ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 28, NO. 2, 69-99

Pre-service Teachers' Subject Matter Knowledge about Normal Distribution

Nadide YILMAZ
Karamanoğlu Mehmetbey University

ABSTRACT

Many countries do not include the normal distribution concept in their middle school mathematics curriculum, but on the grounds that middle school mathematics teachers need to know more than just mathematics, researchers argue that preservice teachers (PTs) ought to have knowledge and skills in this area. This study was aimed to investigate PTs' subject matter knowledge (SMK) about the normal distribution. Data was collected from 120 PTs attending a state university through a questionnaire. PTs were asked to evaluate a variety of examples of the normal distribution. The data were analyzed using descriptive statistics and item-based analysis. It was found that the PTs were more knowledgeable about the features of the normal distribution than about standard deviation. Real-life situations made it difficult for them to interpret the concepts associated with normal distributions. Normal distribution curves were more easily interpreted by the PTs. It was determined that their mistakes were the result of their not establishing connections between statistical concepts, and their operational evaluation of concepts. Additionally, PT errors were caused by failure to establish a relationship between probability values and the area covered by the curve and making errors in operation.

Keywords: Normal distribution, pre-service middle school mathematics teachers, subject matter knowledge

Introduction

The topic of what mathematics teachers should know is an up-to-date area of research attracting a great deal of attention (Ponte & Chapman, 2016; Sullivan & Wood, 2008) because the knowledge and skills of the teacher have a direct effect on students' learning (Ball, 2003; National Council of Teachers of Mathematics [NCTM], 2000; Walshaw, 2012). Teachers should have much more than the knowledge they teach to their students (Ball et al., 2008; Hill et al., 2008). The strength of this knowledge not only triggers the development of pedagogical content knowledge, but also has the potential to support student achievement (Baumert et al., 2010; Rivkin et al., 2005; Shulman 1986). In addition, the fact that subject matter knowledge directly affects the teacher's pedagogical preferences (e.g., Even et al., 1996) indicates the need to investigate this knowledge in more depth. However, research (e.g., Hill et al., 2007; Mewborn, 2003) shows that teachers/pre-service teachers have various difficulties in gaining a deep and rich understanding of the concepts they will teach. A similar situation draws attention in statistics (e.g., Canada, 2008; Groth, 2007; Makar & Confrey, 2005; Mooney et al., 2014), where subject matter knowledge is important. Statistics is important (Konold & Higgins, 2003) because students today need to have a strong understanding of this learning area in order to evaluate and analyze statistical data (e.g., TV news, research results) correctly (Baillargeon 2005; Vermette & Savard, 2019). This makes the knowledge of teachers who directly affect students' learning important

(Franklin et al. 2015). Teachers' subject matter knowledge plays a critical role both in using effective teaching strategies about statistics in the classroom environment and in the development of students' understanding of big ideas related to statistics (Burgess, 2011; Groth, 2013). The normal distribution is an important concept that relies on this knowledge (Batanero et al., 2004) because the normal distribution plays a critical role in representing and understanding data in real-life (Watkins et al., 2008). The normal distribution appears to be a very complex concept that involves the integration and relationship of many different statistical concepts and ideas (Batanero et al., 2004). The normal distribution allows the modelling of many physical, biological, and psychological conditions such as physical measurements, test scores, and measurement errors and is a prerequisite for statistical analyses to be conducted (Batanero et al., 2004). However, research has revealed that there are various difficulties regarding the normal distribution (e.g., Bansilal, 2014). In the current study, it is aimed to reveal the subject matter knowledge (SMK) of preservice teachers (PTs) about the normal distribution.

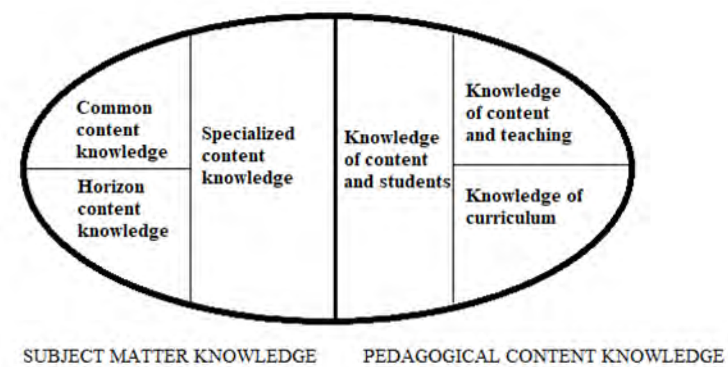
Theoretical Framework

Teacher Subject Matter Knowledge

There has been a significant increase in studies focusing on the knowledge of teachers in recent years and different classifications of this knowledge have been proposed (Tchoshanov et al., 2017). One of the common results of these studies is that the teacher's SMK is of vital importance for the teacher (Ball et. al., 2008; Grossman, 1990; Shulman, 1986, 1987). Shulman (1986) defined the teacher's SMK as "the amount and organization of knowledge per se in the mind of teachers" (p. 9). On the basis of the teacher knowledge proposed by Shulman (1986), Ball et al. (2008) drew attention to the necessity of having SMK and pedagogical content knowledge in order for a teacher to be able to teach mathematics. Ball and colleagues (2008) analyzed SMK by defining it in more depth. See Figure 1 for this information.

Figure 1

Mathematical Knowledge for Teaching (Ball et al., 2008)



Ball et al. (2008) examined SMK by dividing it into three components: common, horizon and specialized content knowledge. While common content knowledge is defined as the knowledge that the teacher should know outside the teaching environment, specialized content knowledge is considered as the knowledge required by the specialization in the field of mathematics teaching. Horizon content knowledge, on the other hand, requires the teacher to be aware of how the subject to be taught is related to other subjects. All these definitions show that both the effectiveness of the teaching process and the development of students' knowledge are directly related to the teacher's SMK

(Hill et al., 2005; Ma, 1999). In order for the teacher to carry out an effective teaching process, they must be able to make in-depth associations about the focused mathematical concepts (Bair & Rich, 2011). Statistics is one of the areas that include such concepts (Groth, 2007).

Individuals have to make evaluations and decisions based on the statistical data surrounding all areas of life. This requires individuals to have knowledge and skills related to statistics (Bargagliotti et al., 2020). Clearly, teacher competence plays a very important role in ensuring students acquire these skills and knowledge. One of these competences is SMK about statistics (Franklin et al., 2015). SMK should be present in the teacher's knowledge of many statistical concepts, one of which is the normal distribution (Batanero et al., 2004). The normal distribution forms the basis for other distributions and provides important clues to the reader at the point of making sense of the data. Many situations in real-life (e.g., biological, psychological) can be modelled using a normal distribution. Different distributions (e.g., Poisson) can be approximated to a normal distribution. Even when the datasets are not normally distributed, it can be assumed that the sample means are normally distributed based on the central limit theorem and analyses can be made based on this (Batanero et al., 2004). The normal distribution is defined as follows: If the probability density function of the continuous random variable X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

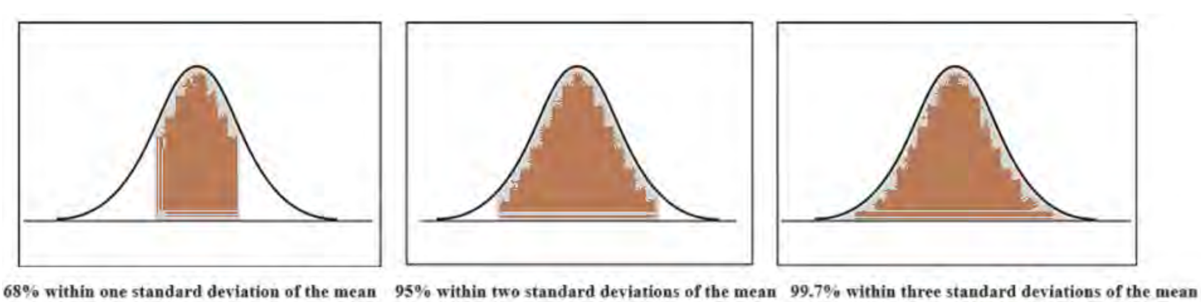
Then the random variable X has a normal distribution. The probability density function f(x) of the normal distribution has the following properties.

i. For each x, $f(x) \geq 0$.
ii. The area under the curve $f(x)$ and bounded by the x-axis is equal to 1.
iii. The curve of $f(x)$ is symmetrical according to $x = \mu$.
iv. The two ends (tails) of the curve go to infinity.

In addition, in a normally distributed dataset, 68% of the data are within one standard deviation of the mean, 95% of the data are within two standard deviations of the mean, and 99.7% of the data are within three standard deviations of the mean. This is called the empirical rule or the 68-95-99.7 rule. See Figure 2 for visuals of this rule.

Figure 2.

Empirical rule or 68-95-99.7 rule (Hinders, 2010, p.75)



The standard normal distribution is a special case of a normal distribution. In other words, a normal distribution with a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$ is called a standard normal distribution.

Although the concept of a normal distribution are not directly included in the middle school mathematics curriculum in many countries and Türkiye (e.g. Bansilal, 2014; Ministry of National Education [MoNE, 2018]), mathematics teachers should possess a broad knowledge and skill set regarding this subject matter, as opposed to merely knowing what their students will learn in class. Despite the importance of a normal distribution, few studies have examined individuals' knowledge of it (Batanero et al., 2004; Pfannkuch & Reading, 2006). These studies revealed deficiencies in the knowledge of teachers/PTs about the normal distribution (Bansilal, 2014; Batanero, et al., 2004; Chaput, et al., 2021; Huck, et al., 1986).

For example, Huck et al. (1986) discovered two erroneous conceptions of normal standard scores in their study of university students. The first is that some students thought that standard scores always vary between –3 and +3. In addition, some students thought that there is no restriction on the maximum and minimum values in these scores. It was concluded that the reasons for these misunderstandings of the students were related to a normal distribution. The students who thought that z-scores always vary between -3 and +3 often used a picture or a table of the standard normal curve with this range of variation. Similarly, the students who thought that there were no maximum and minimum limits for z-scores made an incorrect generalization, thinking that the tails of the normal curve were asymptotic for the abscissa because they did not notice that no finite distribution is exactly normal.

In their study on university students, Batanero et al. (2004) revealed that students had various difficulties regarding normal distribution and many related concepts. Students had difficulties in calculating probabilities under the curve and the graphical representations of the areas under the normal curve. In addition, students misinterpreted the skewness coefficient and assumed that the equities of the mean, median, and mode were sufficient to show the symmetry of the distribution. Some students, on the other hand, were unaware of the effect of range widths on the frequency represented by the area under the histogram. In addition, some students were insufficient in relating other statistical measures (e.g., mode, median) to each other while interpreting the scatter plots.

