroğlu, 2011, "The Nature of Prospective Mathematics Teachers' Pedagogical Content Knowledge: The Case of Multiplication of Fractions", *Journal of Mathematics Teacher Education*, Vol. 14, No. 3, pp. 213-230.

Kieren, T. E. 1976, "On the Mathematical, Cognitive and Instructional, In Number and Measurement", *Papers From a Research Workshop*, Vol. 7418491, pp. 101-123.

Koriat, A., 1997, "Monitoring One's Own Knowledge during Study: A Cue-Utilization Approach to Judgments of Learning", *Journal of Experimental Psychology*, Vol. 126, No. 4, pp. 349-370.

Linnenbrink-Garcia, L., A. M. Durik, A. M. Conley, K. E. Barron, J. M. Tauer, S. A. Karabenick and J. M. Harackiewicz, 2010, "Measuring Situational Interest in Academic Domains", *Educational and Psychological Measurement*, Vol. 70, No. 4, pp. 647-671.

Ministry of National Education (MONE), 2018, "Öğretim Programlarını İzleme ve Değerlendirme Sistemi-Öğretim Programları", Retrieved from, 2023, http://mufredat.meb.gov.tr/Programlar.aspx, accessed May 10, 2023.

Morgan, G. A., N. L. Leech, G. W. Gloeckner and K. C. Barrett, 2011, *IBM SPSS for Introductory Statistics: Use and Interpretation*, Fourth Edition, Routledge, New York.

Newton, K. J. 2008, "An Extensive Analysis of Preservice Elementary Teachers' Knowledge of Fractions", *American Educational Research Journal*, Vol. 45, No. 4, pp. 1080-1110.

Ni, Y. and Y. D. Zhou, 2005, "Teaching and Learning Fraction and Rational Numbers: The Origins and Implications of Whole Number Bias", *Educational Psychologist*, Vol. 40, No. 1, pp. 27-52.

Ojose, B., 2015, *Common Misconceptions in Mathematics: Strategies to Correct Them*, University Press of America, Lanham.

Oudman, S., J. Van de Pol, A. Bakker, M. Moerbeek and T. Van Gog, 2018, "Effects of Different Cue Types on the Accuracy of Primary School Teachers' Judgments of Students' Mathematical Understanding", *Teaching and Teacher Education*, Vol. 76, pp. 214-226.

Pallant, J., 2016, *Statistics is a Powerful Statistical Software (SPSS) Survival Manual: A Step By Step Guide to Data Analysis Using SPSS Program*, McGraw-Hill Education, London.

Pit-ten Cate, I., S. Krolak-Schwerdt, S. Glock and M. Markova, 2014, "Teachers' Judgments: Obtaining Change through Cognitive Processes", *Teachers' Professional Development*, Vol. 1, pp. 45-61.

Ready, D. D. and D. L. Wright, 2011, "Accuracy and Inaccuracy in Teachers' Perceptions of Young Children's Cognitive Abilities: The Role of Child Background and Classroom Context", *American Educational Research Journal*, Vol. 48, No. 2, pp.

335-360.

Rowan, E., 2022, *How Do I Know What You Know? Promoting Teachers' Judgement Accuracy of Their Students' Text Comprehension*, M.S. Thesis, Utrecht University.

Shavelson, R. J., 1983, "Review of Research on Teachers' Pedagogical Judgments, Plans, and Decisions", *Elementary School Journal*, Vol. 83, No. 4, pp. 392-413.

Shavelson, R. J. and P. Stern, 1981, "Research on Teachers' Pedagogical Thoughts, Judgments, Decisions, and Behavior", *Review of Educational Research*, Vol. 51, No. 4, pp. 455-498.

Shulman, L. S., 1986, "Those Who Understand: Knowledge Growth in Teaching", *Educational Researcher*, Vol. 15, No. 2, pp. 4-14.

Siegler, R., T. Carpenter, F. Fennell, D. Geary, J. Lewis, J. Okamoto, and Y. Thompson, J. Wray, 2010, "Developing Effective Fractions Instruction for Kindergarten through 8th Grade", *Institute of Education Sciences Practice Guide*, Vol. 1, pp. 10-39.

Stacey, K., S. Helme, V. Steinle, A. Baturo, K. Irwin and J. Bana, 2001, "Preservice Teachers' Knowledge of Difficulties in Decimal Numeration", *Journal of Mathematics Teacher Education*, Vol. 4, 3, pp. 205-225.

