Spring 2023

K-8 Preservice Teachers' Statistical Thinking When Determining Best Measure of Center

Ha Nguyen
California State University

Eryn M. Stehr Maher
Georgia Southern University, estehr@georgiasouthern.edu

Gregory Chamblee
Georgia Southern University, gchamblee@georgiasouthern.edu

Sharon Taylor
Georgia Southern University, taylors@georgiasouthern.edu

www.ijemst.net

K-8 Preservice Teachers' Statistical Thinking When Determining Best Measure of Center

Ha Nguyen
California State University Dominguez Hills, United States

Eryn M. Maher
Georgia Southern University, United States

Gregory Chamblee
Georgia Southern University, United States

Sharon Taylor
Georgia Southern University, United States

To cite this article:

Nguyen, H., Maher, E. M., Chamblee, G., & Taylor, S. (2023). K-8 preservice teachers' statistical thinking when determining best measure of center. *International Journal of Education in Mathematics, Science, and Technology (IJEMST), 11*(2), 440-454. https://doi.org/10.46328/ijemst.2365

International Journal of Education in Mathematics, Science and Technology

2023, Vol. 11, No. 2, 440-454 https://doi.org/10.46328/ijemst.2365

K-8 Preservice Teachers' Statistical Thinking When Determining Best Measure of Center

Ha Nguyen, Eryn M. Maher, Gregory Chamblee, Sharon Taylor

Article Info

Article History

Received:
08 February 2022
Accepted:
09 November 2022

Keywords

Teacher education
Statistics
Mean
Median

Abstract

The purpose of this study was to determine K-8 preservice teacher (PST) candidates' statistical thinking when selecting the best center representation for the given data. Forty-four PSTs enrolled in a Statistics and Probability for K-8 Teachers course in a university located in the southeastern region of the United States were asked to complete a 2007 National Assessment of Educational Progress test item. All 44 PSTs' data were qualitatively analyzed for correctness and statistical thinking strategies used. Findings were that most PSTs either incorrectly selected the mean, rather than median, as the best measure of center for the given data or did not use appropriate statistical reasoning when explaining their answers. Future research includes modifying the explanation component so PSTs must better explain their statistical thinking for their choice of best measure of center using the context of the problem. Future research could also include implementing a pre- and post-test design with the post-test item embedded in the final exam. This design will provide additional understanding of how much knowledge PSTs bring to the course versus how much they learn in the course and provide incentive for giving thoughtful consideration for their answers.

Introduction

Statistical concepts have been part of the K-12 curriculum for many years. With publications such as the National Council of Teachers of Mathematics (NCTM) *An Agenda for Action* (NCTM, 1980) and *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), statistics education started coming into the mainstream mathematics curriculum. If students needed to learn more statistics, then in-service and pre-service teachers needed to also. The American Statistical Association (ASA) and NCTM advocate for high quality statistics education for in-service and pre-service teachers (PSTs) in order to develop statistical reasoning skills in K-12 students (ASA & NCTM, 2013). Other organizations such as the Association of Mathematics Teacher Educators (AMTE, 2017) and Conference Board of Mathematical Sciences (CBMS, 2012) echo the recommendations for K-12 teachers' statistical understanding. Most studies of PST's statistical knowledge focus on understanding mean or median. Estrada et al. (2004) conducted research on 367 PSTs and found many made errors regarding the mean despite having previous training in statistics. Leavy and O'Loughlin (2006) found that roughly 25% of PSTs had some form of conceptual understanding but the remainder showed limited procedural knowledge. According to Jacobbe and Carvalho (2011), an over-reliance on computation with little focus on conceptual understanding has created barriers to statistical reasoning.

To investigate the statistical understanding of PSTs, the study asked one question from the National Assessment of Educational Progress (NAEP) test from 2007. PSTs were given a data set, the mean, and median and asked to interpret the mean and median and determine the best choice to represent the data. The discussion here is limited to the choice of the best representation of the data. The initial review of PST responses consisted of using a straightforward rubric of correct, partially correct, or incorrect. While using this system gave a count of PSTs in each category, it did not provide an understanding of PSTs' thinking process in making their choice of best measure of center. To better understand the choice process, a more intense review of PST responses began. PST responses were sorted, and similar responses grouped together. Further sorting within each main category allowed researchers to pinpoint specific conceptual errors and where PSTs encountered difficulties. This deeper analysis highlighted flaws in PSTs' understandings that were not possible simply looking at correct, partially correct, and incorrect responses. By examining PST responses as well as an in-depth analysis of their reasoning, a clear picture of PSTs' misunderstandings comes into focus. The discussion of these misunderstandings and implications for curricular changes at the PST level follows. The research question was:

> *What statistical thinking strategies do PSTs use to justify their answers when determining the best measure of center at the end of a statistics and probability course?*

Background Literature

NCTM (1989 and 2000) along with the *Common Core State Standards for Mathematics* (National Governors Association Center for Best Practice & Council of Chief State School Officers, 2010) emphasize the need for K-8 students to master statistical content and thinking. Each of these documents notes having an in-depth knowledge of measures of center and their applications are essential. In order for students to master these ideas, prospective K-8 teachers must master these ideas (CBMS, 2012; ASA & NCTM, 2013; AMTE, 2017).