Bansilal (2014) focused on teachers' knowledge of the concept of a normal distribution in her study conducted with the participation of 290 teachers. The teachers were given two tasks that required the application of the properties of the standard normal distribution curve, and their knowledge was examined based on these tasks. The teachers' success rates for the tasks were 27% and 14%, respectively. When the difficulties they experienced were examined, it was found that they could not establish a connection between the probability values and the area covered by the curve.

Chaput et al. (2021) examined how undergraduate students studying in various departments (statistics, physics, finance) name the vertical axis of a normal distribution and found that only 27 (18.2%) of 148 students could name the vertical axis of a normal distribution correctly and that only five students (3.4%) were able to explain their naming. Students with various misconceptions named the vertical axis as "probability," "count," or "frequency". Few of the students used the correct name as "probability density". These results show the difficulties of both university students and teachers / PTs regarding the concept of normal distribution.

Rationale of the Study

The concept of a normal distribution in statistics education is not part of the middle school mathematics curriculum in many countries (e.g., MoNE, 2018). However, researchers argue that PTs should have knowledge and skills related to this concept (Bansisal, 2014) because of the argument that middle school mathematics teachers should have in-depth knowledge of mathematical concepts (Ball et al., 2008). The normal distribution has an important role in reasoning about real-life situations as well as in understanding other statistical concepts and the relationships between these concepts (Batanero et al., 2004).

Moreover, middle school education plays an important role in developing students' understanding of statistical concepts, because students learn statistical concepts during the middle school years that serve as a foundation for advanced statistical concepts in the later grades. Developing a positive attitude towards statistics and critically evaluating statistical situations begins when students understand the why and how of statistical concepts in the early grades. For them to understand higher-order statistical concepts, they must have a thorough understanding of the underlying statistical concepts. The effectiveness of teaching processes depends on the teachers' ability to establish connections between statistical concepts and allow students to experience higher-order statistical concepts informally, even if not formally (Makar, 2018). Having a limited understanding of the normal distribution by mathematics teachers can negatively impact the teaching process. In this situation, students may only perceive other statistical concepts that shape normal distributions as operations without considering their meaning. As a result, teachers' ability to teach the normal distribution and related concepts are directly related to their subject knowledge. It is therefore important that PTs have the necessary knowledge and skills in the subjects they will teach (Baştürk 2009, 2011). The knowledge and skills PTs possess regarding the normal distribution may provide important clues about what they know about many other statistics concepts as well. Moreover, as teacher education is important for all societies, (Aslan, 2003), teacher training programs should be carefully organized. Accordingly, the results from this study may contribute to the regulation of the content of elementary education mathematics teaching undergraduate courses (e.g., statistics, probability, teaching of statistics and probability). Further, little research has been conducted on normal distributions in the literature (Batanero et al., 2004; Pfannkuch & Reading, 2006). Teachers also have difficulties with these concepts (Bansilal, 2014; Batanero et al., 2004; Chaput et al., 2021; Huck et al., 1986) pointing out the need to reveal PTs' knowledge of the normal distribution. Thus, answers to the following research questions were sought in the current study:

1) What is the knowledge of PTs about normal distribution?
2) What are the reasons for PTs to misinterpret the situations related to normal distribution?

Method

Design of the Study and Participants

Qualitative survey research design allows the researcher to reveal what individuals know about the concept in focus (Fraenkel & Wallen, 2006; Jansen, 2010). This design has been adopted in the current study, as it aimed to reveal the knowledge of PTs regarding the normal distribution. The participants were determined by using the convenience sampling method as it was more suitable for the research design (Fraenkel & Wallen, 2006). The participants of the study were third graders enrolled in the Elementary Education Mathematics Teaching program at a state university located in the Central Anatolia region of Türkiye. Participants' ages vary between 22 and 24, and 86 (71.7%) were female and 34 (28.3%) were male. PTs study normal distribution and related concepts in the probability and statistics courses in their undergraduate curriculum. The PTs took the probability course in the second term of their second year and the statistics course in the first term of their third year. The letter grades of the PTs in these courses are presented in Table 1. As seen in Table 1, nearly half of the PTs (45%) had a grade of 70 or less in the probability course. In the statistics course, half of the PTs (50%) were more successful and had grades in the range of 85-100.

Table 1

Grades of the PTs from the Probability and Statistics Courses

Course	Probability		Statistics	
Letter Grade	**n**	**%**	**n**	**%**
E	32	26	18	15
D	22	19	22	19
C	22	19	20	16
B	26	21	30	25
A	18	15	30	25

Note: 60-69-E, 70-74-D, 75-84-C, 85-89-B, 90-100-A

Data Collection Instrument and Process

In this study, the focus was on the common content and specialized content knowledge of the PTs. In order to reveal the PTs' content knowledge about the normal distribution, questions under four main skills were created. These skills are shown in Table 2.

Table 2

Skills and Questions on Normal Distribution

Skills	Types of SMK (Question number)	
	Common content	**Specialized content**
Knowing properties of normal distribution and related concepts	1,2	-
Interpreting concepts related to normal distribution on the basis of datasets	4,5	3
Interpreting the concept of normal distribution on the basis of real-life situations	7,9	-
Interpreting normal distribution curves	8	10a, 10b

Two questions were asked to the PTs under the skill of "Knowing properties of normal distribution and related concepts". See Figure 3 for this information.

Figure 3

Questions Regarding the Properties of the Normal Distribution and Related Concepts

1)Which of the following are properties of the standard deviation? Explain your answer.
1. The interval ± 2s contains 50% of the data in the distribution.
2. It's resistant to extreme values.
3. If you added 20 to every value in the dataset, the standard deviation wouldn't change.
4. If we subtract 5 from each value in the dataset, the standard deviation will decrease.
5. It's independent of the number of terms in the distribution.
6. It's the square root of the average squared deviation from the mean.

2)Which of the following are properties of the normal distribution? Explain your answer.
1. It has a mean of 0 and a standard deviation of 1.
2. Its mean = median = mode.
3. All terms in the distribution lie within four standard deviations of the mean.
4. It is bell-shaped.
5. The total area under the curve and above the horizontal axis is 1.

Hinders (2010) created the questions and an option was added to the first question (item 4). In the first question, it was aimed to reveal the knowledge of the PTs about the properties of standard deviation, and in the second question about the properties of the normal distribution. Under the second skill, three questions were asked to the PTs to reveal their knowledge and skills on "interpreting concepts related to the normal distribution on the basis of data sets" (Hinders, 2010,). The third question was arranged in a way to reveal the specialized content knowledge of the PTs about the normal distribution. In the other two questions, which to reveal the common content knowledge of the PTs, the phrase "Explain your answer" was added to better reveal the thoughts of the PTs.

In the third question, the PTs were asked to interpret the mean, the standard deviation, and the shape of the distribution. The fourth and fifth questions focused on calculating the concepts related to movement (z score, mean, standard deviation) from a normally distributed dataset (Hinders, 2010). Under the third skill, two questions (Hinders, 2010; Watkins et al., 2008) were asked to the PTs to reveal their knowledge and skills on "interpreting the concept of normal distribution on the basis of real-life situations" (See Figure 4).

Figure 4

Questions Regarding the Interpretation of Concepts Related to the Normal Distribution on the Basis of Datasets

3) Suppose that you are dealing with the subject of normal distribution in your class. You ask following question; "The mean and standard deviation of a normally distributed dataset are 19 anc respectively. 19 is subtracted from each term in the dataset and the result is divided by 4. Which of following best describes the resulting distribution?" You see that the following answers are given your students.
S1. It has a mean of 0 and a standard deviation of 1.
S2. It has a mean of 0 and a standard deviation of 4 and its shape is normal.
S3. It has a mean of 1 and a standard deviation of 0.
S4. It has a mean of 0 and a standard deviation of 1 and its shape is normal.
S5. It has a mean of 0 and a standard deviation of 4 and its shape is not known.
Which of these answer(s) can be an answer to the given question? Explain your answers.

4) One of the values in a normal distribution is 43 and its z-score is 1.65. If the mean of distribution is 40, what is the standard deviation of the distribution? Explain your answer.
a. 3
b. –1.82
c. 0.55
d. 1.82
e. –0.55

5) A college readiness exam scores are known to be approximately normally distributed with mean and standard deviation 6. To the nearest integer value, how many scores are there between 63 and
Explain your answer.
a. 0.6247
b. 4,115
c. 3,650
d. 3,123
e. 3,227

Figure 5

Questions Regarding the Interpretation of the Concept of Normal Distribution on the Basis of Real-life Situations

7) Merve wonders how the size of her beagle Boncuk compares with other beagles. Boncuk is 40.6 cm tall. Merve learned on the internet that beagles heights are approximately normally distributed with a mean of 38.5 cm and a standard deviation of 1.25 cm. Interpret the height of Boncuk.

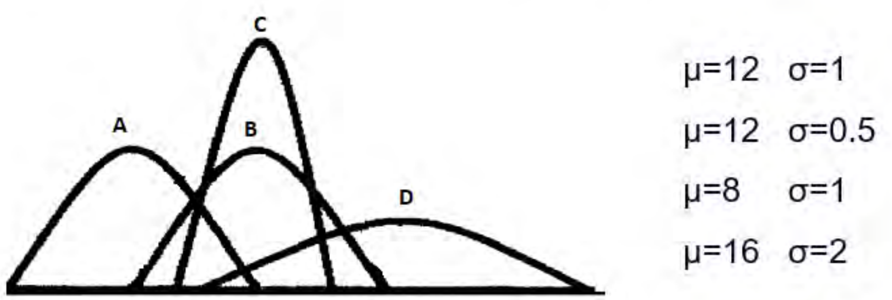
9) Alperen got 70 points from a statistics test whose mean is 60 and standard deviation is 10. He got 50 from a physics test whose mean is 50 and standard deviation is 5. It is known that the scores of the both test are normally distributed. Interpret and compare Alperen's test scores.

In both questions (see Figure 5), it was aimed to reveal the knowledge and skills of the PTs on interpreting real-life situations by associating them with the normal distribution. Under the fourth skill, two questions were asked to reveal the knowledge and skills of the PTs on "interpreting normal distribution curves"(See Figure 6).