Südkamp, A., J. Kaiser and J. Möller, 2012, "Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis", *Journal of Educational Psychology*, Vol. 104, No. 3, pp. 743-762.

Südkamp, A., J. Kaiser and J. Möller, 2014, "Teachers' Judgments of Students' Academic Achievement", *Teachers' Professional Development*, Vol. 2, pp. 5-25.

Thiede, K. W., J. L. Brendefur, M. Carney, J. Champion, L. Turner, R. A. Stewart and R. D. Osguthorpe, 2018, "Improving the Accuracy of Teachers' Judgments of Student Learning", *Teaching and Teacher Education*, Vol. 76, pp. 106-115.

Thiede, K. W., S. Oswalt, J. L. Brendefur, M. B. Carney and R. D. Osguthorpe, 2019, "Teachers' Judgments of Student Learning of Mathematics", *The Cambridge Handbook of Cognition and Education*, Vol. 3, pp. 678-695.

Tirosh, D., 2000, "Enhancing Prospective Teachers' Knowledge of Children's Conceptions: The Case of Division of Fractions", *Journal for Research in Mathematics Education*, Vol. 31, No. 1, pp. 5-25.

Turnuklu, E. B. and S. Yeşildere, 2007, "The Pedagogical Content Knowledge in Mathematics: Pre-Service Primary Mathematics Teachers' Perspectives in Turkey", *Issues in the Undergraduate Mathematics Preparation of School Teachers*, Vol. 3, pp. 1-34.

Van De Pol, J., A. B. H. De Bruin, M. H. Van Loon, and T. Van Gog, 2019, "Students' and Teachers' Monitoring and Regulation of Students' Text Comprehension: Effects of Comprehension Cue Availability", *Contemporary Educational Psychology*, Vol. 56, pp. 236-249.

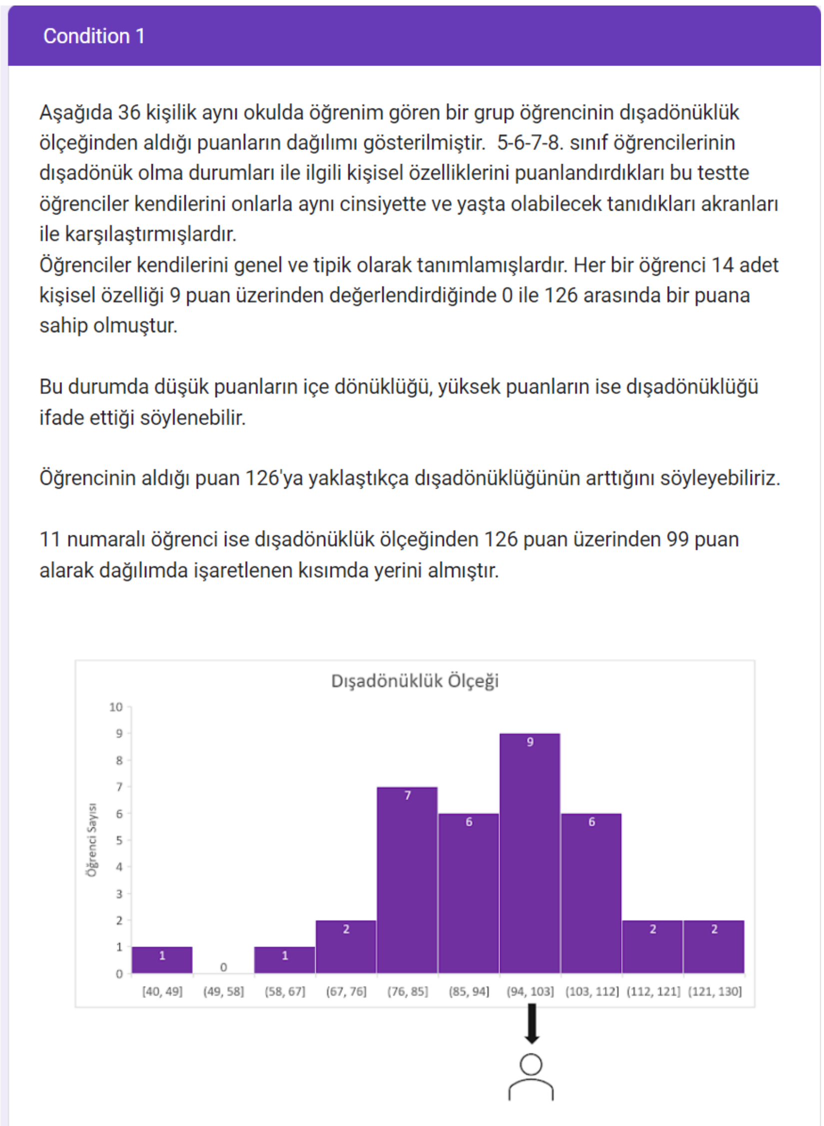
Van de Pol, J., S. N. Van den Boom-Muilenburg and T. Van Gog, 2021a, "Exploring the Relations between Teachers' Cue-Utilization, Monitoring and Regulation of Students' Text Learning", *Metacognition and Learning*, Vol. 16, No. 3, pp. 769-799.