"Statistical literacy is no longer a skill relegated to the few. It is essential knowledge required by all that must be developed beginning at an early age, continuing throughout one's school years" (Metz, 2010, p. 19). Chance (2002) in a historical review of what constitutes statistical thinking/reasoning noted there was no specific definition. However, statistical thinking/reasoning, in general, does involve questioning, justification, and writing in your own words and can only be accomplished via problems that test student reflexes, thought patterns, and creativity in novel situations (Chance, 2002, p. 12). ASA (2016) guidelines build on this idea. More specifically, to develop statistical thinking skills it is crucial to focus on helping students become better educated consumers of statistical information. Students not only need content knowledge, but they also need instruction that emphasizes the use and interpretation of statistics in everyday life.

The National Assessment of Education Progress (NAEP) has been assessing K-12 students' measures of center knowledge since 1990. In 1990, the question in Figure 1 was asked of 12th graders (NAEP, 1990-8M7 #5). NAEP classified item difficulty as Easy; however, only 69% of students answered the question correctly.

The average weight of 50 prize-winning tomatoes is 2.36 pounds. What is the combined weight, in pounds, of these 50 tomatoes?

A. 0.0472

B. 11.8

C. 52.36

D. 59

E. 118 (correct answer)

Figure 1. 1990 NAEP Measure of Center Question

NAEP in 2007 asked eighth graders a statistical thinking question (see Figure 2) that required them to explain which statistic is best in a given situation (NAEP, 2007-8M9 #8). NAEP classified item difficulty as Hard. NAEP stated the correct answer was: Median. A correct explanation of reasoning was noting "the mean is lowered by one low number, 10". A partial answer was: Median with incomplete, incorrect, or missing explanation OR Mean with median explanation. An incorrect answer was: incorrect responses. Six percent answered the question correctly, 21 percent answered the question partially correctly and 67 percent answered the question incorrectly.

The table below shows the number of customers at Malcolm's Bike Shop for 5 days, as well as the mean (average) and the median number of customers for these 5 days.

Number of Customers at Malcom's Bike Shop	
Day 1	100
Day 2	87
Day 3	90
Day 4	10
Day 5	91
Mean (average)	75.6
Median	90

Which statistic, the mean or the median, best represents the typical number of customers at Malcolm's Bike Shop for these 5 days?

Explain your reasoning.

Did you use the calculator on this question? Yes No

Figure 2. 2007 NAEP Measure of Center Question

Teaching Mean and Median in PSTs Textbooks

It is common for K-8 PSTs to take a "Mathematics for Elementary Teachers" course as part of their program of study. Most of these courses use books specifically designed for these students. To see how mean and median were covered in elementary PST textbooks, we examined six texts commonly used in pre-service content courses. In examining these texts for their treatment of mean, median, and which one is most appropriate there were some common themes. In defining the mean, the majority of texts use the formula (O'Daffer et al., 2007; Bennett et al., 2016; Billstein et al., 2020; Musser et al., 2020). After the mathematical definition, several texts also demonstrate balancing as a way to provide context for the mean (Beckmann, 2014; Bennett et al., 2016; Long et al., 2018; Billstein et al., 2020).

Texts for K-8 PSTs have a variety of ways to define median. The simplest definition appears in the Billstein et al. (2020) text and states the median is "the value exactly in the middle of an ordered set of numbers." Some texts discuss the placement of the median based on whether there are an odd or even number of data points (O'Daffer et al., 2007; Long et al., 2014; Bennett et al., 2016). Musser et al. (2020) also bases the definition of median based on an odd or even number of data points but describes the position of the median as a formula. After presenting definitions of mean and median, each text presents an example with a data set with a clear outlier to demonstrate why the median is sometimes more appropriate than the mean as a measure of central tendency. However, after this one example, the concept of an appropriate choice is not revisited. In these texts, emphasis is placed on computational fluency in finding mean and median. There are no opportunities for students to work with mean and median in context. With limited support from texts, it is incumbent on faculty to provide examples not only of which measure is most appropriate, but also how to interpret each measure in the specific context of a problem.

Mean and Median Research

Mean and median research has assessed K-12 students', practicing K-12 teachers', and pre-service teacher candidates' understanding. The research has consistently found all 3 groups have misunderstandings of these concepts with an emphasis on mean understanding. Cai (2000) explored United States (US) and Chinese sixth grade students' understanding and representation of the averaging algorithm using two contextualized problems.

The US and Chinese students were found to use similar strategies (average formula, leveling, and guess-and-check) with those US and Chinese who used an appropriate solution strategy; the majority of them used the averaging algorithm (Cai, p. 852). Tenth graders in Malaysia in nine rural schools were studied to determine students' measures of central tendency knowledge (mean, mode, median). Students were found to have "a high level of understanding regarding the definition; a moderate level of understanding regarding the procedure; a low level of understanding on properties; and a very low understanding on the problem and representation, as well as argument and proof" (Saidi & Siew, 2019, p. 78). Whitaker et al. (2015) tested middle grades and secondary students' statistical knowledge. Questions assessing data analysis were found to be the most difficult for students to answer.

