



## Primary Pre-service Teachers' Knowledge of the Concepts of Mean and Median

Cathrine Kazunga, Sarah Bansilal & Lytton Chiromo

**To cite this article:** Cathrine Kazunga, Sarah Bansilal & Lytton Chiromo (2023) Primary Pre-service Teachers' Knowledge of the Concepts of Mean and Median, African Journal of Research in Mathematics, Science and Technology Education, 27:3, 367-382, DOI: [10.1080/18117295.2023.2277984](https://doi.org/10.1080/18117295.2023.2277984)

**To link to this article:** <https://doi.org/10.1080/18117295.2023.2277984>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 28 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 1391



View related articles [↗](#)



View Crossmark data [↗](#)



# Primary Pre-service Teachers' Knowledge of the Concepts of Mean and Median

Cathrine Kazunga, Sarah Bansilal <sup>\*</sup>, and Lytton Chiromo

*University of KwaZulu–Natal, Durban, KwaZulu–Natal*

*\*Corresponding author. Email: [bansilals@ukzn.ac.za](mailto:bansilals@ukzn.ac.za)*

The development of statistical literacy has become a key outcome of mathematics curricula in various countries. The purpose of the study was to explore primary pre-service teachers' (PSTs) knowledge of aspects of measures of central tendency, which is a central concept in statistical literacy. The 183 PSTs' written responses to four tasks were analysed in terms of demands elucidated by the Mathematical Knowledge for Teaching (MKT) framework. The results showed that almost all the participants were able to carry out simple computations of the mean and median and were generally able to identify learner errors. However, they found it much harder to solve problems that required unpacking of the procedures and comparison of different measures of central tendency. They also struggled with providing suitable feedback based on learner errors. This study shows that primary school PSTs need more support in developing these MKT skills with respect to the concepts of the mean and median.

**Keywords:** *Mean; median; statistical literacy; pre-service teachers; mathematical knowledge for teaching*

---

## Introduction

Statistical literacy refers to knowledge of statistical concepts, skills in statistical methods and the capacity to critically evaluate statistical information (Garfield & Ben-Zvi, 2009). It has gained increasing attention in recent years especially for dealing with the demands of the twenty-first century. Many mathematics school curricula across the world have been revised to reflect the importance of developing statistical literacy amongst their learners (Burrill et al., 2023), including that of South Africa (Bansilal, 2023). Hence, it is important to teach basic statistical concepts in a manner that encourages the interpretation of information and critical thinking. Among these basic concepts are measures of central tendency, which are some of the earliest statistical concepts that learners encounter in school.

Teachers also need a broad personal understanding of statistical concepts that will enable them to develop statistical literacy skills in their learners. Kalobo (2016) asserts that it is necessary for teachers to first understand and to be able to describe the statistical learning outcomes themselves before they can develop their learners' statistical literacy skills and statistical reasoning (Kalobo, 2016). Hence it is important that PSTs are given opportunities to develop their understanding of the key ideas in statistical literacy. Ball et al. (2008: 395) use the construct of 'Mathematical Knowledge for Teaching' (MKT) to capture the essence of the teacher knowledge that is needed to mediate effective learning in the classroom. These authors have identified some demands faced by mathematics teachers as they teach, organised according to sub-domains of the MKT model. It is these demands, related to the teaching of the concepts of mean and median as measures of central tendency, that are the focus of our interest in this paper. The research question that underpins this study is: what do pre-service teachers' responses reveal about their readiness to handle demands

associated with mathematical knowledge for teaching subdomains with respect to the mean and median? (Ball et al., 2008)

It is hoped that this study will help unpack some of the demands associated with teaching the concept of measures of central tendency, while also identifying how prepared PSTs are to handle these teaching challenges that they will face.

## Literature Review

There are three main statistical measures of central tendency: the mean, median and mode. Each of these measures describes a slightly different central position within a dataset (Holmes et al., 2018). Kristanto (2018) suggested that the interpretation of the mean can be done under the backdrop of connecting the mean to contexts or models. One well-known model is that of fair share, where the values in a dataset are shared from one to another in order to obtain the same value (Kristanto, 2018; Leavy & O'Loughlin, 2006). Related to the idea of fair share is that of the mean as a balance point on a number line, similar to that of a centre of gravity (Leavy & O'Loughlin, 2006), where the sum of the distances from the mean of all the points above the mean is equal to the sum of the distances from the mean of all the data points below the mean. Computationally, the arithmetic average is the score around which deviations in one direction exactly equal deviations in another direction. These links of the mean with fair sharing and number line representations point to ways in which learning in the Foundation Phase grades can be connected with teaching and learning linked to this central tendency measure. The median is based on ordering data and is thus also founded on fundamental Foundation Phase skills. Similarly, the mode depends on counting for the most frequently occurring phenomena in a dataset. Hence it is clear that even though measures of central tendency are not introduced formally at Foundation Phase, the big ideas that underpin the concept start to be developed at this phase.

Groth and Bergner (2006), in a study on PSTs' responses to the assessment of their knowledge of the three measures of central tendency, identified different levels of thinking in regards to comparing and contrasting the measures of central tendency. A low level of understanding (unistructural) involves just knowing the definitions and being able to describe the procedures for finding the different measures of centre. A multistructural level is when they are able to see the three measures as objects for data analysis rather than just procedures to be done. A more sophisticated level of understanding (relational level) seems to be when students come to realise that 'measures of central tendency tell what is "typical" or "average" about a set of data' (Groth & Bergner, 2006: 47). Being able to distinguish between and select the most suitable measure of centre for the dataset being analysed marks an even higher level of understanding (extended abstract level) (Groth & Bergner, 2006).