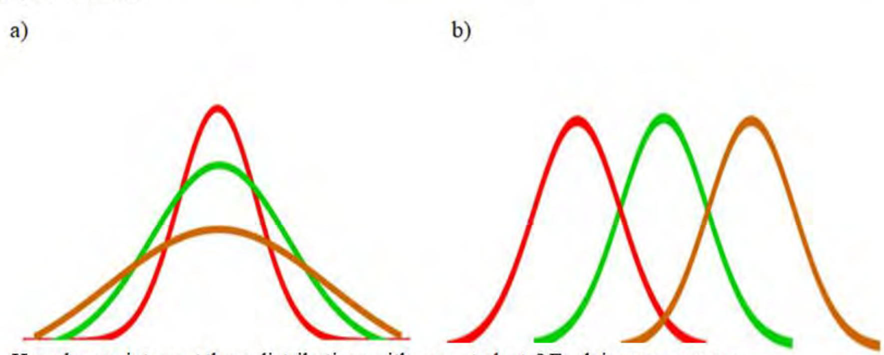
Figure 6

Questions Regarding the Interpretation of the Normal Distribution Curves

8) The populations to which the parameters given below belong show a normal distribution. Match these parameters with the normal distribution curves that you think are appropriate (A, B, C, D). Explain your answer.



10) Imagine a lesson in which you talk about distributions. How do you interpret these distributions with your students?



How do you interpret these distributions with your students? Explain your answers.

The eighth question was created based on the study by Özmen (2015). The PTs were asked to explain their answers to examine their thoughts in more detail. The tenth question, which was created by the researcher, and was prepared to reveal the specialized content knowledge of the PTs.

Translations of the original English questions into Turkish were verified by language experts. Following the correction of words or phrases that might cause semantic shifts or misunderstandings, the final forms of the questions were created. Three experts working in statistics education were asked to review the questions for relevance to the objectives and evaluate content validity. A pilot study was conducted with the PTs at the end of the third year, followed by the actual study. Thirty third-year students from a small, state university in Central Anatolia were recruited to participate in the pilot study. Following the pilot study, the actual study was carried out and the PTs were surveyed. It took the PTs 30-45 minutes to complete the questions.

Data Analysis

Descriptive statistics and item-based analysis were performed. Each question was assessed using an evaluation rubric. Each of the PT's answers to each question was categorized as correct and incorrect, then the reasons for each answer were presented as frequencies and percentages. An expert examined 25% of the data for validity and reliability, and the level of agreement between the expert and the study was determined. A 90% inter-coder reliability was found between the researcher (author) and the expert. As a result of discussions, a consensus coding was reached after a number of differences were identified. A description of how the research was conducted was also provided. The results obtained were further supported by several direct quotations. Accordingly, the study's validity and reliability were established (Miles & Huberman, 1994).

Research Ethics

This study adhered to all ethical procedures. Ethical compliance approval was obtained for this research in accordance with the decision of the Scientific Research and Publication Ethics Committee in the field of Social and Human Sciences of Karamanoğlu Mehmetbey University Ethics Committee dated on 17.05.2022 and numbered with 04-2022/80

Findings

The PTs' SMKs concerning the normal distribution was classified under four headings. As a first step, information about PTs' understanding of the normal distribution properties is given, as well as samples of their responses. Secondly, PTs' interpretations of concepts related to the normal distribution are presented along with samples from the PT responses. The third heading presents the PTs' interpretation of the normal distributions based on real-life situations. The last heading presents the interpretations of PTs about normal distribution curves.

PTs' Knowledge on "Properties of Normal Distribution and Related Concepts"

The PTs were asked to answer two questions in the context of the properties of normal distribution and related concepts. The evaluation rubric given in Table 3 was prepared. Answers given to the first and second questions by the PTs were analyzed and the results are presented in Table 4. A review of the PTs' responses to the first question revealed that 11.9% were able to accurately identify all the properties of standard deviation. One property of standard deviation was misidentified by 20%

of participants, two by 36.9%, and three by 23.3%. A mere 6% of PTs could identify only one property correctly, and 1.9% could not identify any of them.

Table 3

Evaluation rubric for the first and second questions

Correctness of the answers	Codes
All the items are correct	5
1 item is incorrect	4
2 items are incorrect	3
3 items are incorrect	2
4 items are incorrect	1
All the items are incorrect	0

Table 4

Answers Given to the First and Second Questions by the PTs

Codes	First Question Frequency/Percentage	Second Question Frequency/Percentage
5	14 (11.9%)	24 (20.0%)
4	24 (20.0%)	60 (50.0%)
3	44 (36.9%)	36 (30.0%)
2	28 (23.3%)	-
1	8 (6.0%)	-
0	2 (1.9%)	-

In the study, 81.6% of PTs misunderstood the statement "The standard deviation is independent of the number of terms in the distribution" when it was examined which statements related to the standard deviation the PTs tended to misidentify. Upon being asked to explain why this statement is incorrect, all of the PTs who made mistakes considered the standard deviation from an operational perspective. PT2 justified their answer by saying "the standard deviation cannot be considered independently of the number of data because we also use the number of data in the formula" and didn't explain what the standard deviation means.

The following statement was misinterpreted by 68.3% of the PTs; “The standard deviation is equal to the square root of the average squared deviation from the mean.” The PTs were observed to evaluate the standard deviation operationally when examining how they justified their answers. Taking PT8, for example, it was only the formula of the standard deviation that was discussed: “There is division by the number of data in the formula. So this statement is wrong.”

There were 20% of PTs who misunderstood the statement, "if we subtract 5 from each value in the dataset, the standard deviation will decrease". PT41 stated that “if we subtract 5 from each value, the arithmetic mean decreases by 5. Therefore, the standard deviation decreases." Thus, it was apparent that they believed the decrease in the arithmetic mean had a direct impact on the standard deviation. The standard deviation, however, offers information about how the data are distributed, as opposed to the arithmetic mean, which is a measure of central tendency.

About 15% of respondents misinterpreted another statement, "The standard deviation won't change if you add 20 to every value in the dataset". As an example, PT2 replied, “Adding 20 to each value will increase the standard deviation because it will increase the arithmetic mean.” The PTs followed the same logic and made a direct connection between the arithmetic mean and the standard deviation, disregarding the fact that the arithmetic mean indicates where the dataset is centered, whereas standard deviations indicate where the dataset is distributed.

According to the results, 13.3% of the PTs misinterpreted the following statement; "The standard deviation is resistant to extreme values." As with other wrong reasons, the association between standard deviation and the arithmetic mean in operational terms emerged when the answers of the PTs were evaluated. In evidence of this, PT27 explained: "The standard deviation depends on the extreme values since the extreme values change the arithmetic mean and the standard deviation depends on the arithmetic mean". There was only one PT who misinterpreted the statement "The interval ± 2s contains 50% of the data in the distribution". This misinterpretation can be attributed to the inability to establish a relationship between distribution and standard deviation.

According to the results of the second question in which the PTs were quizzed about their knowledge of the normal distribution, 20% of them were able to evaluate all normal distribution statements correctly. Among the PTs, 50% misinterpreted one statement and 30% misinterpreted two statements. Upon evaluating which statement(s) the PTs interpreted incorrectly, it was observed that the majority (68.3%) misinterpreted the following statement: "It has a mean of zero and a standard deviation of one." A possible explanation for why this statement was evaluated incorrectly is that the PTs were confused between the normal distribution and standard normal distribution concepts.

The following statement was misinterpreted by 30% of the PTs; "All terms in the distribution lie within four standard deviations of the mean." PT30, for example, responded, "Yes, all data are within four standard deviations of the mean." Based on the theory, it is known that almost all data are within four standard deviations of the mean, however some data are further from it. Another misinterpreted statement is "The total area under the curve and above the horizontal axis is 1". There were 18.3% of PTs who misinterpreted this statement. PT26, for example, stated, "The area under the normal curve is 1, but the area above the curve is greater than 1." Due to a misunderstanding of the normal distribution curve, the PT gave the wrong answer. Normal distributions have this property, which makes them useful for calculating probability density curves. It is clear from evaluating the PTs' answers to the two questions that they have a better understanding of normal distribution properties than standard deviation properties.

PTs' Knowledge on "Interpreting Concepts Related to the Normal Distribution on the Basis of Datasets"

PTs were asked to answer three questions related to interpreting the normal distribution concepts using the datasets. There were two questions (4, 5) that aimed to expose common content knowledge, and one (3) that aimed to expose specialized content knowledge. The evaluation rubric in Table 5 was developed for this purpose.

Table 5

Evaluation Rubric for the Third, Fourth and Fifth Questions

Correctness of the answers	Codes
Correct answer & correct justification	3
Correct answer & wrong/incomplete justification	2
Wrong answer	1
No answer	0

On the basis of the evaluation rubric prepared, the answers given by the PTs to the third questions are shown in Table 6. Nearly half of the PTs (48.1%) answered the third question correctly. However, when asked to justify their answers, approximately one-third (27%) of the PTs who gave correct answers could not justify their answers.

Table 6

PTs' Answers to the Third Question

Codes	Answers	Frequency/Percentage
3	Option D	42 (35%)
2	Option D	16 (13.3%)
1	Option B	38 (31.7%)
	Option E	16 (13.3%)
	Option A	8 (6.7%)

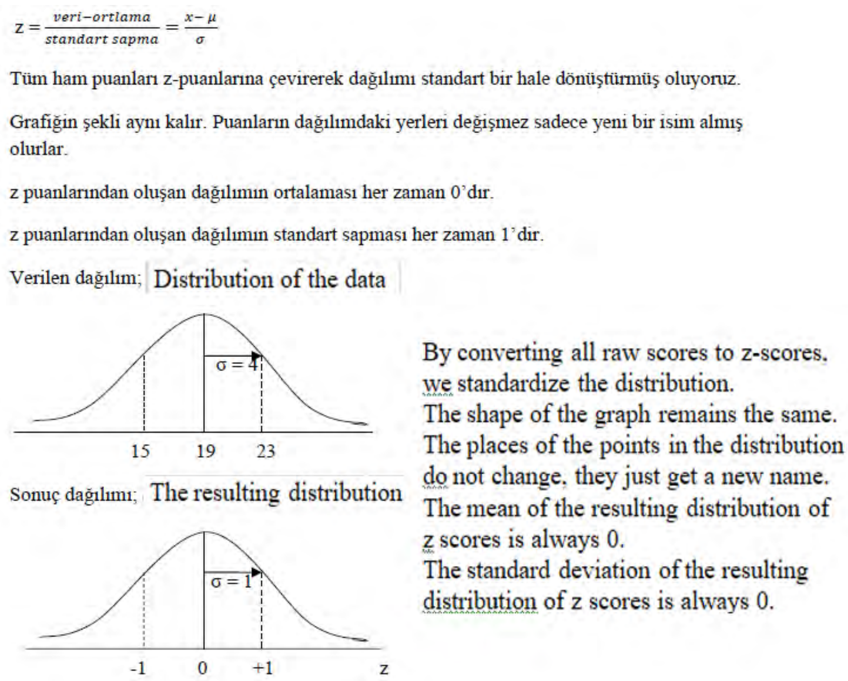
According to PT4 who gave the correct answer and justified it correctly:

> The student who chose option D chose the right answer because subtracting the same value from the mean of a dataset as the mean is to reduce the old mean by the same amount. The first mean was 19, and when 19 is subtracted from each term, the new mean becomes zero. The effect of dividing each term by the same value on the standard deviation of a dataset is to divide the standard deviation by that value. The first standard deviation is four, a new standard deviation of one is obtained when each term is divided by four. If we look at the z-score, its mean is zero and its standard deviation is one. Since the previous distribution is a normal distribution, this distribution will be normal, too.