Estrada et al. (2004) surveyed PSTs to determine their level of statistical knowledge prior to the teaching of a statistics unit. Estrada et al. found only 46.9% of PSTs correctly answered a mean question when it included outliers. Groth (2009) conducted an online asynchronous case study to document nine practicing elementary and middle grades teachers' conversations about teaching mean, median, and mode. Teachers did not oftentimes require students to reason about contextualized data (Groth, 2009, p. 715). A review of research on teachers' understanding of mean found that teacher and student understanding of average was very similar along with an exaggerated reliance upon procedural algorithms and a general lack of conceptual understanding by both students and teachers" (Jacobbe & Carvalho, 2011, p. 207).

Groth and Bergner (2006) conducted a qualitative study of preservice elementary and middle grades teachers' content knowledge about mean, median, and mode. Most teachers were found to equate the mean with the average of a data set and the median described how one calculates the measure only in cases having an odd number of values in a data set" (Groth & Bergner, p. 30). Leavy and O'Loughlin (2006) researched elementary teachers' understanding of mean. Mastery of computational skills relating to the mean were evident, but all had difficulties in applying their knowledge of the mean to unfamiliar tasks (Leavy & O'Loughlin, p. 84). Reaburn (2013) asked 32 first- and second-year PSTs in Australia to determine the best measure between mean and median in a contextualized problem. Half of the students did not provide an answer. Of those responding, half chose mean and half chose median. Mean reasoning included "takes all values into account", "more correct", and "because it is the average". Median reasoning included "median is more representative", median – no further answer', and 'median because of soy sauce [outlier]" (Reaburn, p. 566).

Method

Participants, Context, and Procedures

Participants in the study were 44 senior- and junior-level PSTs, preparing to teach ages 5-14, enrolled in two senior-level *Probability and Statistics for K-8 Teachers* courses, taught by the same instructor, at a university in the southeastern United States. PSTs included 32 Elementary (ages 5-11) PSTs, three Middle Grades (ages 11-14) PSTs, three Special Education PSTs, and six Dual (Elementary / Special Education) PSTs. Prior to enrollment, all PSTs completed at least: (a) *College Algebra*, *Mathematics Modeling*, or higher-level mathematics; (b) *Numbers and Operations*; and (c) *Data Analysis and Geometry*. Measures of center are introduced in (c), emphasizing computing and defining mean, median, and mode. The *Probability and Statistics* course requires PSTs to apply and interpret the measures across a variety of situations and contexts. Table 1 displays the course overview.

Table 1. Probability and Statistics for K-8 Teachers Course Overview

Unit	Topics Taught
1 (4.5 Weeks)	Sampling Design: Types of sampling designs, sampling errors, non-sampling errors, margin of errors, confidence statements Study Design: Surveys; Experimental vs. observational studies
2 (5 Weeks)	Analysis: Representations, Measures of center and variation Two-Variable Data: Scatterplots, Correlation, Regression
3 (4.5 Weeks)	Probability: Counting methods and use for probability and multi-stage experiments, Two-way tables, Odds, Expected value

The PSTs in this course had taken a *Foundations of Data Analysis and Geometry* course (see (c) above). There are several instructors in our department who teach this course, so we cannot describe each PST's experience with mean and median. However, the course used Beckmann (2014) textbook and tasks from that book, which includes a treatment of conceptual meanings of mean: balancing, leveling out, etc. In the final meeting of *Probability and Statistics for K-8 Teachers* (taught by the first author), before the final exam, PSTs completed the item (see Figure 2) individually, using any type of calculator. No incentives were given, and the item was ungraded.

Analysis

Data were analyzed using the constant-comparative analysis technique to determine the best measure of center scoring results. Each author individually classified PSTs' answers into two categories: response matched the NAEP key (correct) or did not match the NAEP key (not correct). Each author then classified all not correct responses using the NAEP key as a guide as either partially correct or incorrect. The authors met to discuss their findings. Inconsistent classifications were discussed as a team. PST responses that did not receive a consensus were re-classified individually and an additional meeting was held to discuss re-classifications. This consensus process was continued until all PST responses received a consensus vote.

After the initial analysis, the same overarching technique was used to categorize statistical thinking strategies used. Items classified as correct and incorrect were grouped together. Each author individually categorized PSTs' explanations into themes. The authors met to discuss their findings. Inconsistent themes were discussed as a team. Themes that did not receive a consensus were re-analyzed individually and an additional meeting was held to discuss theme changes. This consensus process was continued until all themes received a consensus vote. A final meeting was held to discuss consolidating themes.

Results

Two analyses were conducted to answer the research question: scoring the best measure of center responses and coding statistical thinking strategies used by PSTs. Results from these analyses are presented below.