Goth and Bergner's (2006) research suggests that primary PSTs tend to stick to procedural conceptions of measures of centre; do not have an understanding of multiple measures of centre at their disposal; and struggle with selecting the most appropriate measure of centre in describing a given dataset. Other results about PSTs indicate that they are more pre-occupied with the procedures to calculate the mean and less concerned with its interpretation and they also show a preference for standard algorithms in calculating the mean and mode (Kristanto, 2018). Research also suggests that PSTs have mainly computationally based, rather than conceptually based understandings; that they confuse the mean with the mode; and that their mastery of mathematical content knowledge is inadequate to achieve effective teaching (Leavy & O'Loughlin, 2006).

One method that can be used to judge the reasonableness of results across different procedures is estimation, yet it has a limited representation in many curricula (Dowker, 1992; Sunde et al., 2022). Estimation can be seen as a systematic process that starts off with an informed guess using the data available and then adjusting the guess as more evidence is taken into account. Computational estimation (as opposed to measurement or quantity estimation) involves making 'reasonable guesses as to the approximate answers to arithmetic problems, without or before actually doing the calculations' (Dowker, 1992: 45).

There is a growing emphasis on the development of statistical literacy skills which include the ability to understand and interpret statistical data. It also involves engaging with the meaning, interpretation

and selection of the most suitable calculations such as those for the measure of centre (Groth & Bergner, 2006). The sophisticated ways in which mathematics and statistics concepts are used in the media mean that ordinary people are required to have knowledge of the limitations of using certain procedures and methods of representing data (Gal & Geiger, 2022). However, Lampen (2015) notes that the instructional approaches used by teachers in teaching statistics are mostly restricted to computing well-defined mathematical procedures such as the mean. Many mathematics teachers still rely on traditional methods instead of employing data-driven methods focusing on the development of statistical reasoning (Wessels & Nieuwoudt, 2011). Harrell-Williams et al. (2019) in their study about teachers' self-efficacy to teach statistics identified three levels of engagement with the data ranging from being able to read data to being able to read between the data and being able to read beyond the data. Teachers' self-efficacy estimates were lowest in the reading beyond the data level, which requires critical interpretation of the data. A related finding by Umugiraneza et al. (2022) suggests that teachers are not as confident in teaching statistics in a way that develops critical thinking as they are in teaching well-known mathematics procedures. These studies indicate that there is a need for a deeper understanding of the specific outcomes within statistics as well as how these concepts develop from Foundation Phase across to higher phases.

## Framework

Many researchers in the field of teacher education have generated descriptions and definitions to try to encapsulate the kind of knowledge that is needed to mediate learning in the classroom. The term 'Mathematical Knowledge for Teaching' is used to describe the nature of the knowledge needed to carry out the work of teaching mathematics (Ball et al., 2008). The MKT model includes two domains: Subject Matter for Teaching and Pedagogical Content Knowledge. Subject Matter for Teaching is further divided into two subdomains called common content knowledge (CCK) and specialised content knowledge (SCK). Common content knowledge is 'mathematical knowledge and skill used in settings other than teaching' and includes knowing the content to be taught as well as being able to recognise learners' wrong answers (Ball et al., 2008: 399). Specialised content knowledge refers to the 'mathematics knowledge and skill unique to teaching'. Specialised content knowledge goes beyond the knowledge taught to learners and 'is not typically needed for purposes other than teaching' (Ball et al., 2008: 400). It includes the ability to look at patterns in errors and to 'unpack' the mathematics which is taught (Ball et al., 2008: 399–400).

The Pedagogic Content Knowledge (PCK) domain includes knowledge of content and students (KCS), which requires an 'interaction between specific mathematical understanding and familiarity with students and their mathematical thinking' (Ball et al., 2008: 401). Although recognising an answer as incorrect is a skill related to CCK, being able to recognise common errors and explain possible causes is a characteristic feature of KCS. An important task of the teacher is that of providing feedback to students about their errors or misconceptions and advice on how to improve their understanding, which we believe is a feature of KCS.

This study is concerned with demands relating to the three subdomains (CCK, SCK and KCS) elaborated above with respect to the concepts of mean and median using the four tasks in Table 1.

Ball and colleagues caution that there are overlaps across the subdomains depending on the context of analysis and the person's perspective. However, the importance of the model lies in its elaboration of the notion of pedagogic content knowledge, which helps us to better understand some of the demands of teaching the particular concept. While the items in Table 1 were meant to target a particular domain, PST responses may or may not be related to that domain depending on what they said.