Another PT who gave the correct answer and correctly justified his/her answer was PT43. See Figure 7 for their response.

Figure 7

PT43's Answer



A good example of a PT who gave the wrong justification despite giving the correct answer is PT49. According to this PT, the student who chose option D gave the correct answer because even when a small sample of data is used, it is clear that this option is correct. PTs who gave wrong answers mostly tended to choose statements like: "It has a mean of zero, a standard deviation of four, and a normal shape" (31.7%). PT24 explained, for example, that:

> The answer of the student who chose option B is correct because, starting from the arithmetic mean of the dataset, the sum of the data in the dataset is 76, and when 19 is subtracted from each data and the arithmetic mean is calculated, the new result will be zero. Since it will be calculated by subtracting 19 from each data, the arithmetic mean will also decrease by 19. Since the number of data remains the same, the standard deviation will not change. That is, the standard deviation will remain as four. As the standard deviation does not change, the distribution is still a normal distribution.

Thus while they was able to determine how the mean was affected by the changes to be made in the dataset, they ignored the possible changes in the standard deviation. On the other hand, 13.3% of the PTs chose the following statement: "It has a mean of zero, a standard deviation of four and its shape is unknown." The PT50 giving this answer justified his/her answer as follows in Figure 8:

Figure 8

PT50's Answer

x= veri toplamı
y= veri sayısı
x/y=19 (ortalama)
x-19y/y= x/y - 19y/y
= 19 -19
=0 (ortalama)
Standart sapmanın değişmeyeceğini düşünüyoruz. Dağılımın şekli hakkında ise yorum yapamayız.

x=sum of data
y=number of data
x/y=19(mean)
x-19y/y=x/y – 19y/y
=19-19
=0 (mean)
We think that the standard deviation will not change.
Do not make any comment on the shape of distribution

Here, in relation to the previous answer (option B), although they took into account how changes in the data set affect the arithmetic mean, they ignored the standard deviation and how the shape of the distribution would change. Although they were able to determine the standard deviation and mean correctly, the students who could not determine the shape of the distribution and gave wrong answers constituting 6.7% of all the students. For example, PT33 made the following explanation:

> If we assume that all four samples are 19 and subtract 19 from each, we see that the mean is zero. We see that the standard deviation is one. So the standard deviation also becomes one. When tested with other samples, it is easily concluded that the arithmetic mean is zero.

The answers of the PTs given to the fourth question, which is another question in which the PTs are required to interpret the concepts related to the normal distribution on the basis of the datasets, are shown in Table 7.

Table 7

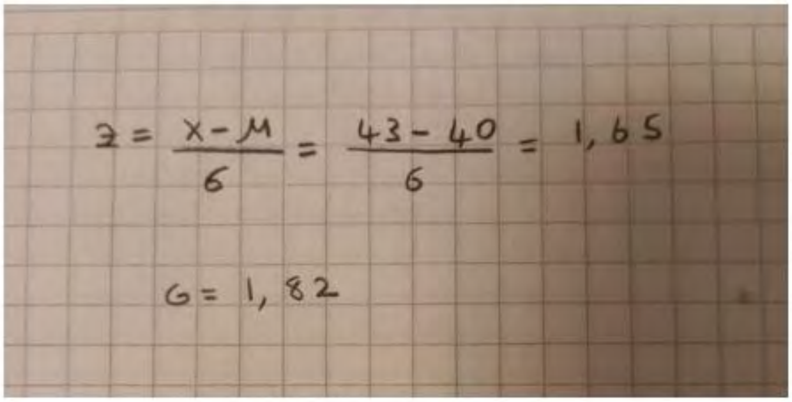
Answers Given to the Fourth Question by the PTs

Codes	Answers	Frequency/Percentage
3	Option D	96 (80%)
2	Option D	-
1	Option C	24 (20%)

It has been revealed that all of the PTs who answered this question correctly were able to justify their answers correctly. For example, PT9 was able to correctly calculate the standard deviation of the dataset as shown in Figure 9.

Figure 9

PT9's Answer



It was observed that the PTs who gave wrong answers gave these wrong answers because they made a mistake in the operation. For example, PT19 gave the following answer as shown in Figure 10.

Figure 10

PT19's Answer

Option C.

1.65=43-40/standard deviation

Standard deviation=0.55

When the answer given is examined, it can be said that the PT made an operational mistake. The answers given by the PTs to the fifth question are shown in Table 8.

Table 8

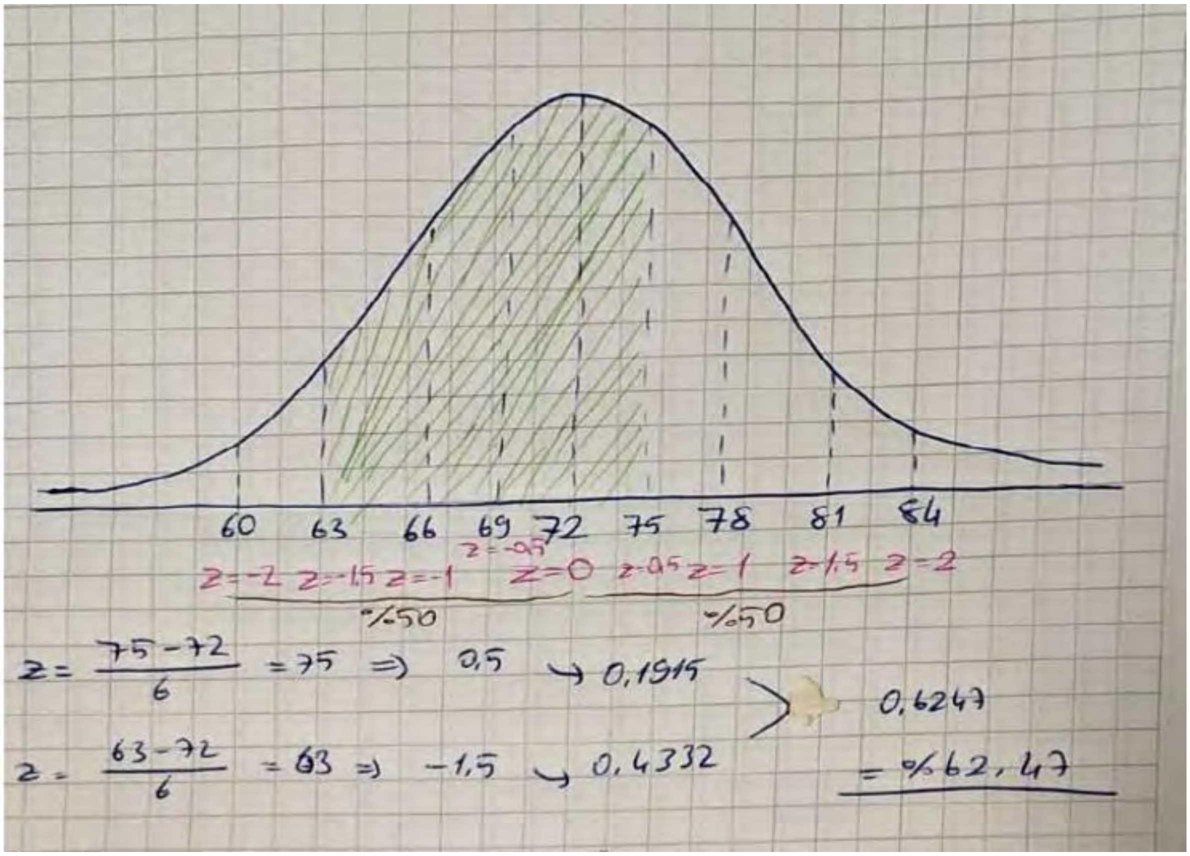
Answers Given to the Fifth Question by the PTs

Codes	Answers	Frequency/Percentage
3	62.47	66 (55%)
2	62.47	-
1	57.95	16 (13.3%)
	57.5	8 (6.65%)
	24.17	16 (13.3%)
	65.86	6 (5%)
	17.5	6 (5%)
	80	2 (1.75%)

For this question, it was revealed that more than half of the students could both give the correct answer and justify their answers. The answer of PT23, one of the PTs who gave both the correct answer and correct justifications, is given in Figure 11.

Figure 11.

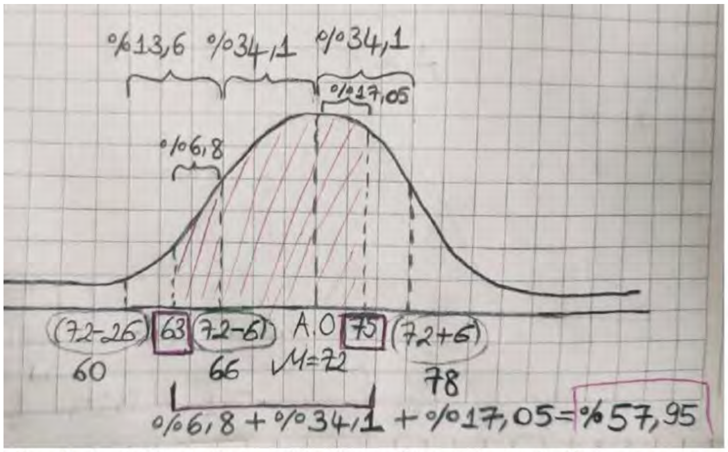
PT23's answer



It was revealed that the wrong answers given by the PTs who gave wrong answers varied. Of the PTs, 13.3% found the answer as 57.95. The answer given by PT1 is as follows in Figure 12

Figure 12

PT1's Answer



Görselde de görüldüğü gibi normal dağılım grafiğini çiziyoruz. Grafiğin tam ortasına verilen aritmetik ortalamayı yazıyoruz yani 72 değerini yazıyoruz. Daha sonra ise standart sapması 6 değeri verildiği için grafiğin sağ tarafına (72+1 standart sapmayı) yazıyoruz. Burada standart sapma 6 verildiği için 72+6 'dan 78 değerine ulaşıyoruz. Bu kısımda 75 değeri aritmetik ortalama ile +1 standart sapmanın tam ortasında kaldığından yani 75 ile 78 değeri arasında bulunduğundan buradaki yüzde değer olan %34,1' i ikiye bölüyoruz ve %17.05 değerine ulaşıyoruz. Şimdi ise grafiğin sol tarafına geçelim. Burada (72 - 1 standart sapmayı) yazıyoruz. Burada standart sapma 6 verildiği için 72 - 6 'dan 66 değerine ulaşıyoruz. Ve devam ediyoruz bu kez (72 - 2 standart sapmayı) yazıyoruz. Burada standart sapma 6 verildiği için 72 - 12 'den 60 değerine ulaşıyoruz.