Best Measure of Center Scoring Results

PSTs were asked to choose the best measure for the situation and explain their reasoning. Hence correct answers include two parts: (a) "median" and (b) an explanation that indicates statistical reasoning that the best choice for measure of center is the more fair representative of the data and that, in this situation, the low value signifies that the mean will be less fair than the median. For example, an expected correct answer is, "The median is a better measure of center. Day 4's customers are much fewer than those on other days, so the mean is smaller than it should be to fairly represent a typical day at the bike shop." Of the 44 PSTs, 21 gave correct responses, while nine gave partially correct responses, and 14 gave incorrect responses (see Table 2).

Examples of correct and partially correct responses from the data are also shown in Table 2.

Table 2. Results for Choice of Statistic and Reasoning

Scoring	Common Responses	Frequency	%
Correct	median as best measure because of outlier "Median because there is an extreme low value of 10 which pulls the mean down"	21	48%
Partially Correct	median as best measure; flawed reasoning "median because it describes the majority"	9	20%
Incorrect	Mean as best measure and/or explanation indicated misconception(s)	14	32%

Incorrect responses included PSTs choosing the mean as the best measure (rather than median) and/or showing misconceptions. For example, one PST wrote, "The mean because it most directly reflects the middle of the data." This PST mistook the meaning of median for the mean (i.e., "middle") along with the phrase "most directly reflects" which may indicate *most efficient* rather than *fairest* representative of the data.

Statistical Thinking Strategies Used

In response to the question asking PSTs to decide on the statistic that best represents the typical number of customers at Malcolm's Bike Shop and to share their reasoning, 30 PSTs answered "median" and 14 answered "mean" (see Figure 3).

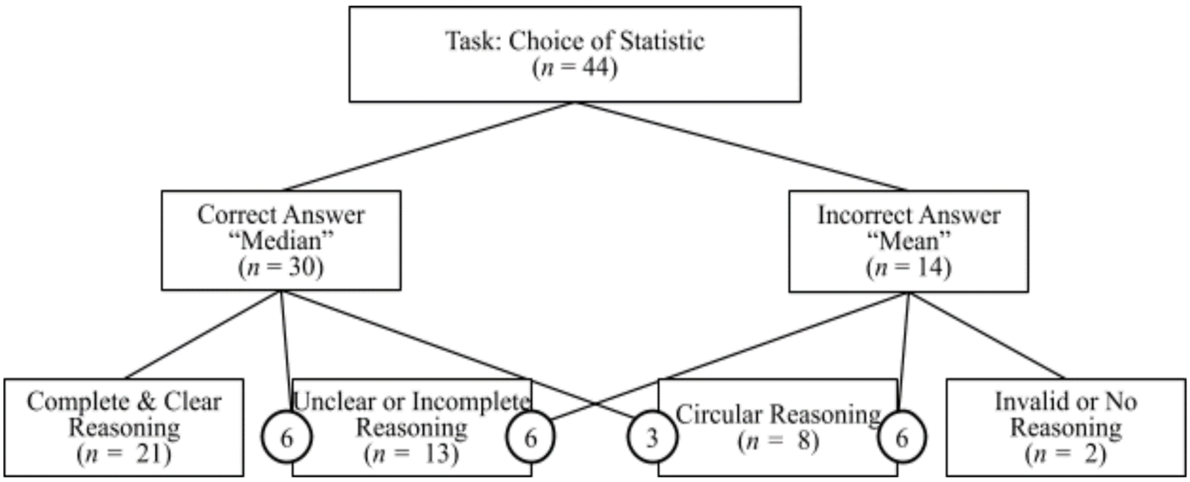

Figure 3. Overview of PSTs' Answers and Reasoning

All but one PST provided reasoning in support of their response. To identify statistical thinking strategies used by PSTs to respond to this question, we organize responses in two ways. We first describe general strategies based on PSTs' choice of best measure (i.e., *mean* or *median*). We then describe PSTs' reasoning in more detail, categorizing aspects of statistical reasoning that were visible in any PST explanations: (a) attention to context, (b) meaning of central measures, (c) attention to the low value, and (d) attention to the low value's impact on the mean.

Median and Statistical Thinking

Median responses were sorted into those that exhibited *complete and clear* reasoning (n = 21), *circular* reasoning (n = 3), or *unclear or incomplete* reasoning (n = 6) (see Figure 3). In our analysis, we defined *complete and clear* reasoning as an explanation that supported their answer and that made sense to us in the way it was written. For example, a PST wrote, "90 or median [because] 10 is an outlier that brings down the mean." We categorized this response as complete and clear because the PST gave a reason (i.e., "outlier that brings down the mean") that justified their response and indicated practical statistical thinking. Three responses were categorized as circular reasoning. These responses restated either the meaning of median (e.g., "because it is the middle") or a part of the question (e.g., "shows around what # is typically there") without providing evidence of additional statistical thinking.