## Methodology

This qualitative study utilised an interpretive approach because the main goal of this study was to understand the students' interpretations of reality (Cohen et al., 2018) concerning the concepts of

**Table 1.** Summary of demands of the four tasks in the study

Tasks	Subdomain of Ball et al.'s (2008) framework and the associated demands	Coding rubric used in analysis
<p><b>Task 1</b></p> <p>Peter teaches karate classes. One of his classes has the following ages (in years): 4, 29, 4, 3, 4, 11, 16, 14, 17, 3.</p> <p>1.1. Calculate the mean of the ages. 1.2. Calculate the median of the ages. 1.3. Explain why the mean is not a good measure of central tendency for the ages.</p>	<p>This task targets demands within the CCK subdomain:</p> <ul style="list-style-type: none"> <li>calculating the mean and median of a given dataset.</li> </ul> <p>Item 1.3 also targets the SCK subdomain:</p> <ul style="list-style-type: none"> <li>recognising why a certain measure of central tendency is more appropriate in a context than another.</li> </ul>	<p>1.1. Responses coded as correct or incorrect. 1.2. Correct or incorrect. 1.3. Codes developed on an inductive basis individually by authors and condensed to five codes.</p>
<p><b>Task 2</b></p> <p>The mean score of a group of 17 learners is 65. Two other learners whose scores are 89 and 85 were added to the group. What is the new mean of the group of learners?</p>	<p>This task targets demands associated with SCK subdomain:</p> <ul style="list-style-type: none"> <li>unpacking the mathematics behind the calculation of the mean.</li> </ul>	<p>Coding: completely correct; slip or careless error in calculation; tried to generate a new dataset; number grabbing; blank.</p>
<p><b>Task 3</b></p> <p>Learners were asked this question: 'Find the mean of the dataset: 12; 11; 15; 28; 39'.</p> <p>Ayanda found the mean to be 105.</p> <p>3.1. Discuss Ayanda's possible understanding of the mean 3.2. Explain how the use of estimation and a conceptual understanding could have helped Ayanda realise that her answer was incorrect.</p>	<p>This item targets demands related to KCS subdomain:</p> <ul style="list-style-type: none"> <li>identifying the learner's misconception—the learner did the first step of finding the sum of the numbers but did not divide the sum by the number of scores;</li> <li>providing hints about how learners could use estimation to recognise their error.</li> </ul>	<p>3.1 codes: identified or did not identify cause of error. 3.2 codes developed on an inductive basis individually by authors—condensed to seven codes.</p>
<p><b>Task 4</b></p> <p>Learners were asked to find the median of the following dataset: 18; 15; 23; 12; 19. Some students provided the answers given below. Evaluate each of the following learner responses to identify what misconception, if any, the learner had. How can you address each misconception?</p> <p>4.1. Median is 23 because 23 is the middle number. 4.2. The median is 17.4. 4.3. The median is 18.</p>	<p>This task targets demands related to the KCS subdomain, although this depends on how the students respond to the task:</p> <ul style="list-style-type: none"> <li>identifying the cause of the learner error—the learner has worked with the data list as it was, without ordering it first, producing incorrect answers for the median (4.1);</li> <li>identifying the cause of the error—the given number is the mean and not the median (4.2);</li> <li>identifying that a given number is the correct answer for the median (4.3);</li> <li>providing feedback on how these errors can be addressed.</li> </ul>	<p>Coding for 4.1 and 4.2 identified the cause of error and did not go further; identified the error and described feedback they would give to help learner; identified the error and explained the computation of the median. 4.3 identified or did not identify that 18 was the correct answer.</p>

CCK, Common content knowledge; KCS, knowledge of content and students; SCK, specialised content knowledge.

mean and median. The participants in the study were 183 primary Bachelor of Education PSTs specialising in the Foundation and Intermediate Phases (FP and IP), at a South African University. A case study research design was employed where the case was the group of participants. The intention was to understand whether PSTs could problem solve and engage at higher levels of cognitive reasoning in relation to central measures of tendency. It is important to note that the big ideas behind central measures including counting to identify the number that appears the most often, ordering sets of numbers and the notion of fair share are meant to be developed informally from the Foundation Phase. In South African schools, learners are introduced to the formal description of the mode in Grade 5, the median in Grade 6 and the mean in Grade 7, at which point they are also expected to select the most appropriate measure in a situation. Although theoretically Grade 7 forms part of the Senior phase, because of its location in the primary school, Grade 7 learners are usually taught by Intermediate Phase teachers. It is therefore important for Intermediate Phase teachers to be able to teach Grade 7 content. The participants in the study were being trained across both the FP and IP, making it imperative for them to understand the trajectories between these phases. In this study, we decided to focus on two measures of central tendency. Since the mode of a dataset does not require any calculations but simply involves identifying the data point that appears most frequently, we chose to focus on the two more complex measures of mean and median, enabling us to go into greater depth. In addition, we explore the participants' identification of errors and misconceptions related to these measures of central tendency and their ability to help their learners to address the errors and misconceptions.

Data were generated from 1647 written responses, consisting of 183 answers to the nine items organised as four tasks. Depending on the items, different coding schemes were used as detailed in [Table 1](#). Images of the exact responses of the students are presented in those cases where the writing is legible. When the images were not sufficiently clear, we reproduced the responses.

## Results

The results of each of the tasks are arranged according to those that target the CCK, SCK and the KCS subdomains respectively.

### **Results for Task 1 (CCK; SCK)**

Task 1 (see [Figure 1](#)) targeted demands from the CCK and SCK subdomains. It focused on a context of participants in a karate class. Students were asked why the mean was not a good measure of central tendency for the ages of participants, where half the group were younger than 5 years and the other half were 11 years or older.

#### *Results for Items 1.1 and 1.2*

There were 176 participants who correctly calculated the mean, while 180 participants correctly calculated the median.

Peter teaches karate classes. One of his classes has the following ages (in years):

4, 29, 4, 3, 4, 11, 16, 14, 17, 3.