Böylece 63 değeri -1 standart sapma ile -2 standart sapmanın tam ortasında bulunduğundan buradaki yüzde değer olarak %13,6 'yı ikiye bölüyoruz ve %6,8 değerine ulaşıyoruz. Sonuç olarak 63 ile 75 arasındaki yüzdeleri topluyoruz bu hesaba aritmetik ortalama ile -1 standart sapma arasındaki yüzdelik değeri katmayı unutmuyoruz böylece %6,8 + %34,1 + %13,6 = %57,95 olarak bulmuş oluyoruz. Sonuç olarak verilerin %57,95 'i 63 ile 75 arasındadır diyoruz.

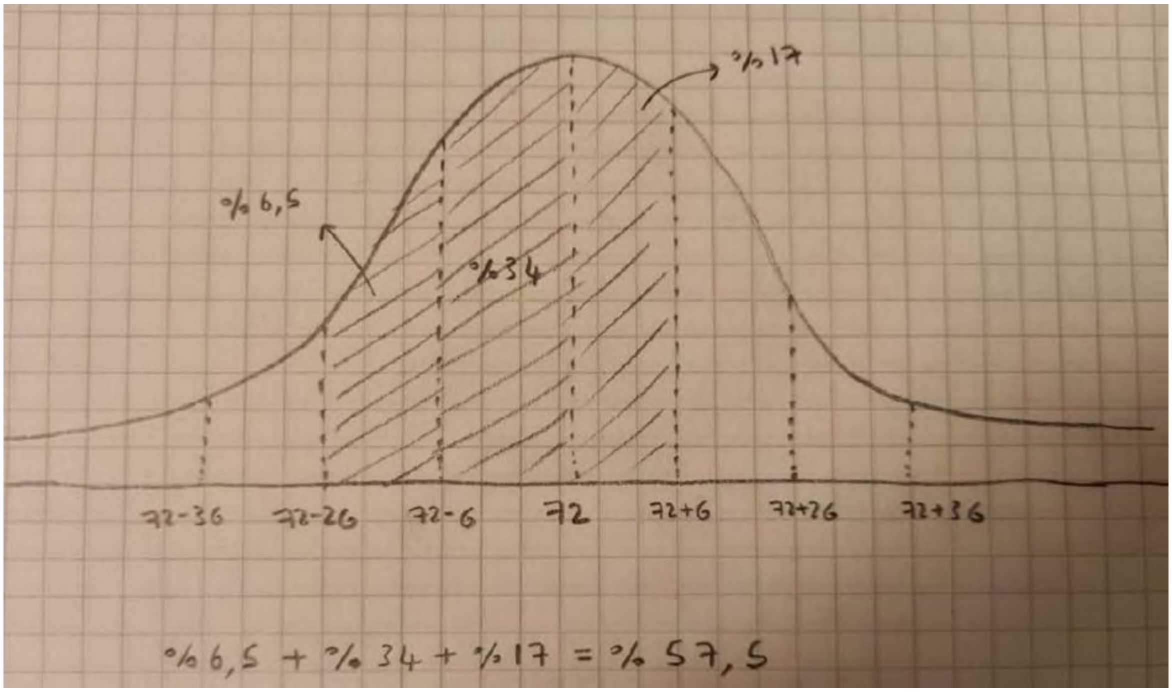
As seen in the figure, we draw the normal distribution graph. We write the arithmetic mean given in the middle of the graph, that is, we write 72. Then, since the value of the standard deviation is given as 6, we reach the value of 78 from 72+6. In this part, since the value of 75 is in the middle of the arithmetic mean and +1 standard deviation (78), that is, it is between the values of 75 and 78, we divide the percentage value here, 34.1%, into two, and we reach the value of 17.05%. Now let's move on to the left side of the graph. Here we write (72 – 1 standard deviation). Since the standard deviation is given as 6 here, we arrive at the value 66 from 72 - 6.

Thus, since the value of 63 is in the middle of -1 standard deviation and -2 standard deviation, we divide 13.6% into two as the percentage value here and we get 6.8%. As a result, we add the percentages between 63 and 75, do not forget to add the percentage value between the arithmetic mean and -1 standard deviation, so we get 6.8% + 34.1% + 13.6% = 57.95%. As a result, we say that 57.95% of the data are between 63 and 75%.

The PTs who gave this answer established a direct relationship with the distances of the data from the arithmetic mean and their areas. They correctly defined the area between 72 and 78, which is one standard deviation (standard deviation six) away from 72, as 0.34. Since the z score of the 75 points is 0.5, they divided the area into two equal parts and found it to be 0.17. A similar situation was observed for 63 points, which were found to be at a distance of -1.5z score. Here, they directly correlated the z score with the area. However, such a relationship is not always the case. For example, the area in the 0.5z score is 0.1925. The PTs who acted with a similar logic and found the answer as 57.5% constituted 6.65% of all the PTs. These PTs acted with similar logic and reached this conclusion because they did not consider the values after the comma.

Figure 13

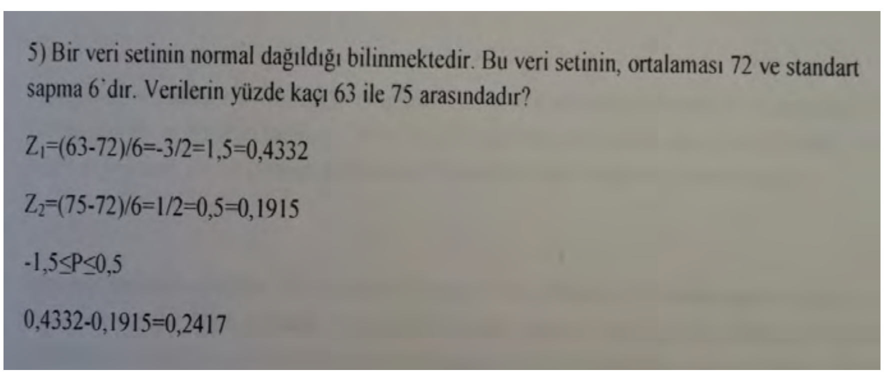
PT34's Answer



The PTs who gave the answer 24.17, which is another wrong answer, constituted 13.3% of all the PTs. The answer given by the PT42 is as follows:

Figure 14

PT42's Answer



5) Bir veri setinin normal dağıldığı bilinmektedir. Bu veri setinin, ortalaması 72 ve standart sapma 6`dır. Verilerin yüzde kaçı 63 ile 75 arasındadır?

Z1=(63-72)/6=-3/2=1,5=0,4332

Z2=(75-72)/6=1/2=0,5=0,1915

-1,5≤P≤0,5

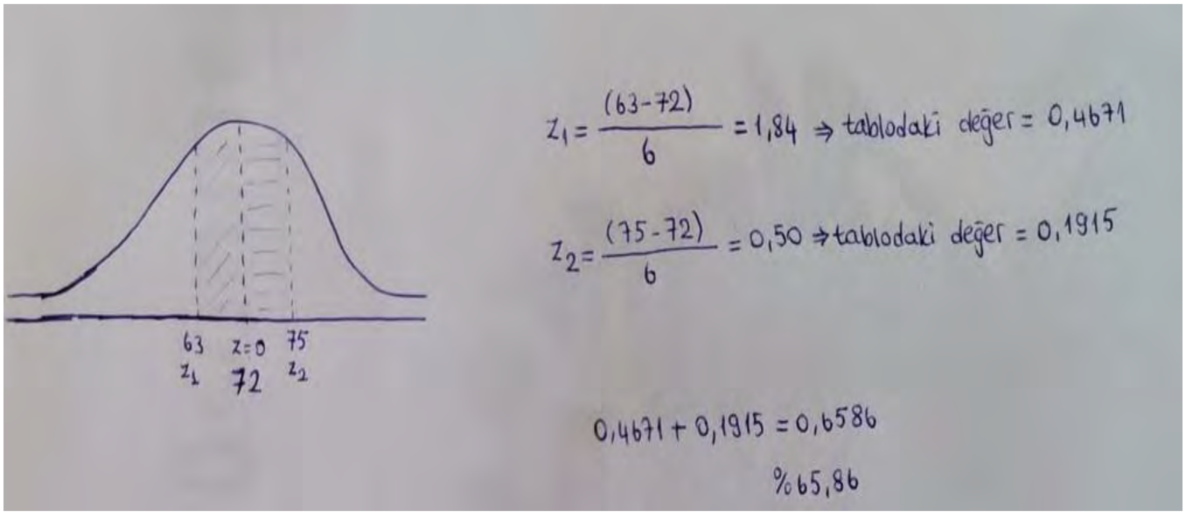
0,4332-0,1915=0,2417

In evaluating the PT''s answer, it appears that although she was able to find the z-scores accurately, she was unable to determine how the areas in the distribution were distributed. In spite of the fact that the data were both to the right and left of the arithmetic mean, they acted as if they were only to the right.

Of the total number of PTs, 5% found the answer 65.86. The answer given by PT29 is shown in Figure 15.

Figure 15

PT29's Answer



As a result of an operational mistake, the PT calculated the z-scores incorrectly and found the areas incorrectly. A total of 5% of the PTs found the answer as 17.5% without explaining why. A student also found the answer to be 80 but did not explain why.

In this section, the highest number of PTs answered the fourth question correctly and justified their answers correctly. It was revealed that although this rate decreased, more than half of students gave correct answers and justified their answers in the fifth question. About one-third of the students with the correct answer gave incomplete or incorrect justifications for their answers in the third question. This question was answered incorrectly by more than half of the PTs. Additionally, the third and fifth questions showed a greater degree of variation in the errors made by the PTs.