If the reasoning did not make sense to us in some way or if it seemed to lack information, we marked it as *incomplete or unclear* reasoning. For example, one PST wrote, "Median because as the mean goes down the median will stay the same." This reasoning was categorized as unclear or incomplete despite the indication of some understanding (e.g., that the mean is affected while the median is not) because the PST did not state why the mean might go down. An example of an unclear response is, "The median because it describes the majority." In this analysis, we cannot differentiate between a PST using valid, but poorly communicated, reasoning and one using invalid reasoning. Of the 30 PSTs that chose median, 21 communicated complete and clear reasoning while nine did not.

Responses that we identified as *complete and clear* varied in the amount of detail and sophistication of reasoning that was included. For example, compare two PSTs' responses: "median because there is an extreme value" and "Median. The extreme low of 10 caused the mean to be much lower than it should [be]. The mean is not a good representation [because] of the extreme low." Note that both responses mention an extreme value, but the second response further explains the impact of the extreme value on the mean and affirms that the impact causes the mean to be not as good of a representation of the center of the data.

Mean and Statistical Thinking

Considering responses of PSTs who answered "mean," we found PSTs used *circular* reasoning (n = 6), provided an *unclear or incomplete* explanation (n = 6), an *invalid* explanation (n = 1), or *no* explanation (n = 1) (see Figure 3). We defined *circular* reasoning as any explanation that the mean is the best measure of center because it "is the average" (e.g., providing a synonym) or "is the typical number" (e.g., restating the question). For example, two PSTs wrote, "mean – it's the average" and "Mean. It's the typical/average number." PSTs using the former justification focused on the idea that *mean* and *average* are synonymous, and so mean must be the best measure of center, without necessarily considering the data values.

Other PSTs provided explanations that we categorized as *unclear or incomplete*. For example, the mean "shows how many people came in," "takes all numbers into account," or "is a bigger range." These written responses could reflect valid reasoning, but do not provide enough detail to clearly communicate it. Finally, we classified one explanation as *invalid*, because the PST used the meaning of median, writing "because it most directly reflects the middle of the data." The PST might have valid reasoning but wrote "middle of the data," so her response indicates a misconception about the meaning of *mean* as opposed to *median*. Figure 4 shows that, regardless of whether or not they gave a correct answer or clear and complete reasoning, PSTs differed in their attention to the meaning of measure of center, their attention to the low value, and their explanation of the impact of the low value on the mean (and not on median).

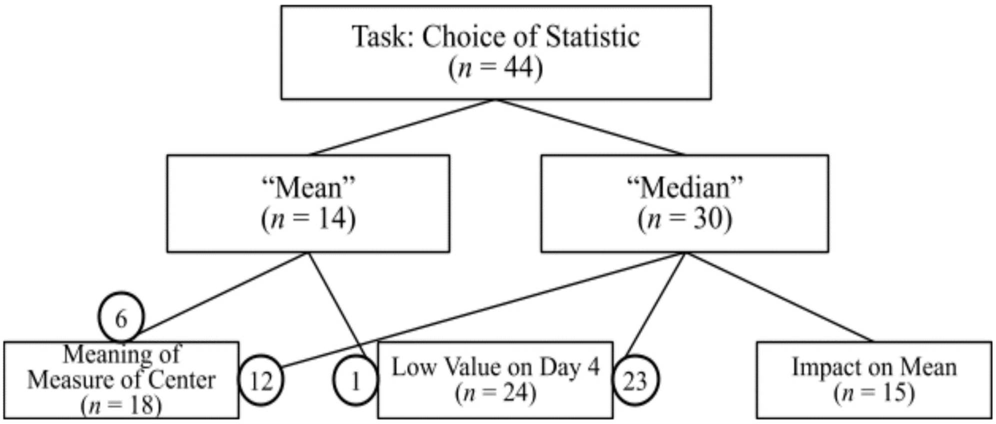

Figure 4. Responses Sorted by Measure of Center, Low Value, and Impact

In the following sections, we consider all PSTs' responses to explore these aspects of their statistical thinking. Particularly, we explain and share student examples of the categories shown in Figure 4. We begin by exploring students' use of the meaning of mean or median in their responses. We then give examples of students' attention to the low value on Day 4, when the shop had 10 customers rather than the 100, 87, 90, or 91 customers on the other days. Finally, we explore students' responses when they reference the low value's impact on the mean and why they led them to believe that median was the best choice.

Meaning of Measuring Central Tendency

Mean and median are measures of center, intended to provide a fair representation of what is a *typical* or *fair* representation of all data values. Because the question includes this wording, it is difficult to parse the difference between circular reasoning (e.g., restating the question) and understanding the measure of center. In this section, we consider only the wording used by PSTs. In their written justifications, many PSTs (n = 12) clearly referenced the idea of "best represents the typical number of customers" using phrases such as: accurate representation (n = 2), typical number of customers (n = 5), good or reliable description of the average number of customers (n = 5). For example, one PST wrote, "The median because on day 4 the store had a very low number of customers that is not a reliable description of the average number of customers at the store."