1.1 Calculate the mean of the ages.

1.2 Calculate the median of the ages.

1.3 Explain why the mean is not a good measure of central tendency for the ages.

**Figure 1.** Task 1

**Table 2.** Results for Item 1.3

Description of category	Frequency
MA: Large or huge gaps between the dataset or mean not accurate measure because of extremes	40
MA2: Suggested that it was not a balance point or true point of the distribution	23
MA3: Wide range between upper half and lower half	31
MA4: General description or comment not related to why mean is not suitable	35
MA5: No response	54

### Results for Item 1.3

The summary of the participants' explanations of why the mean would not be a good measure of central tendency in this context, appears in [Table 2](#).

Almost 30% of the class (MA6) did not provide a response to the item, suggesting that they found it harder to provide an explanation related to the context, in comparison with doing a calculation as in Items 1.1 and 1.2, where more than 90% got the calculations correct.

There were 35 students (MA4) who gave a description not related to the question (see [Figure 2](#)).

There were 40 explanations (MA1) that were related to the large gaps between the data points, as explained by S121 ([Figure 3](#)). Student S121 seems to suggest that because the range of the data is very big, the number 10.5 cannot represent a central point of the data.

There were 23 explanations that drew upon the balance point analogy (MA2). Many students alluded to the meaning of mean as a point that balances the other values, e.g. student S74 ([Figure 4](#)) said that it is not easy to share the ages to get the balance point of the distribution.

Some students (31) alluded to the fact that there was a large gap between the upper and lower halves of the dataset (MA4) as illustrated in [Figure 5](#).

### Responses to Task 2 (SCK)

The students were asked to respond to the following task which targeted SCK skills.

The mean score of a group of 17 learners is 65. Two other learners whose scores are 89 and 85 were added to the group. What is the new mean of the group of learners?

There were five categories of responses as detailed in [Table 3](#).

Most students (95) solved this problem correctly (NM1), while 10 students (NM2) made a small calculation slip. Ten students (NM3) tried to generate a list of 17 numbers first whose mean was 65, before adding the two new numbers, but most of them struggled with this approach (see [Figure 6](#)).

Many students used a 'number grabbing' approach where they 'grab' numbers appearing in the instructions and carry out arbitrary operations (Schoenfeld, 1988; Mbonambi & Bansilal, 2014). There were 59 responses taken as number grabbing ([Figure 7](#)).

### Responses to Task 3 (KCS)

This task ([Figure 8](#)) targeted KCS skills such as identifying the possible misconception behind the error, and providing feedback on how estimation could be used to recognise that the answer was incorrect.

### Responses to Item 3.1

Most of the participants (171 or 93%) were able to clearly explain Ayanda's error that she added all the scores but did not divide the sum by the number of scores. However, 12 participants (7%) did not

*It is because it is not arranged in one range.*

**Figure 2.** Response by student S166 to Item 1.3, showing a general or unrelated statement

*The gap between the lowest value and the highest value is great and a mean of 10.5 years does not represent a central point.*

**Figure 3.** Response by S121 to Item 3.3, about the large gaps between the numbers

$$10.5 \approx 11$$

*It is not a good representation with age as its not easy to share the ages and for the balance point of a distribution since there are different there would be no equal distribution.*

**Figure 4.** Responses of S74 to Item 1.3, alluding to the balance point understanding of the mean

*There is a wide range between the lower half and the upper half of the values, the mean is affected, it is not a good average*

**Figure 5.** Response by student S120 to Item 1.3 about the lower and upper halves of the data

**Table 3.** Responses to Task 2.

Category	Frequency
NM1: Correct calculation of the new mean	95
NM2: Slip or careless error in calculation	10
NM3: Generated a new dataset	10
NM4: Engaged in number grabbing—doing operations using numbers which may not make sense	59
NM5: No response	9

address her error explicitly. They either gave a general formula for calculating the mean or they illustrated how this example should have been done (Figure 9).

*Responses to Item 3.2*

Item 3.2 probed them further about the feedback they could give to Ayanda about how she could have recognised her answer was incorrect using estimation. The students’ explanations varied and were placed into seven categories, which are not in any particular order, as summarised in Table 4.

90, 25, 40, 18, 109, 125, 25, 189, 61, 99  
 83, 69, 24, 55, 65, 48, 24  
 $90 + 25 + 40 + \dots + 48 + 24$   
 $1117 \div 17 = 65,74588$   
 $\approx 65$   
 If 89 and 85 were added  $1117 + 89 + 85 = 1291$   
 Its 19 because 2 terms were added  $1291 \div 19 = 67$

**Figure 6.** Response of student S84 to Task 2, struggling to generate 17 numbers whose mean was 65

The mean score of a group of 17 learners is 65. Two other learners whose scores are 89 and 85 were added to the group. What is the new mean of the group of learners? (

$$\begin{aligned}
 &65 + 89 + 85 \\
 &= \frac{239}{19} \\
 &= 12,6
 \end{aligned}$$

**Figure 7.** Example of number grabbing from S47

The following item appeared in a quiz after instruction about the mean and median:

*Find the mean of the dataset: 12; 11; 15; 28; 39*

*Ayanda found the mean to be 105.*

3.1 Discuss Ayanda's possible understanding of the mean, based on her answer.

3.2 Explain how estimation and conceptual understanding would have helped Ayanda to realise her answer was incorrect.

**Figure 8.** Instructions for Task 3

$$\begin{aligned}
 &11; 12; 15; 28; 39 \\
 &11 + 12 + 15 + 28 + 39 \\
 &= \frac{105}{5} \\
 &\text{mean} = 21 \rightarrow
 \end{aligned}$$

**Figure 9.** The response of S58 to Item 3.1 showing the computation of the mean

Fifty students identified Ayanda's error but did not go further to explain how estimation could have helped her realise the answer was not correct (Figure 10).