PTs' Knowledge on "Interpreting the Concept of the Normal Distribution on the Basis of Real-life Situations"

The PTs were asked to answer two questions in order to reveal their knowledge and skills about interpreting the concept of the normal distribution on the basis of real-life situations. The evaluation rubric for these questions is given in Table 9.

Table 9

Evaluation Rubric for the Seventh and Ninth Questions

Correctness of the answers	Codes
Correct answer & correct justification	4
Correct answer & incomplete justification	3
Correct answer & superficial justification	2
Wrong answer	1
No answer	0

Table 10 presents the PTs' responses to the seventh question, as evaluated by the prepared evaluation rubric.

Table 10

Answers to the Seventh Question

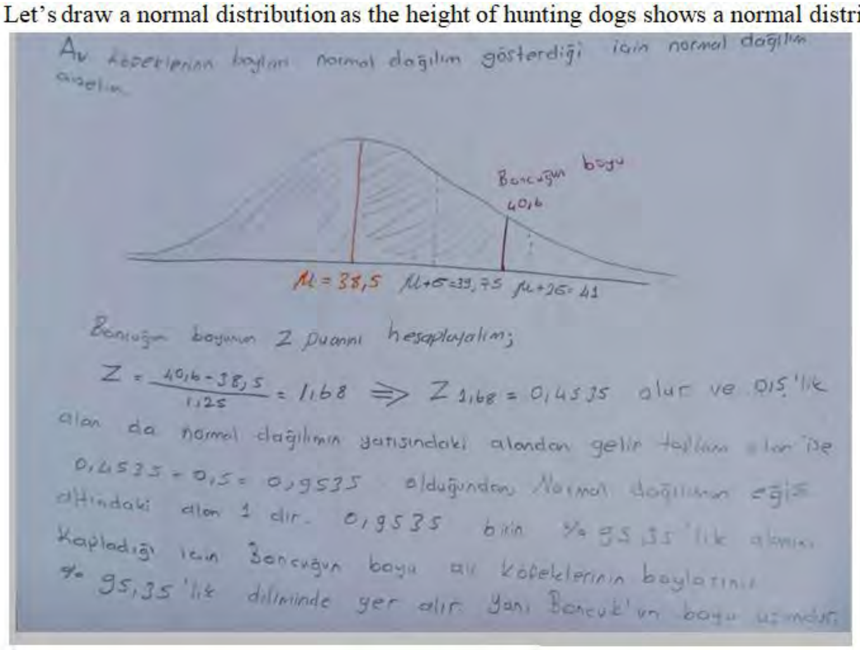
Codes	Answers	Frequency/Percentage
4	%95.35	22 (18.3%)
3	Between the standard deviations of +1 and +2	28 (23.3%)
2	Over the mean	38 (31.7%)
1	In the 45% percentile	14 (11.7%)
	In the 2.28% percentile	8 (6.7%)
	In the +2 standard deviation	8 (6.7%)
0	-	2 (1.6%)

According to the results, almost 75% of the PTs were able to answer the question correctly (73.3%). When asked to justify their correct answers, the PTs were unable to demonstrate the same level of success. It was found that only 25 percent of the PTs who gave correct answers were able to justify their answers in the correct way. Figure 16 shows the answer of PT26, who not only gave the correct answer, but also justified their answer.

Figure 16

PT26's Answer

Let's draw a normal distribution as the height of hunting dogs shows a normal distribution.



Height of Boncuk is 40.6

Let's calculate the z score of the height of Boncuk:

0.5 area comes from the area in the middle of the normal distribution, as the total area is 0.4535 + 0.5 = 0.9535, the area under the curve of normal distribution is 1. As 0.9535 covers 95.35% of 1, the height of Boncuk is in the 95.35% slice of the heights of hunting dogs. That is, Boncuk is tall.

Despite providing accurate solutions, nearly a third of the PTs (32%) failed to provide complete justifications for their answers. PT21 provided the following explanation as shown in Figure 17.

Figure 17

PT21's Answer

Boncuk is quite tall compared to other hunting dogs.

If we were to draw a normal distribution graph showing the height of the dogs;

Mean of the graph 38.5/+1 standard deviation is 39.75 / +2 standard deviation is 41

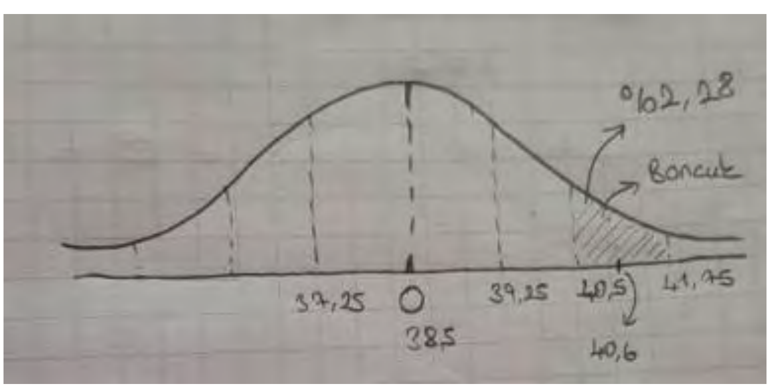
- In this case, we see that the height of Boncuk (40.6) is between +1 standard deviation and +2 standard deviation and is closer to +2 standard deviation. In other words, we can say that Boncuk is taller than 85% of all dogs.

By using correct reasoning, PT21 was able to conclude that the height of Boncuk lies in the range of +1 and +2 standard deviations, and is most likely closer to +2 standard deviations due to the right reasoning they demonstrated. Due to the lack of explanation of the point Boncuk corresponds to on the z score distribution, they justified their answer incompletely. A total of 43% of the PTs who gave the correct answer gave superficial justifications for their answers despite giving the correct answer. Although these PTs attempted to justify Boncuk's height by associating it with the standard deviation, their explanations remained quite superficial. PT56 used a statement such as "Boncuk's height is above the mean according to the standard deviation", but did not explain how this was determined.

The PTs who made mistakes constituted 25.1% of all the PTs. When the types of the mistakes made by these PTs were examined, three different types of mistakes were noticed. PS30, one of the PTs who found the answer as 45%, gave the following answer: "The z-score of the height of Boncuk is 1.68. The percentage of Boncuk among other hunting dogs is in the 45.35% slice". When this answer was evaluated, it was observed that although the PT found the z-score correct, they determined the percentage by only taking into account the upper half of the distribution but not the lower half. The PTs who gave the answer 2.28%, which is another wrong answer, constituted 6.7% of all the PTs. For example, PS13 stated that "When compared to the average height of other hunting dogs, the height of the hunting dog, which Merve named Boncuk, is tall and is in the 2.28% slice." and drew the graph shown in Figure 18.

Figure 18

PT13's Answer



As a result of evaluating the PT's answer, it becomes apparent that they estimated Boncuk's height as between +2 and +3 standard deviations, which constitutes an operational error. There were 6.7% of PTs who determined the answer to be +2 standard deviation. As an example, PT41 explains:

> If hunting dogs have an average height of 38.5 cm and a standard deviation of 1.25 cm, Boncuk's height of 40.6 cm falls within the +2 standard deviation percentile. Because of this, we can say that Boncuk is taller than the average hunting dog.

Here, they made a mistake by saying that the height of Boncuk is within +2 standard deviation. No answers were given by two PTs. Table 11 shows the answers given by PTs to another question designed to reveal their understanding of interpreting the concept of a normal distribution in real-life situations.

Table 11

Answers to the Ninth Question

Code	Answers	Frequency/Percentage
4	Alperen is more successful in the physics test than the statistics test.	84 (70%)
3	Alperen is successful in both the physics and statistics tests	20 (16.6%)
2	-	-
1	Alperen scored 70 in the statistics test and 60 in the physics test. Therefore, he is more successful in the physics test.	14 (11.7%)
0	-	2 (1.7%)

Seventy percent of the PTs answered correctly and justified their answer to the ninth question. According to PT3, the following answer and rationale was given:

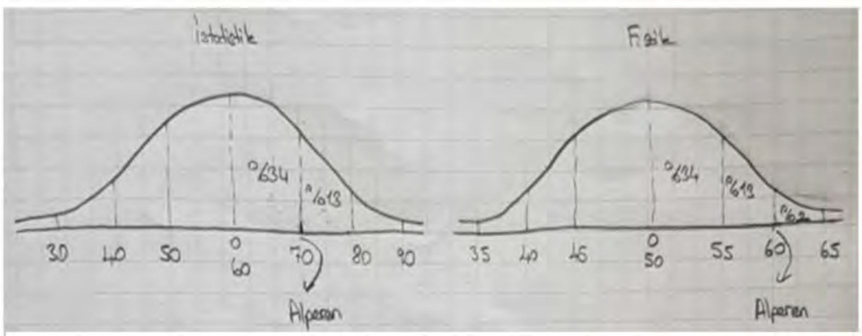
> We look at the normal distributions in both tests. We can say that Alperen is more successful in physics, even though he scored 70 in statistics and 60 in physics. The mean in the statistics test is 60 and the standard deviation is 10. In Alperen's case, the grade is 70. In the physics test, Alperen's grade is 60, with a mean of 50 and a standard deviation of five. Alperen's grade in the statistics test is in the range of +1 standard deviation. Alperen's grade in the physics test is in the range of +2 standard deviation. In physics, Alperen does better because when the standard deviation goes in the positive direction from the mean, this indicates that the student will perform better.

Although some PTs accurately commented on Alperen's state in the two tests separately, it was observed that they did not compare the two tests. For example, PT15 provided the answer shown in Figure 19. It is evident that although PT15 interpreted Alperen's statistics and physics grades, he/she did not compare them. The question was incorrectly answered by approximately one out of ten PTs (11.7%). The PTs failed to consider the standard deviation and arithmetic mean when interpreting the grades. PT50 explained that the student got 70 on the statistics test, thus making him more successful.

According to PT50, the physics test was less successful for him since he only scored 60. Two PTs didn't answer this question. According to the results of the tests, when the PTs were challenged on the basis of their real-life situations to interpret the concept of the normal distribution, the second question proved to be more successful than the first question in terms of interpreting the concept of the normal distribution. There was a greater variation in the mistakes made in the first question.

Figure 19

PT15's Answer



Alperen got 70 points in the statistics test and is in the 18% slice of the successful students. In the physics test, he is in the 2% slice of the successful students as he got 60. In general, we can say that Alperen is successful in both of these tests.

PTs' Knowledge on "Interpreting Normal Distribution Curves"

In order to reveal their understanding and skills about interpreting normal distribution curves, the PTs were asked two questions. They were asked a question to reveal their common content knowledge (Question 8), and a second question to reveal their specialized content knowledge (10a, 10b). These three questions were evaluated according to the rubric in Table 12.