The inclusion of the idea of "reliable description of the average" shows some awareness of the meaning of measure of center. Other PSTs (*n* = 6) less clearly referenced the meaning of measure of center, but still seemed to reflect the need for a measure of center to describe all data fairly, using phrases such as: describes the majority or all of the data (*n* = 5) or directly reflects the middle (*n* = 1). For example, one PST wrote, “there are more numbers in the 90’s – 100’s than 75.6.” This explanation is incomplete and unclear, but seems to reference the idea of fairness, that a measure of center should represent “more numbers.”

Attention to Low Value

In statistics, *extreme value* and *outlier* can have different meanings, but formally *outlier* refers to a datum outside the inner fences and *extreme value* outside the outer. In the task given, the outlier calculation shows that 10 is not an outlier because it is not less than the lower fence: Q1 - 1.5* IQR = 48.5 - 1.5*47 = -22, where Q1 is the lower quartile, and IQR is the interquartile range, where Q1 is the lower quartile, and IQR is the interquartile range. Because it is not an outlier, 10 is also not an extreme value: Q1 - 3*IQR = 48.5 - 3*47 = -92.5.

Using these formal meanings, Day 4’s data point of 10 customers is neither an outlier nor an extreme value. Because data are assumed to be clustered around a central value, however, statisticians and laypeople can informally use the terms *extreme values* or *outliers* to refer to data values that seem unusually far from the main cluster. Most PSTs (*n* = 25) used one or more of the following terms as a part of their reasoning: extreme value (*n* = 11), outlier (*n* = 10), very low value (*n* = 2), or Day 4’s 10 customers (*n* = 17). Even though the low value is not an outlier or an extreme value, the question is structured to draw attention to it and to encourage PSTs to choose median as a measure of center because of its impact on the mean.

Impact of Low Value on Mean and Median

Many PSTs (*n* = 15) included the unusually low value’s impact on the mean in their reasoning. Of these, most (*n* = 9) also mentioned the direction of the impact (i.e., “drags down” or “lowers”), but several did not. Two PSTs (one PST from each group) described the impact as skewing the mean or results. Other PSTs used a variety of language and metaphors that we discuss here.

PSTs who did not indicate a direction of impact, used “affected by” (*n* = 2), “throws off” (*n* = 1), or “causes inaccuracy” (*n* = 1) to explain the impact of the low value on the mean. PSTs who included direction used words such as “brings” or “pulls” the mean down (*n* = 5), “caused the mean to be lower” (*n* = 3), or “mean goes down” (*n* = 1). Although one PST used the description of median as a “resistant summary measure” to justify their response, and it does not explicitly reference an impact on the mean, we felt it was appropriate to mention here as acknowledgement that the mean is a sensitive summary measure while median is resistant.

Discussion

Almost half of the PSTs (48%) provided correct responses (“median” and complete explanation) while about a third (32%) gave incorrect responses (“mean” and/or explanation indicating misconceptions). Compared to PSTs in Estrada et al.'s (2004) research, a lower percentage of PSTs in this study chose the mean as the best center measure (32% versus 46.9%, respectively). Compared to eighth graders who were given the same task, PSTs in this study scored higher (48% versus 6% correct responses and 32% versus 67% incorrect responses, respectively) (NCES, 2007). PSTs had difficulty with the data analysis and conceptual understanding aspect of the question. This finding extends Whitaker et al.'s (2015) and Jacobbe and Carvalho's (2011) studies to PSTs. We agree, a focus on correctness without also attending to reasoning does not demonstrate mastery of the concept.

Several PSTs used circular logic to explain their reasoning for choosing the median (n = 2) or the mean (n = 5). Examples of those include, "The median because it is the middle of all the customers," and "mean - it's the average." Using "middle" to explain the median and "average" to explain the mean do not provide insights into their reasoning for choosing the center measure. Several PSTs explained the mean was better because it "shows the range." Researchers assumed that by "range," PSTs meant the distance between extreme values, but such an interpretation would not align properly with the meaning of mean (or center). Our PSTs' statistical reasoning echoes Reaburn's (2013) students' responses; although, Reaburn's students were first- and second-year PSTs while ours are in their third or fourth year of the program. However, our PSTs' level of reasoning also resonates with responses from practicing teachers (e.g., Engledowl & Tarr, 2020; Groth & Bergner, 2006).

While the scoring results gave a clear picture of correct, partially correct, and incorrect responses, the mapping revealed that regardless of their choice for best statistic, the PSTs in this study displayed (mis)conceptions that a correct, partially correct, and incorrect response model would otherwise miss. Specifically, about 41% of the PSTs gave responses that have potentially valid measure of center meanings by including in their reasoning phrases such as: accurate representation, typical number of customers, good representation of the data points, or a reliable description of the average number of customers. A response to this NAEP 2007 question can be superficial (e.g., "median because of the outlier") but PSTs showed glimpses of other understandings as they attempted to explain using statistical thinking what they knew based on memorized procedures.