In Category A6, 44 students provided general descriptions of what Ayanda could have done, without referring specifically to the error or any estimation skills (Figure 11).

There were 30 students in Category A4 whose response was to repeat the explanation of how to go about finding the mean of a set of numbers (Figure 12). This seems to be a common response of many teachers when their learners get a wrong answer — they tend to explain the calculation procedure again. Perhaps some teachers believe that the more the learners are told something, the easier it is for them to learn (Umugiraneza et al, 2017).

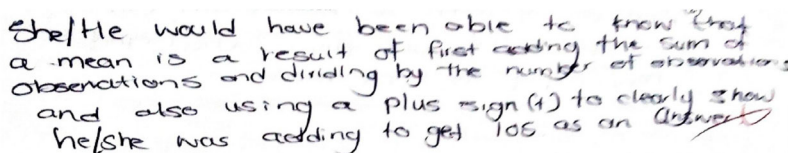
*She didn't divide the sum by the total values*

**Figure 10.** Response of S146 to Item 3.2, where she identified Ayanda's error (Category A5)

**Table 4.** Summary of responses to Item 3.2 where students were advised how estimation could be used to recognise that the answer was wrong

Description of each category	Frequency
A1—Explained that estimation will help one realise mean should be around 20 or 21	5
A2—Explained that estimation will help one realise that the value of 105 was too large to be the mean or will help you to realise that the mean should be lower than the highest number in the set	30
A3—Explained that estimation will help one realise that the mean lies between the largest and smallest number in a dataset	20
A4—Explained the computation of the mean	30
A5—Identified the error of not dividing the sum by the number of score	50
A6—Gave a general description of what the learner should have done which was not related specifically to the errors	44
A7—No response (blank)	4

*Conceptual understanding would have help Ayanda in realising that her answer was incorrect because understand the concept is beneficial and cost effective. Estimation would have also helped Ayanda because by estimating it means the concept is understood and estimation would have given her a closely related answer.*

**Figure 11.** Response of S75 to Item 3.2 who provided a general description (Category A6)


*She/He would have been able to know that a mean is a result of first adding the sum of observations and dividing by the number of observations and also using a plus sign (+) to clearly show he/she was adding to get 105 as an answer.*

**Figure 12.** Response of student S103 to Item 3.2 explaining the computation of the mean (A4)

*Ayanda's estimation was going to lies around 20 for example  $100 \div 5 = 20$*

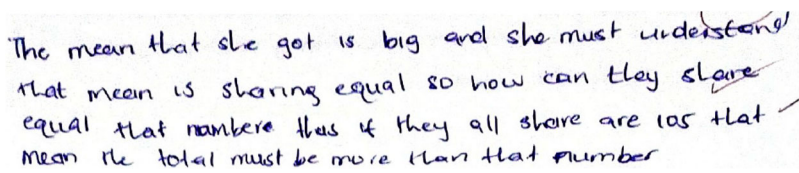
**Figure 13.** Response of student S26 to Item 3.2, explaining that by estimating one could realise that the answer lies around 20 or 21 (A1)

*The numbers in the dataset are too small to have such a big mean. She must have also checked if she had divided*

**Figure 14.** Response of student S78 to Item 3.2, explaining that by estimating one could realise that the answer given was too large (A2)

*Using estimation ... she should realise that the mean must lie between the highest and lowest score as it is a representative of all the scores*

**Figure 15.** Response of student S146 to Item 3.2, explaining that by estimating one could identify an upper and lower bound for the mean (A3)



The mean that she got is big and she must understand that mean is sharing equal so how can they share equal that number thus if they all share are 105 that mean the total must be more than that number

**Figure 16.** Use of fair share reasoning by S108 to argue that the calculated mean value was too high

Learners were asked to find the median of the following dataset: 18; 15; 23; 12; 19. Some students provided the answers given below.

Evaluate each of the following learner responses to identify what misconception, if any, the learner had. How can you address each misconception?

4.1 Median is 23 because 23 is the middle number.

4.2 The median is 17.4.

4.3 The median is 18.

**Figure 17.** Task 4

Many students did provide specific advice about how estimation could have helped Ayanda realise her error. Five of these responses (A1) referred to the mean being around 20 or 21 (Figure 13), while 30 students (A2) explained that estimation would help you realise that the mean should be lower than the highest value in the set (Figure 14). Twenty students (A3) pointed out that the mean should lie between the highest and lowest numbers of the set (Figure 15).

An interesting observation across the different categories in this group of responses about how estimation can help one narrow down the possible values, was that many students (25) used a fair share explanation. These students explained that if Ayanda knew that the mean involved putting together what is there and redistributing equally, she would have realised that out of five numbers there is no value close to 105, so there is no way you can share equally and get such a big value (Figure 16). The idea of fair sharing starts being developed at FP and it is important for PSTs to trace how this idea is taken up from an informal treatment in FP and applied to the more formal concept of the mean much later.

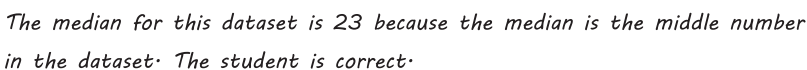
In Figure 16, the student has explained that if fair sharing is used and each share is 105, then the total of the set would be more than 105.