Van de Pol, J., T. Van Gog and K. Thiede, 2021b, "The Relationship between Teachers' Cue-Utilization and Their Monitoring Accuracy of Students' Text Comprehension", *Teaching and Teacher Education*, Vol. 107, pp. 482-506.

Van Hoof, J., L. Verschaffel and W. Van Dooren, 2015, "Inappropriately Applying Natural Number Properties in Rational Number Tasks: Characterizing the Development of the Natural Number Bias Through Primary and Secondary Education", *Educational Studies in Mathematics*, Vol. 90, No. 1, pp. 39-56.

Van de Walle, J. A., K. S. Karp and J. M. Bay-Williams, 2012, *Elementary and Middle School Mathematics: Teaching Developmentally*, Teaching Student-Centered Mathematics Series, Pearson, Upper Saddle River.

APPENDIX A: THE EXAMPLE FOR THE PRESENTATION OF STUDENT CUES TO PRESERVICE TEACHERS ON A GOOGLE FORMS


Condition 1

Aşağıda 36 kişilik aynı okulda öğrenim gören bir grup öğrencinin dışadönüklük ölçeğinden aldığı puanların dağılımı gösterilmiştir. 5-6-7-8. sınıf öğrencilerinin dışadönük olma durumları ile ilgili kişisel özelliklerini puanlandırdıkları bu testte öğrenciler kendilerini onlarla aynı cinsiyette ve yaşta olabilecek tanıdıkları akranları ile karşılaştırmışlardır.
Öğrenciler kendilerini genel ve tipik olarak tanımlamışlardır. Her bir öğrenci 14 adet kişisel özelliği 9 puan üzerinden değerlendirdiğinde 0 ile 126 arasında bir puana sahip olmuştur.

Bu durumda düşük puanların içe dönüklüğü, yüksek puanların ise dışadönüklüğü ifade ettiği söylenebilir.

Öğrencinin aldığı puan 126'ya yaklaştıkça dışadönüklüğünün arttığını söyleyebiliriz.

11 numaralı öğrenci ise dışadönüklük ölçeğinden 126 puan üzerinden 99 puan alarak dağılımda işaretlenen kısımda yerini almıştır.

Figure A.1. The example for the presentation of student cues to preservice teachers on a google forms.

Aşağıda 36 kişilik aynı okulda öğrenim gören bir grup öğrencinin içinde bulunduğumuz eğitim-öğretim yılı boyunca okulda tamamladıkları 6 adet matematik yazılı sınavından aldıkları puanların ortalamasının dağılımı gösterilmiştir. 5-6-7-8. sınıf öğrencileri test, açık uçlu, boşluk doldurma ve doğru-yanlış soru türlerinden oluşan bu sınavlardan ortalamada 0-ile 100 arasında bir puana sahip olmuştur.

Bir öğrencinin puanı 100'e yaklaştıkça matematik dersinden başarısının arttığını söyleyebiliriz.

11 numaralı öğrenci ise matematik yazılı sınavlarının ortalaması alındığında 45 puana sahip olduğu için dağılımda işaretlenen kısımda yerini almıştır.