Approximately 57% of the PSTs noted an extreme value, an outlier, a very low value, and/or Day 4's 10 customers in their explanations. The term "outlier" was mentioned 10 times while "extreme value" was mentioned 11 times and "Day 4 or 10 customers" 17 times. This reasoning indicates that many PSTs acknowledged 10 customers on Day 4 impacted their choice. However, the uncertainty about whether 10 should be classified as an outlier or extreme value shows that the PSTs had not yet totally conceptualized how these ideas are similar and different. The following PST's response, "Median as 10 was an extreme low value, maybe an outlier," illustrates PSTs' awareness of the difference between "outliers" and "extreme values."

Overall, based on the PSTs' responses in this study, the researchers found in terms of statistical thinking a correct answer with complete, clear, and sophisticated reasoning consists of all the following components: 1) correct answer, 2) reasoning, and 3) supporting concepts. The correct answer is "median." Clear and complete reasoning uses valid mathematical language and provides concepts to support the PSTs' choices. Supporting concepts convey the meaning of center measures, meaning of outlier as compared to extreme value and unexpectedly low value, and impact of data values that are different than the "typical" value (e.g., median as a resistant summary measure in contrast to mean as sensitive to values that are much higher or lower than the typical values).

Conclusions, Limitations, and Future Research

In this section, we first present conclusions of the study. We connect our findings back to NCES (2007). We then discuss limitations and suggest future research directions.

Conclusions

PSTs continue to have difficulty determining the best measure of center when presented with a statistical thinking problem (NCES, 2007; Jacobbe & Carvalho, 2011), and their explanations tend to focus on procedures (Landtblom & Sumpter, 2021; Landtblom, 2018). Textbooks typically used in classes for K-8 PSTs continue to primarily teach measures of center from a computational fluency perspective. PSTs need to encounter a more robust introduction to statistics course curriculum where statistical thinking, discussions of typical measures of center misconceptions are addressed, and how these constructs are similar/different in context.

As the researchers classified and discussed PSTs' correct responses, the idea of how faculty teach PSTs extreme values and outliers in the context of choosing the best measure of center was discussed. The researchers noted in PSTs responses the idea of one or more extreme values *pulling* the mean in the direction of that value arose. Is the description "the mean goes down" mathematically correct to describe the impact on the mean?

The authors of this paper, instructors all, admit that we use the action verb when talking to PSTs about this concept. In some sense, the verb *pull* personifies the extreme value and provides a dynamic and visual explanation as to why some distributions are said to be skewed left or skewed right. On the other hand, the data is not active, it is static (nothing is actually moving!), so the phrase *pull* in this context may obscure a conceptual understanding of the relationship between the mean, the appropriate measure of center, and the distribution.

In the original NAEP item, as well as in this study, the word mean is followed by the parenthetical "(average)." Providing this connection in the stem of the question may reinforce PSTs' (sometimes false) expectation that the mean is the more appropriate measure of center because it is emphasized that the mean is the average, and, as such, perhaps the only true measure of center.

Since the mean and median were given in the NAEP item, the intent was to focus on students' statistical thinking rather than performing computations to explain their best choice of center. Outliers, in general, are a major construct when determining the best center measure. Thus, the low value "10 customers" was designed to draw attention to the eighth graders and PSTs in this study and to prompt them to choose median as the best center measure for the data without performing any calculations.

The NAEP key referred to "10 customers" as an outlier. However, based on the formal meanings, the data point of 10 customers is neither an outlier nor an extreme value. Thus, PSTs can possibly make an argument that the mean could be a reasonable representation for the data. We agree with Metz (2010) that extra attention be given to designing well-constructed tasks where statistical thinking explanations require PSTs to demonstrate interconnectedness of statistical concepts.

Limitations

There are a few possible limitations to this research. First, the sample size of 44 PSTs is small. Second, the 2007 NAEP item was given on the last day of the semester, along with several other tasks, when PSTs were ready to be done with the course and might have not taken the time to work on the NAEP item. Third, incentives were not offered, so it is possible that PSTs might have not tried their best, even though they were given ample time to work on all the tasks.

Recommendations

Future research includes modifying the explanation component so PSTs must better explain their statistical thinking for their choice of best measure of center using the context of the problem. Additionally, removing the word "average" from the question may reduce the number of circular arguments given by PSTs. Future research could also include implementing a pre- and post-test design with the post-test item embedded in the final exam. This design will provide additional understanding of how much knowledge PSTs bring to the course versus how much they learn in the course and provide incentive for giving thoughtful consideration for their answers.

References

American Statistical Association & National Council of Teachers of Mathematics [ASA & NCTM]. (2013). *Preparing preK-12 teachers of statistics: A joint position statement.* Retrieved from https://www.nctm.org/Standards-and-Positions/Position-Statements/Preparing-Pre-K-12-Teachers-of-Statistics/

Association of Mathematics Teacher Educators. (2017). *Standards for Preparing Teachers of Mathematics,* Raleigh, NC: Author.

Beckmann, S. (2018). *Mathematics for Elementary Teachers with Activities* (5th ed.). New York: Pearson.

Bennett, A., Burton, L., Nelson,T., & Ediger, J. (2016). *Mathematics for Elementary Teachers: A Conceptual Approach* (10th ed.). New York: McGraw Hill.