### **Results for Task 4 (KCS)**

Task 4 (Figure 17) targeted KCS skills.

#### *Results for Item 4.1*

For this item, students were asked to evaluate and then explain how they would respond to a learner who found the median of the given set to be 23. Ninety per cent of the participants (164) recognised that the answer of 23 was incorrect and explained the error that the set was not ordered. Ten per cent (19) of the students did not seem to recognise the error and confirmed that the method of picking out the middle number of the set was correct (Figure 18).



The median for this dataset is 23 because the median is the middle number in the dataset. The student is correct.

**Figure 18.** Response by student S149 to Item 4.1

**Table 5.** Summary of 164 responses to Item 4.1 about how students would respond to a learner who said the median was 23.

Description and frequency for each category	Example of a response in the category
B1 (45): Identified the possible cause of the error and explained what they would do to help the student address the error	<p><del>Median is incorrect</del></p> <p><del>Misconception:</del> Median is incorrect. Learner did not arrange the numbers/data in ascending order.</p> <p><u>Address misconception:</u> By teaching learners to first arrange data by ascending or descending order. I will give them activity of this. I will explain to them the definition of median.</p> <p>S12 [note that strikeout text is part of student's response]</p>
B2 (84): Identified the cause of the error and did not go further	<p>The learner misconception is that the learner did not put the data in ascending order therefore the learners order of numbers were incorrect therefore the middle numbers would be incorrect.</p> <p>S33</p>
B3 (35): Identified the cause of the error and explained the computation of the median	<p>Emphasise to the learners that before they can look for a middle number they must first arrange the numbers in an ascending order which will be 12, 13, 18, 19 ,23 leaving 18 being the median because now the numbers are in a correct order</p> <p>S132</p>

Further details about the responses of the 164 students about possible feedback they would give to address the misconceptions appear in [Table 5](#).

[Table 5](#) shows that students were easily able to identify the cause of the error. There were 84 students (B2) who just identified the error and 45 students (B1) who went on to explain in general terms what they would do to help remediate the error. A large number (35) explained how to calculate the median, once again showing a strong computational focus (B3).

#### Results for Item 4.2

For this item, students were asked to evaluate the response of a learner who said the median was 17.4; however, this value was actually the mean of the set. There were 67 participants (34%) who did not realise that the given value was the mean while 42 students wrote that the learner added two middle numbers  $23 + 12$  and divided them by 2. It seems that they incorrectly applied the method of finding the median of a set of even numbers (finding the average of the two middle numbers). Five students felt that the learner had considered numbers that were not in the dataset. There were 20 students who did not try to identify the error but simply described how a person was supposed to calculate the median of a dataset.

**Table 6.** Summary of 116 responses (64%) to Item 4.2 where the misconception was identified

Description and frequency for each category	Example of response
C1 (41): Identified the error that the learner confused mean and median and explained how they would help the learner address the misconception	<p><i>The student is confused between mean and median. The student has added all the numbers and divided by the number of values. Learners should be taught the definitions of each term, they should be given extensive exercises where they work with measure of centre through experimentation to give them a better understanding.</i></p> <p>S188</p>
C2 (44): Identified the error and explained the computation of the median	<p><i>A learner confuse the median with mean, learner does not understanding on what is a mean. I would explain to the learner that median is the middle number that you get after you arranged your data by ascending order or descending order.</i></p> <p>S33</p>
C3 (31): Identified the error and did not go further	<p><i>This student does not understand what a median is therefore the misconception here is that the student calculated the mean.</i></p> <p>S176</p>

Overall, there were 116 participants (64%) who correctly identified the underlying error that 17.4 was the mean instead of the median, and some of them provided some explanations of what could be done, which is summarised in [Table 6](#).

The summary in [Table 6](#) shows that 31 students (C3) just identified the error and did not go further, while 44 students (C2) identified the error and then explained the procedure of how to find the median, showing a computational focus. Forty-one students identified the error and explained what they would do to help the learner address the problem (C1).

#### *Results for Item 4.3*

Students were asked to respond to the answer of 18 as the median, which was the correct value. There were 157 students who explicitly identified 18 as the correct answer, while 26 (14%) did not recognise it was correct. Of this group of 26, 15 wrote that the learner had a misconception that

the first number in the list was the median. Two students believed that the learner had guessed the answer while nine participants wrote a general explanation of what could be done.

## Discussion

In this section we first discuss the extent to which the participants were able to meet specific demands embedded within the CCK, SCK and KCS subdomains. We then look across these demands to identify those that the participants seem to be most comfortable with and those that seem to present challenges to them. The findings show that certain demands were easier to meet than others and hence point to areas which teacher educators could target in developing the pedagogic content knowledge of primary school teachers in mathematics.

### ***Demands Associated with Common Content Knowledge***

Common content knowledge refers to knowing concepts in a straightforward manner as an ordinary person, without any special mathematics training, may know it. In this study, the participants found it easy to do straightforward calculations of the median (1.1) and that of the mean (1.2) with over 90% doing the correct computation. This finding is supported by Lampen (2015), who cautioned that teachers generally focus on explicit straightforward calculations of the mean instead of emphasising connections across the big ideas. The fact that the PSTs were able to compute the mean and median with ease is a positive development. This is noteworthy because there is extensive data from South Africa indicating that primary teachers are unable to perform the mathematics they are required to teach (Taylor, 2011; Venkat & Spaul, 2015). It is encouraging that PSTs can perform basic computations in these two concepts.