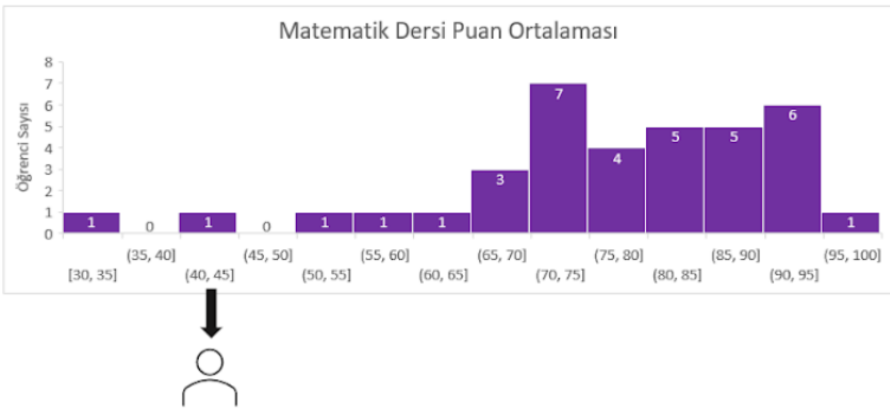


Figure A.2. The example for the presentation of answer cues to preservice teachers on a google forms.

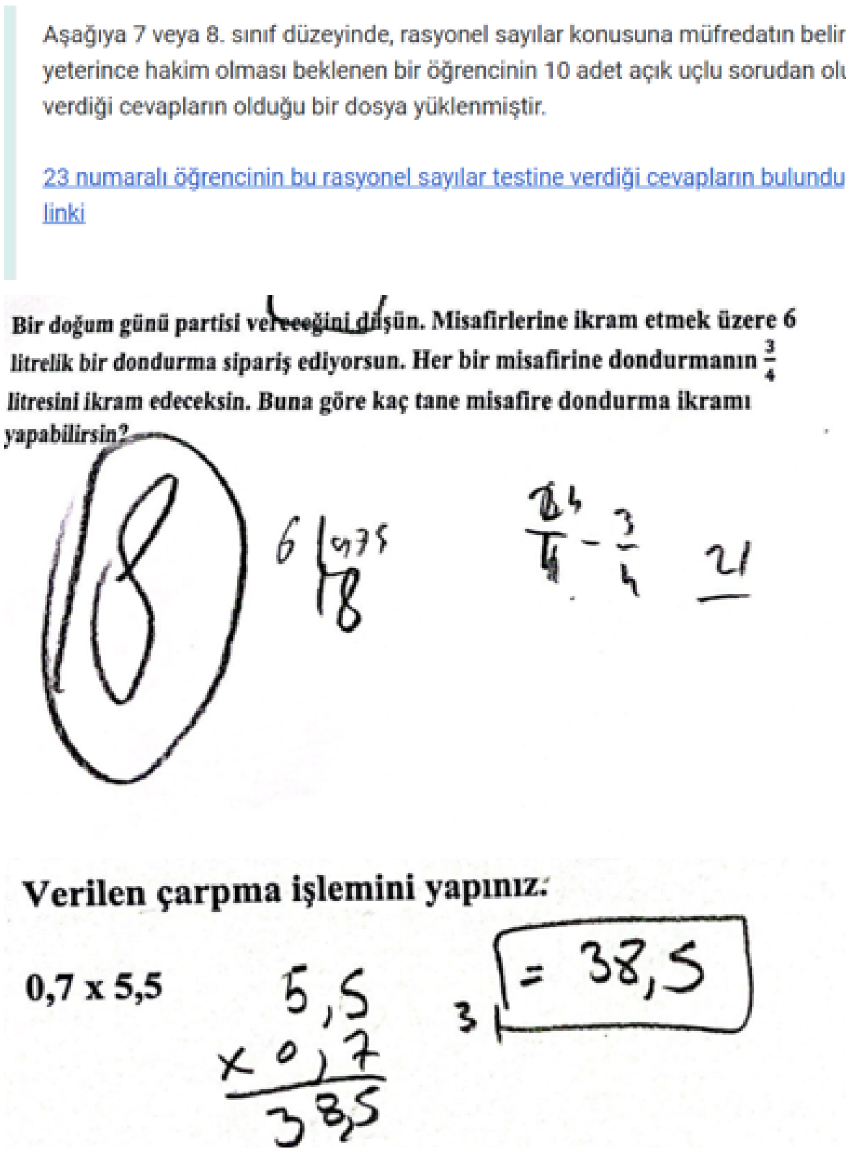
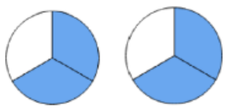


Figure A.3. The example for the presentation of answer cues to preservice teachers on a google forms 1.