Billstein, R., Boschmans, B., Libeskind, S., & Lott, J. (2020). *A Problem Solving Approach to Mathematics for Elementary Teachers* (13th ed.). New York: Pearson.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of U.S. and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology, 31*(6), 839-855.

Chance, B. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, *10*(3). DOI: 10.1080/10691898.2002.11910677

Conference Board of the Mathematical Sciences. (2012). *The Mathematical Education of Teachers II.* Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.

Engledowl, C., & Tarr, J. E. (2020). Secondary Teachers' Knowledge Structures for Measures of Center, Spread & Shape of Distribution Supporting Their Statistical Reasoning. *International Journal of Education in Mathematics, Science and Technology*, *8(*2), 146-167.

Estrada, A., Batanero, C., & Fortuny, J. M. (2004). Un estudio comparado de las actitudes haciala estadística en profesores en formación y en ejercicio [A study comparing the attitudes towards statistics of training and in-service teachers]. *Enseñanza de las Ciencias*, *22*(2), 263- 274.

ASA. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016*. Retrieved from http://www.amstat.org/education/gaise

Groth, R., & Bergner, J. (2006). Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode. *Mathematical Thinking and Learning*, *8*(1), 37-63. DOI: 10.1207/s15327833mtl0801_3

Groth, R. (2009). Characteristics of Teachers' Conversations About Teaching Mean, Median, and Mode. *Teaching and Teacher Education*, *25*, 707-716.

Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 199-209). New York: Springer.

Landtblom, K. (2018). Prospective teachers' conceptions of the concepts mean, median and mode. In H. Palmér & J. Skott (Eds.), *Students' and teachers' values, attitudes, feelings and beliefs in mathematics classrooms* (pp. 43–52). Cham, Switzerland: Springer

Landtblom, K., & Sumpter, L. (2021). Teachers and prospective teachers' conceptions about averages. *Journal of Adult Learning, Knowledge and Innovation*, *4*(1), 1-8. DOI: 10.1556/2059.03.2019.02

Leavy, A., & O'Loughlin, N. (2006). Preservice teachers' understanding of the mean: Moving beyond the arithmetic average. *Journal of Mathematics Teacher Education, 9*(1), 53-90.

Metz, M.L. (2010). Using GAISE and NCTM Standards as Frameworks for Teaching Probability and Statistics to Pre-Service Elementary and Middle School Mathematics Teachers. *Journal of Statistics Education, 18*(3). DOI: 10.1080/10691898.2010.11889585

Long, C., DeTemple, D., & Millman, R. (2014). *Mathematical Reasoning for Elementary Teachers* (7th ed.). New York: Pearson.

Musser, G., Peterson, B., & Burger, W. (2020). *Mathematics for Elementary Teachers: A Conceptual Approach* (10th ed.). Hoboken, NJ: Wiley.

National Council of Teachers of Mathematics. (1980). *An Agenda for Action.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics.* Reston, VA: Author.

National Governors Association Center for Best Practice & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington D.C.: Author.

O'Daffer, P., Charles, R., Cooney, T., Dossey, J., & Schielack, J. (2007). *Mathematics for Elementary Teachers* (4th ed.). New York: Pearson.

Reaburn, R. (2013). Pre-service Teachers' Understanding of Measures of Centre: When the Meaning Gets Lost?. In V. Steinle, L. Ball and C. Bardini (Eds.), *Mathematics Education: Yesterday, Today, and Tomorrow (Proceedings of the 36th Annual Conference of the Mathematics Education Research Group of Austral*asia) (pp. 562-569). Melbourne, VIC: MERGA.

Saidi, S.S., & Siew, N.M. (2019). Assessing Students' Understanding of the Measures of Central Tendency and Attitude towards Statistics in Rural Secondary Schools. *International Electronic Journal of Mathematics Education*, *14*(1), 73-86.

National Center for Education Statistics (NCES). (1990). *National Assessment of Educational Progress (NAEP): Mathematics Assessment*. U.S. Department of Education, Institute of Education Sciences.

NCES. (2007). NAEP: *Mathematics Assessment*. U.S. Department of Education, Institute of Education Sciences.

Whitaker, D., Foti, S. & Jacobbe, T. (2015). The levels of conceptual understanding in statistics (LOCUS) project: Results of the pilot study. *Numeracy*, 8(2), Article 3. http://dx.doi.org/10.5038/1936-4660.8.2.3

Author Information

Ha Nguyen
https://orcid.org/0000-0002-9940-7761
California State University, Dominguez Hills
1000 E. Victoria Street | Carson, CA 90747
United States
Contact e-mail: *hnnguyen@csudh.edu*

Eryn M. Maher
https://orcid.org/0000-0002-2833-4614
Georgia Southern University
PO Box 8093 | Statesboro, GA 30460
United States

Gregory Chamblee
https://orcid.org/0000-0003-0438-4059
Georgia Southern University
PO Box 8134 | Statesboro, GA 30460
United States

Sharon Taylor
https://orcid.org/0000-0002-3325-1266
Georgia Southern University
PO Box 8093 | Statesboro, GA 30460
United States