Table 12

Evaluation Rubric for the 8th and 10th Questions

Correctness of the answers	Codes
Correct answer & correct justification	3
Correct answer & incomplete justification	2
Wrong answer	1
No answer	0

By using the evaluation rubric above, the answers to the eighth question by the PTs are listed in Table 13.

Table 13

Answers to the Eighth Question

Codes	Frequency/Percentage
3	94 (78.4%)
2	10 (8%)
1	14 (12%)
0	2 (1.6%)

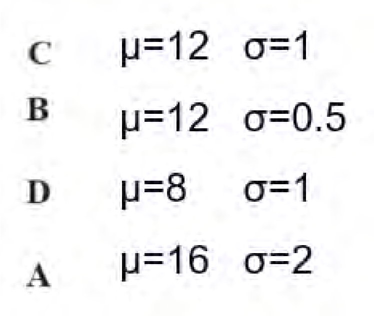
Most PTs answered the eighth question correctly and justified their answers (78.4%). According to PT38, graphs A and B have different means, but they have the same standard deviations. In their opinion:

Both C and B have the same mean, but C's standard deviation is smaller than B's. Therefore, when the standard deviation of graphs A, B, and C is compared, it is $\sigma_A = \sigma_B > \sigma_C$. When the standard deviation of the graph D is compared to those of the other graphs, it is $\sigma_A = \sigma_B > \sigma_C > \sigma_D$. Based on the means of the graphs A, B, C, and D, we can conclude that $\mu_D > \mu_C = \mu_B > \mu_A$.

Despite the fact that 8% of PTs gave correct answers, they did not justify their answers. Twelve percent of the PTs provided incorrect answers. PT55's response is shown in Figure 20.

Figure 20

PT55's Answer



Based on this answer, it can be concluded that the PT had difficulty understanding the standard deviation. A narrower curve (C) is chosen when the standard deviation is large, while a wider curve (B) is chosen when the standard deviation is small. The student was unable to understand the standard deviation and the arithmetic mean, while evaluating the A and D curves. Table 14 shows the answers provided by the PTs to question 10a.

Table 14

Answers to the Tenth Question (10a)

Codes	Frequency/Percentage
3	88 (73.3%)
2	22 (18.3%)
1	8 (6.6%)
0	2 (1.8%)

Many PTs were able to justify their answers and provided correct answers. According to PT21, all three distributions are normal. They stated that *"In all three graphs, the arithmetic mean, mode, and median are the same. The standard deviations are different from each other and the brown graph has the largest standard deviation, while the red graph has the smallest standard deviation."* About one out of every five PTs had the correct answer, but they failed to justify their answer (18.3%). Eight PTs answered the question incorrectly. For example, PT36 explained:

> It is a pointed distribution shown in red because it is compressed into a small area, which gives it a pointed appearance. Therefore, its standard deviation is small, and when its standard deviation is small, the group scores are close to each other, resulting in pointedness. Green indicates the normal distribution, which is characterized by a concentration of measurements at the middle and sparse measurements at the extremities. In addition, it is symmetrical. Due to the wide spread of the distribution, the brown distribution appears flattened. In this way,

we can make sense of it. The scores of the groups differ greatly from each other, and the standard deviation is high.

Here, the PT doesn't know what a normal distribution is. All of the distributions given in the questions are normal distributions, but they are all referred to in a different way. There was no response to this question from two PTs. Table 15 shows the answers of the PTs to the second part of the question.

Table 15

Answers to the Tenth Question (10b)

Codes	Frequency/Percentage
3	110 (91.6%)
2	8 (6.6%)
1	-
0	2 (1.8%)

This question was answered correctly and justified by the vast majority of PTs. According to PT7, their standard deviations are the same since both graphs have the same dispersion and kurtosis. They stated that "However, their arithmetic means differ. Their arithmetic means are as follows, from largest to smallest: brown>green>red because if we think of it like a number line, we can say that the brown distribution has the largest arithmetic mean." Although 6.6% of the PTs gave correct answers, they incompletely justified their answers. There was no PT who gave an incorrect answer to this question. Only one PT did not answer the question. When the PTs are evaluated in general on their ability to interpret normal distribution curves, they are considered successful. PTs generally justified their answers and provided correct answers. There was a minority of PTs who provided incorrect answers.

Conclusion and Discussion

This study aimed to reveal PTs' SMK about the normal distribution and to examine their reasons for making mistakes. According to the results, the PTs were better at identifying the normal distribution properties than standard deviation properties. It was noted in some statements that PTs tend to make more mistakes than others. As examples of these statements, we can give the statements "Standard deviation does not depend on the number of terms in the distribution" and "Standard deviation is the square root of the average squared deviation from the mean". It can be considered that the PTs misinterpreted these statements because they presented their justification on the basis of the standard deviation formula, which meant they made the evaluations only operationally.

Furthermore, it has been reported in the literature that PTs/teachers consider statistical concepts mainly operationally, without considering their meaning (Groth & Bergner, 2006; Ijeh & Onwu, 2012; Salinas-Herrera & Salinas-Hernández, 2022; Savard & Manuel, 2015). For example, Salinas-Herrera and Salinas-Hernández (2022) found that students had difficulty in understanding that the area between the horizontal axis and the curve is equal to one. Additionally, the PTs made errors when attempting to correlate the measures of central tendency and dispersion. As an example, in the statement "When we subtract five from each value in the dataset, the standard deviation decreases", the PTs were unable to determine that the standard deviation did not change, because the centre of the dataset did not change despite the decrease in the arithmetic mean. Thus, the measure of central tendency could not be related to the measure of dispersion. Batanero et. al. (2004) reported that PTs have difficulty relating statistical concepts to each other in the literature. Overgeneralizations can be found in some of the PT's statements. A property of the standard normal distribution, "It has a mean of zero and a standard deviation of one", was generalized into a property of all normal distributions by PTs. There is also evidence that pre-service teachers tend to generalize certain claims in the literature (Burgess, 2002; Kurt, 2015; Sorto, 2004).

When asked about interpreting concepts related to the normal distribution based on datasets, the answers of the PTs varied greatly. When it comes to finding the standard deviation, they are most successful. Eight out of ten students answered this question correctly, and PTs who made mistakes in this question did so because they performed operations incorrectly. A question asking where the data are located in the percentage resulted in nearly half the PTs (45%) making mistakes, and these mistakes showed greater diversity. There were two mistakes that were made the most. PTs were observed to establish a direct relationship between the data distances from the arithmetic mean and their areas in the first study. There is 34% of the data located one standard deviation away from the arithmetic mean, while 19.15% of the data are located 0.5 standard deviations away from it. In spite of this, the PTs believed that when the z score was half, 17% of the data fell into the 0.5 standard deviation range since the data was also halved. A similar overgeneralization can be attributed to the PTs here. There was a belief that there was a similar ratio between z scores and the area under the normal distribution. In their opinion, halving the z score results in halving the area. A possible explanation is that the PTs had difficulty relating probability values to the area covered by the curve. There have been similar difficulties emphasized in the literature (Bansisal, 2014; Batanero et al., 2004). The study conducted by Batanero et al. (2004) revealed that undergraduate students had difficulty calculating probabilities under normal curves and presenting them graphically.

Furthermore, the PTs often made mistakes when evaluating normal distributions. It was found that although the PTs were able to calculate z-scores correctly, they evaluated the data as only being located on the right of the arithmetic mean, despite the fact that the data were located both right and left of it. When extracting values from the dataset and analyzing them, PTs had the greatest difficulty determining the curve shape, standard deviation, and arithmetic mean of the new distribution. Over half of the PTs answered incorrectly (51.7%), and about one third of the PTs who answered correctly justified it incorrectly. We observed that only 35% of the PTs understood what changes to the dataset would mean for the shape of the distribution and other statistical concepts. A common characteristic of those who gave incorrect answers was that they could not relate statistical concepts together. An important aspect of this subject is illustrating the areas under a normal distribution and connecting them to various concepts. In the literature, similar findings have been reported, and students are reported to have difficulty determining probabilities and associating them with other concepts (Batanero et al., 2004).

Analysis of PTs' interpretations of the normal distribution by referring to real-life situations revealed remarkable results. According to the results, both questions asked in different contexts resulted in differing rates of correct answers by the PTs. The PTs who answered the question regarding dogs' height correctly accounted for 73.3% of the correct answers, but only 18.3% were able to justify their answers. Taking the grades related question into account, it was found that 86% of the PTs responded correctly, and 70% of them could justify their responses. As shown here, the PTs evaluated the situations differently depending on the context. The PTs are more familiar with the context of grades from different tests, which may explain their greater success rates. In examining the mistakes made, it was found that the reasons were similar to those in the previous section. These reasons are primarily based on the inability to make associations between concepts and operational evaluation of concepts. Further, PTs may experience epistemological anxiety as a result of their difficulties (Wilensky, 1995, 1997). Epistemological anxiety is described by Wilensky (1995; 1997) as the feeling of indecision and confusion experienced by individuals when faced with normal distribution problems.

In evaluating the PTs' interpretations of normal distribution curves, more than 70% answered all three questions correctly, along with acknowledging that their answers were justified. Two categories can be used to evaluate the reasons for the mistakes made. As a starting point, PTs have trouble associating concepts like standard deviation and arithmetic mean. It may also be due to the prototype perceptions developed by PTs. Different words, such as pointed and flattened, were used by PTs to describe the various representations of the normal distribution. There may have been an emergence of such a result since the PTs only created one prototype for the normal distribution. A literature review revealed that prototype perceptions may prevent individuals from expanding their understanding of related concepts.

PTs responded differently to both questions (three and 10a-10b) in which specialized content knowledge was questioned. Despite being able to interpret the given visual representations more accurately, they were unable to create them themselves with the same success. According to the literature, individuals have difficulty transforming datasets into visual representations (Bruno & Espinel, 2009; Meletiou-Mavrotheris & Lee, 2005).

Researchers recommend using and associating different statistical concepts (e.g., centre, skewness) and evaluating distributions as a whole. It is also important to establish a relationship between real data and distributions (Batanero et al., 2004). PTs were required to evaluate and interpret multiple situations related to normal distributions in the current study. This study concludes that normal distributions, as described in other studies (e.g., Batanero et al., 2004), involve many different statistical concepts and ideas that need to be integrated and interrelated. The main emphasis is not on developing computational skills related to statistical concepts, but rather on making profound associations between concepts and taking into account many concepts at once (Batanero et al., 2004). Prior to studying the normal distribution, students should understand probabilities, density curves, kurtosis and skewness, as well as histograms (Batanero et al., 2004). Lack of knowledge about these concepts may also have contributed to PTs' difficulties.