### ***Demands Associated with Specialised Content Knowledge***

Specialised content knowledge is referred to as the specialised mathematics knowledge needed for teaching, such as knowing different interpretations of operations and procedures. Item 1.3 probed the students about why the mean was not a good indicator of a central tendency in the given context. This question required a deep understanding of what a measure of central tendency is. Groth and Bergner (2006) argue that the ability to select the most appropriate measure of centre in a context marks the highest level of understanding of these measures. In this study almost half the class were unable to interpret the differences between the measures of central tendency or explain why the mean was not the most suitable in conveying an average value in the context. This low rate of engagement in this item is a concern since teachers are being called upon to develop statistical literacy skills with their learners. If preservice teachers are not able to evaluate the appropriateness of statistical procedures (Gal & Geiger, 2022), it suggests that they will struggle to support their learners in developing statistical literacy skills.

Another facet of specialised content knowledge is being able to solve problems involving higher cognitive skills of unpacking mathematical procedures (Ball et al., 2008), such as knowing why a procedure works and when it works. Task 2 required an understanding of the mathematics behind the calculation of the mean. There were 95 students (52%) who solved this problem while a further 10 made a slight slip. Some students were able to generate 17 numbers whose mean was 65, then considered the two additional numbers and found the new mean of the 19 numbers. These students chose to use a tedious calculation but they showed an understanding of the procedures used in calculating the mean. The rest of the group (37%) did not respond or they used number grabbing (Schoenfield, 1988; Mbonambi & Bansilal, 2014), which is usually used when students have no other option that makes sense. Solving such problems with a high level of cognitive demand is important for developing statistical literacy skills which teachers are required to foster in their learners.

### ***Demands Associated with Knowledge of Content and Students***

The subdomain knowledge of content and students includes knowing about students' misconceptions. Although recognising that an answer is wrong can be seen as a CCK competence, familiarity

**Table 7.** Summary of responses to some demands

	Correct calculation	Identified error	Identified error only	Gave a general or specific explanation	Repeated the explanation of the computation
Mean	180 (98%)	171 (93%)	50	99	30
Median	176 (96%)	164 (90%)	84	45	30

with common errors and understanding the nature of the error fall within KCS. A further skill required by teachers is that of providing feedback to learners that could address the misconception and help them improve their understanding (Brodie, 2014). In this study, we looked at some of these demands within the KCS domain as revealed in the responses to Tasks 3 and 4.

In terms of KCS skills with respect to the mean (Item 3.1), 171 students (93%) were able to identify Ayanda's misconception; however, only 55 students (30%) provided feedback about different ways in which estimation could have alerted the learner that the answer of 104 was incorrect. In Table 4, 50 students stopped short of identifying the error, while 44 provided general statements and four did not give a response at all. A further 30 students simply explained the computation of the mean.

In terms of diagnosing errors related to the median and providing feedback, in Item 4.1, 90% of the participants correctly identified the cause of the error, while 89% identified that the response was incorrect in Item 4.2. However diagnosing the error was not as easy in Item 4.2 since 67 participants (37%) provided an incorrect diagnosis of the error, while 41 (23%) identified the error and explained the feedback they would give to the learner.

These results reveal differences between the students' proficiency in carrying out a straightforward calculation, identifying the cause of errors and providing feedback on the learner errors. This study indicates that when dealing with the concepts of the mean and median, the easiest task for the PSTs was to directly calculate these without thinking about the links between the conceptual and procedural foundation. The next level of complexity involves identifying an error of misconception in someone else's work. Giving relevant feedback about learner errors was the most complex task for these PSTs. They are unable to link the computations to the conceptual and procedural foundations of data interpretation in the Foundation Phase by thinking about central tendency measures and the procedures needed to do this. These findings are illustrated in Table 7. More than 95% of students were able to calculate the mean and median respectively. When presented with an error in the calculation of the mean or median, over 90% were able to recognise the error. There were 129 (70%) who went on to provide some feedback related to the error in the mean value, while 75 (41%) provided some sort of feedback about the error in the calculation of the median.

In this study, the PSTs found it much harder to provide feedback or to respond to learners about their errors, with many students stopping short after identifying the errors. Many of the PSTs seem to believe that in order to help their learners recognise their mistakes, they simply need to explain the procedure again. This approach also seems to be favoured by experienced teachers who reported that they prefer teacher explanations to any other approaches (Umugiraneza, et al., 2017). An important role of the teacher concerns the feedback given to learners about their errors or misconceptions and using the errors as an opportunity to improve their understanding of the concepts at hand (Mutambara & Bansilal, 2022). These findings suggest that primary school PSTs need opportunities to engage with the ideas about feedback.

## Conclusion

The role of statistical literacy in meeting the demands of the twenty-first century has been recognised globally. In this study we analysed the responses of a group of 183 primary school PSTs to items on measures of central tendency, a central concept in statistical literacy. The purpose was to try to understand the extent to which these participants could problem solve and engage at higher levels of cognitive reasoning as well as their awareness of learner errors and the related feedback.

An important contribution of this study is that the results can be used to distinguish between and quantify students' readiness to meet different demands of teaching the mean and median as measures of central tendency in this context. The results showed that more than 95% of the PSTs were able to carry out simple computations of the mean and median. More than 90% identified the cause of the error or underlying misconception in a learners' incorrect response about a mean or median. Seventy per cent were able to give feedback about the error in the mean while 41% did the same for the median. Effective teaching requires useful feedback to learners when they commit errors or display misconceptions, which must include remediation plans on how these can be addressed. This study shows that preservice teachers need more support in developing these skills with respect to the concepts of the mean and median.