APPENDIX B: A SAMPLE ITEM THAT PRESERVICE TEACHERS MAKE JUDGMENTS FOR STUDENTS' ANSWERS IN STUDENT-ONLY CONDITION

DOĞRU CEVAP : B *

1. Aşağıda verilen şeklin tamamı bir bütünü temsil etmektedir. Verilen şeklin taralı olan kısmı aşağıda verilen kesirlerden hangisi ile ifade edilir?



A) $\frac{6}{6}$ **B)** $\frac{2}{3}$ **C)** $\frac{4}{3}$ **D)** $\frac{3}{4}$

○ Doğru cevaplamıştır

○ Yanlış cevaplamıştır

Figure B.1. A sample item that preservice teachers make judgments for students' answers in student-only condition.

DOĞRU CEVAP: B *

5. Bir öğretmen her öğrencisine dağıtmak için 10 litre limonata hazırladı. Her bir öğrencisine limonatanın $\frac{2}{5}$ litresini dağıtacağına göre bu öğretmenin toplamda kaç öğrencisi vardır?
Bu sorunun cevabını bulmak için aşağıda verilen işlemlerden hangisi yapılmalıdır?

A) $10 \times \frac{2}{5}$ **B)** $10 : \frac{2}{5}$ **C)** $10 - \frac{2}{5}$ **D)** $\frac{2}{5} : 10$

- ◯ A işaretlemiştir.
- ◯ B işaretlemiştir.
- ◯ C işaretlemiştir.
- ◯ D işaretlemiştir.

Figure B.2. A sample item that preservice teachers make judgments for students' answers in answer-only and student+answer condition.

APPENDIX C: THE EXAMPLE FOR THE SCALE THAT PRESERVICE TEACHERS GIVE SCORES FOR THEIR CUE UTILIZATION ON A GOOGLE FORMS

Aşağıda sizden 32 numaralı öğrencinin rasyonel sayılar testindeki performansını değerlendirirken bu öğrenci hakkında sizlere verilen bilgileri ve öğrencinin benzer bir teste verdiği cevapları inceleyerek kullanmış olabileceğiniz bilgileri ne ölçüde kullandığınızı belirtmeniz istenecektir.

Verilen bilgileri ne ölçüde kullandığınızı belirtirken 0-1-2-3 puanlarından birini vermeniz gerekmektedir. "0" puan "öğrenci performansını değerlendirirken bu bilgiyi hiç kullanmadım" anlamına gelirken "3" puan "tamamen bu bilgiyi kullanarak değerlendirme yaptım" anlamına gelir.

Figure C.1. The example for the scale that preservice teachers give scores for their cue utilization on a google forms.

Matematik Dersi İçin Durumsal İlgi *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Matematik Dersi Puan Ortalaması *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Öğrencinin ayrı ayrı her bir sorudaki performansı *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Öğrencinin genel test performansı *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Öğrencinin kavram yanılgıları (misconceptions) *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Öğrencinin soruya tahminen veya rastgele cevap vermesi *

0 1 2 3

Bu bilgiyi hiç kullanmadım ○ ○ ○ ○ Tamamen bu bilgiyi kullandım

Figure C.2. The example for the scale that preservice teachers give scores for their cue utilization on a google forms 1.