Suggestions and Limitations

For the SMK, the normal distribution is an important concept that the teacher should be familiar with (Ball et al., 2008). As a result, the findings of this study suggest important implications that are supported by other research findings (e.g., Delport, 2022). Students and PTs who have difficulties understanding the normal distribution may have difficulty understanding other statistical concepts as well. Moreover, Batanero et al. (2004) and McLeod (2019) claim that it can be used to model many natural and psychological phenomena. This concept needs to be supported by applications in order for PTs to gain a deeper understanding. In order to improve their understanding of this concept, it may be helpful to work on different tasks to overcome the difficulties they experience. The current study also did not include any computer aided software due to technical difficulties, which may be considered as a limitation of the study. A comparative study could be conducted in the future to find out how these software programs affect PTs' understanding.

A "no answer" code was applied to PT responses if they failed to answer the question. As for the student's knowledge of the questions, it's unclear if they are aware of them or prefer to not write anything. It is possible to consider this a limitation of the study. The identification of these two situations can be improved by conducting qualitative research in future studies. Further, the study examined the reasons for PTs' mistakes. However, the study was unable to determine the origins of the reasons mentioned. PTs' thinking in depth could be analyzed by conducting semi-structured interviews in future studies as a way to understand the reasons behind their mistakes in relation to these concepts.

Different forms of representations can also help PTs overcome the difficulties they encounter. As Wood et al. (2018, p.299) suggests a transition from "concrete to visual and abstract representations" can help students grasp the concept of the normal distribution better. It is possible for PTs to make more in-depth associations between concepts by connecting datasets and representations of these datasets.

The author received no financial support for the research, authorship, and/or publication of this manuscript.

Nadide Yılmaz (nadideylmz70@gmail.com) is an associate professor in the Department of Mathematics and Science Education at the Karamanoğlu Mehmetbey University, Karaman, Turkey. She received her doctorate from the Hacettepe University, Ankara, Turkey. She is interested in mathematics teacher knowledge and teacher education. Her research focuses on statistics knowledge for teaching and teacher training programs especially lesson study

References

Aslan, K. (2003). Eğitim fakültelerinin yeniden yapılandırılmalarına ilişkin bir değerlendirme [An evaluation of the restructuring of education faculties]. *Balıkesir University Journal of Social Sciences Institute, 6*(9), 23–37.

Baillargeon, N. (2005). *Petit cours d'autodéfense intellectuelle* [Short intellectual self-defense course]. Lux Publisher.

Ball, D. L. (2003). *What mathematical knowledge is needed for teaching mathematics?* Secretary's Summit on Mathematics, U.S. Department of Education.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177/0022487108324554.

Bair, S. L., & Rich, B. S. (2011). Characterizing the development of specialized mathematical content knowledge for teaching in algebraic reasoning and number theory. *Mathematical Thinking and Learning, 13*(4), 292-321. https://doi.org/10.1080/10986065.2011.608345.

Bansilal, S. (2014). Using an APOS framework to understand teachers' responses to questions on the normal distribution. *Statistics Education Research Journal, 13*, 42-57.

Batanero, C., Tauber, L., & Sanchez, V. (2004). Students' reasoning about the normal distribution. In D. Ben-Zavi, & Garfield, J. (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking.* (pp. 257-276). Kluwer Academic Publishers.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education (GAISE) report II.* American Statistical Association and National Council of Teachers of Mathematics.

Baştürk, S. (2009). Perspectives of student teachers of secondary mathematics education on mathematics teaching in faculty of arts and science. *İnönü University Journal of the Faculty of Education, 10*(3), 137–160.

Baştürk, S. (2011). Mathematics teacher candidates' evaluations of teaching and learning process in faculty of education. *International Journal of Human Sciences, 1*(8), 58–94.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157.

Bruno, A. & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training, *International Journal of Mathematical Education in Science and Technology, 40*(4), 473-493. https://doi.org/10.1080/00207390902759584.

Burgess, T. (2002). Investigating the "data sense" of preservice teachers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics* (pp. 1-6). International Association for Statistics Education.

Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 259–270). Springer. https://doi.org/10.1007/978-94-007-1131-026

Canada, D. L. (2008). Conceptions of distribution held by middle school students and preservice teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference* (pp.1-6). ICMI and IASE.

Chaput, J.S; Crack, T.F. & Onishchenko, O. (2021). What quantity appears on the vertical axis of a normal distribution? A student survey. *Journal of Statistics and Data Science Education, 29*(2), 192-201. https://doi.org/10.1080/26939169.2021.1933658.

Delport, D. H. (2022). How the number line can be used to promote students' understanding of the normal distribution. *Numeracy, 15*(2), 1403. https://doi.org/10.5038/1936-4660.15.2.1403.

Even, R., Tirosh, D., & Markovits, Z. (1996, June). *Teacher subject matter knowledge and pedagogical content knowledge* [Paper presentation]. Conference of the International Group for the Psychology of Mathematics Education (PME 20), Valencia, Spain.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education.* McGraw-Hill.

Franklin, C., Kader, G., Bargagliotti, A., Scheaffer, R., Case, C., & Spangler, D. (2015). *Statistical education of teachers*. American Statistical Association.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* Teachers College Press.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*(5), 427-437. https://www.jstor.org/stable/30034960.

Groth R. E. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning, 15*(2), 121-145. https://doi.org/10.1080/10986065.2013.770718

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning, 8*, 37-63. https://doi.org/10.1207/s15327833mtl0801_3.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235.

Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371.

Hill, H., Sleep, L., Lewis, M., & Ball, D. (2007). Assessing teachers' mathematical knowledge. What knowledge matters and what evidence counts? In F. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111-155). Information Age Publishing.

Hinders, D. (2010). *5 steps to AP statistics*. McGraw Hill Companies.

Huck, S.W., Cross, T.L. & Clark, S.B. (1986). Overcoming misconceptions about z-scores, *Teaching Statistics, 8*(2), 38-40. https://doi.org/10.1111/j.1467-9639.1986.tb00624.x.

Ijeh, S. B. & Onwu, G.O.M. (2012). *Using PCK to deal with learner's conceptions (Misconceptions and Preconceptions) in statistics teaching.* University of Africa

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. *Forum: Qualitative Social Research*, *11*(2), 1450. https://doi.org/10.17169/fqs-11.2.1450

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D.E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). National Council of Teachers of Mathematics.

Kurt, G. (2015). An investigation of understanding of preservice elementary mathematics teachers (PEMT) about data displays. In K. Krainer & N. Vondrová (Ed.), *Ninth Congress of the European Society for Research in Mathematics Education (CERME 9)* (pp. 615-621). Hall Archives.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Erlbaum.

Makar, K. (2018). Theorising links between context and structure to introduce powerful statistical ideas in the early years. In A. M. Leavy, M. Meletiou-Mavrotheris, &E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking*. Springer.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

McLeod, S. A. (2019). *What is a normal distribution in statistics?* Simply Psychology. https://www.simplypsychology.org/normal-distribution.html

Meletiou-Mavrotheris, M. & Lee, C. (2005, February). *Exploring ıntroductory statistics students' understanding of variation in histograms* [Paper presentation]. 4th Congress of the European Society for Research in Mathematics Education, San Feliu de Guixols.

Mewborn, D. (2003). Teaching, teachers' knowledge, and their professional development. In N. C. Mathematics, J. Kilpatrick, W. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 45-53). The National Council of Teachers of Mathematics.

Miles, M, B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded Sourcebook*. Sage.

Ministry of National Education [MoNE]. (2018). *Matematik dersi (İlkokul ve ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar) öğretim programı* [Primary and middle school mathematics curricula for grades 1, 2, 3, 4, 5, 6, 7, and 8]. Author.

Mooney, E., Duni, D., van Meenen, E., & Langrall, C. (2014). Preservice teachers' awareness of variability. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics* (pp. 1-4). International Statistical Institute.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Author.

Özmen, Z. M. (2015). *Examination of the statistical literacy levels of students from different undergraduate programs* (Publication no. 407693) [Doctoral dissertation, Karadeniz Technical University]. Council of Higher Education Thesis Center.

Pfannkuch, M., & Reading, C. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal*, *5*(2), 4–9.

Ponte, J. P., & Chapman, O. (2016). Prospective mathematics teachers' learning and knowledge for teaching. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 275-296). Routledge.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Salinas-Herrera, J., & Salinas-Hernández, U. (2022). Teaching and learning the notion of normal distribution using a digital resource. *Canadian Journal of Science, Mathematics and Technology Education. 22*, 576–590. https://doi.org/10.1007/s42330-022-00226-1.

Savard, A. & Manuel, D. (2015*)*. Teaching statistics in middle school mathematics Classrooms: Making links with mathematics but avoiding statistical reasoning. In M. Marshman, V. Geiger & A. Bennison (Eds.), *Proceedings of the 38th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 547-554). MERGA.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.https://doi.org/10.3102/0013189X015002004.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411.

Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its application to teaching* (Publication No. 3146114) [Doctoral dissertation, Michigan State University]. ProQuest Dissertations and Theses Global.

Sullivan, P., & Wood, T. (Eds.). (2008). *The international handbook of mathematics teacher education. Volume 1: Knowledge and beliefs in mathematics teaching and teaching development.* Sense.

Tchoshanov, M., Cruz, M.D., Huereca, K., Shakirova, K., Shakirova, L. & Ibragimova, A. N. (2017). Examination of lower secondary mathematics teachers' content knowledge and its connection to students' performance. *International Journal of Science and Mathematics Education*, *15*, 683–702. https://doi.org/10.1007/s10763-015-9703-9.

Walshaw, M. (2012). Teacher knowledge as fundamental to effective teaching practice. *Journal of Mathematics Teacher Education, 15*, 181-185. https://doi.org/10.1007/s10857-012-9217-0.

Watkins, A. E, Scheaffer, R. L., & Cobb, G. W. (2008). *Statistics in action*. Key Curriculum Press.

Wilensky, U. (1995). Learning probability through building computational models. In D. Carraher & L. Meira (Eds.), *Proceedings of the 19th PME Conference* (Vol. 3, pp. 152–159). PME.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics, 33*, 171–202.

Woods, D. M., Leanne, K. G. & Deni, B. (2018). Number sense on the number line. *Intervention in School and Clinic, 53*(4), 229–236. https://doi.org/10.1177/1053451217712971.

Vermette, S. & Savard, A. (2019). Necessary knowledge for teaching statistics: Example of the concept of variability. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and Trends in Current Statistics Education ICME-13 Monographs* (pp.225-244). Springer.