In terms of the higher-level cognitive demands related to the SCK subdomain for the mean, 61% were able to solve the problem requiring them to calculate a new mean for a dataset that was extended by two numbers. In terms of reasoning about why the mean was not a good indicator of the average in a particular context, only 50% of the class offered relevant explanations. Solving such problems with a high level of cognitive demand is important experience that can develop key statistical literacy skills that teachers are required to foster in their learners. Granting PSTs these additional opportunities to engage with and effectively solve such problems will enhance their readiness to support their learners in developing statistical literacy skills.

### Acknowledgements

Funding for this study was given by the National Research Foundation (South Africa), grant no. 129308.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### ORCID

Sarah Bansilal  <http://orcid.org/0000-0002-5445-5612>

### References

- 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.
- Bansilal, S. (2023). Statistics and probability in the curriculum in South Africa. In G. F. Burrill, L. de Oliveria Souza, & E. Reston (Eds.), *Research on reasoning with data and statistical thinking: International perspectives*, Advances in Mathematics Education. Springer Nature.
- Brodie, K. (2014). Learning about learner errors in professional learning communities. *Educational Studies in Mathematics*, 85, 221–239. <https://doi.org/10.1007/s10649-013-9507-1>
- Burrill, G. F., de Oliveria Souza, L., & Reston, E. (Eds.). (2023). *Research on reasoning with data and statistical thinking: International perspectives*. Springer Nature.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*, 23, 45–55.
- Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: Mathematics, statistics, evidence literacy, and the Coronavirus pandemic media. *Educational Studies in Mathematics*. <https://doi.org/10.1007/s10649-022-10151-7>
- Garfield, J. & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning Learning Environment. *Teaching Statistics: An International Journal for Statistics and Data Science Teaching*, 31(3), 72–77. <https://doi.org/10.1111/j.1467-9639-2009.00363.x>
- Groth, R. E. & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median and mode. *Mathematical Thinking and Learning*, 8(1), 37–63.

- Harrell-Williams, L. M., Lovett, J. N., Lee, H. S., Pierce, R. L., Lesser, L. M., & Sorto, M. A. (2019). Validation of scores from the high school version of the self-efficacy to teach statistics instrument using preservice mathematics teachers. *Journal of Psychoeducational Assessment, 37*(2), 194–208.
- Holmes, A., Illowsky, B., & Dean, S. (2018) *Introductory business statistics*. OpensStax.
- Kalobo, L. (2016). Teachers' perceptions of learners' proficiency in statistical literacy, reasoning and thinking. *African Journal of Research in Mathematics, Science and Technology Education, 20*(3), 225–233.
- Kristanto, Y.D. (2018). Mathematics pre-service teachers' statistical reasoning about meaning. *IOP Conference Series: Materials Science and Engineering, 296*, <https://doi.org/10.1088/1757-899x/296/1/012037>
- Lampen, E. (2015). Teacher narratives in making sense of the statistical mean algorithm. *Pythagoras, 36*(1). <https://doi.org/10.4102/pythagoras.v36i1.281>
- Leavy, A. M. and O'Loughlin, N. (2006). Moving beyond the arithmetic average: Pre-service teachers' understanding of the mean. *Statistics Education Research Journal, 9*(1), 53–90.
- Mbonambi, S., & Bansilal, S. (2014). Comparing Grade 11 mathematics and mathematical literacy learners' algebraic proficiency in temperature conversion problems. *African Journal for Research in Science, Mathematics and Technology Education, 18*(2), 198–209.
- Mutambara, L. H. N., & Bansilal, S. (2022). A case study of in-service teachers' errors and misconceptions in linear combinations, *International Journal of Mathematical Education in Science and Technology, 53*(11), 2900–2918. <https://doi.org/10.1080/0020739X.2021.1913656>
- Schoenfeld, A. H. (1988). Problem solving in context(s). *The Teaching and Assessing of Mathematical Problem Solving, 3*, 82–92.
- Sunde, P. B., Peterson, J., Nosrati, M., Rosenqvist, E., & Andrews, P.(2022). Estimation in the mathematics curricula of Denmark, Norway, and Sweden: Inadequate conceptualisations of an essential competence, *Scandinavian Journal of Educational Research, 66*(4), 626–641.
- Taylor, N. (2011). *The national school effectiveness study (NSES): Summary for the synthesis report*. JET Education Services.
- Umugiraneza, O., Bansilal, S., & North, D. (2017). Exploring teachers' practices in teaching Mathematics and Statistics in KwaZulu–Natal schools, *South African Journal of Education, 37*(2), article 1306, 13 pages. <https://doi.org/10.15700/saje.v37n2a1306>
- Umugiraneza, O., Bansilal, S., & North, D. (2022). An analysis of teachers' confidence in teaching mathematics and statistics. *Statistics Education Research Journal, 21*(3). <https://doi.org/10.52041/serj.v21i3.422>
- Venkat, H., & Spaull, N. (2015). What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *International Journal of Educational Development, 41*, 121–130. <https://doi.org/10.1016/j.ijedudev.2015.02.002>
- Wessels, H., & Nieuwoudt, H. (2011). Teachers' professional development needs in data handling and probability. *Pythagoras, 32*(1). <https://doi.org/10.4102/pythagoras.v32i1.10>