APPENDIX D: A DIAGRAM FOR DATA COLLECTION PROCEDURE

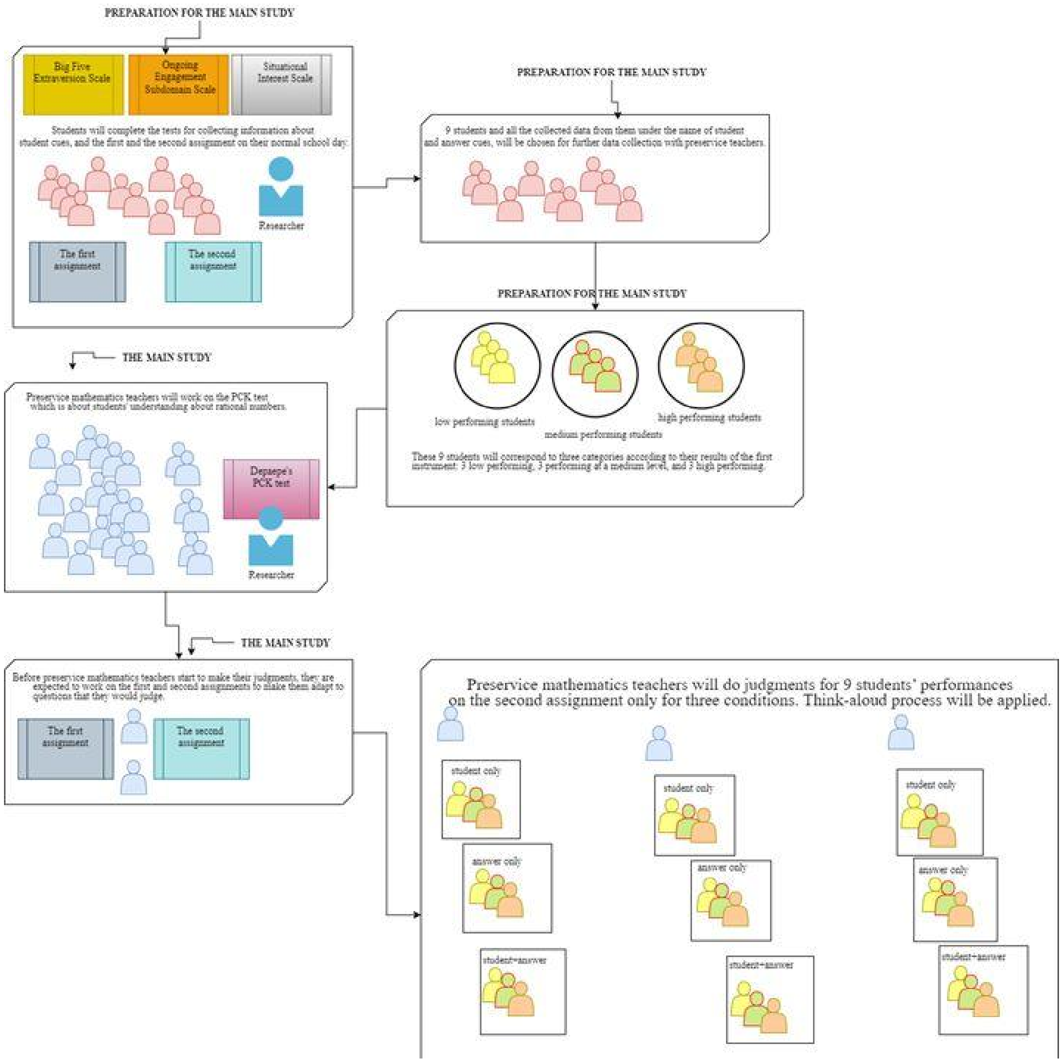


Figure D.1. A Diagram for Data Collection Procedure.

APPENDIX E: ETHICS COMMITTEE APPROVAL

Evrak Tarih ve Sayısı: 14.03.2023-117524

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu
(FMİNAREK)

Sayı : E-84391427-050.01.04-117524 14.03.2023
Konu : 2023/07 Kayıt no'lu başvurunuz hakkında

Sayın Doç. Dr. Nizamettin Engin ADER
Matematik ve Fen Bilimleri Eğitimi Bölüm Başkanlığı - Öğretim Üyesi

"Ortaokul Matematik Öğretmen Adaylarının Öğrencilerin Rasyonel Sayılar Anlayışını Yargılamalarındaki İpucu Kullanımı ve Yargı Doğruluğu" başlıklı projeniz ile Boğaziçi Üniversitesi Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu (FMİNAREK)'e yaptığınız 2023/07 kayıt numaralı başvuru 06.03.2023 tarihli ve 2023/03 No.lu kurul toplantısında incelenerek etik onay verilmesi uygun bulunmuştur.

Bu karar tüm üyelerin toplantıya on-line olarak katılımıyla ve oybirliği ile alınmıştır. Onay mektubu tüm üyeler adına Komisyon Başkanı tarafından e-imzalanmıştır.

Saygılarımızla bilginize sunarız.

Prof. Dr. Tınaz EKİM AŞICI
Başkan

Bu belge, güvenli elektronik İmza ile imzalanmıştır.

Doğrulama Kodu :BS943UBT73 Pin Kodu :19052 Belge Takip Adresi : https://turkiye.gov.tr/ebd?eK=4787&eD=BS943UBT73&eS=117524

34342 Bebek-İstanbul
Telefon No:0212 287 17 53 Faks No:0212 265 70 06
İnternet Adresi:www.boun.edu.tr
Kep Adresi:bogaziciuniversitesi@hs01.kep.tr
Bilgi için: Nurşen MUNAR
Unvan: Mühendis

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Figure E.1. Ethics Committee Approval.