ASSESSMENT IN STATISTICS: *MOVE BEYOND PROCEDURAL FLUENCY AND GET STUDENTS THINKING CRITICALLY*

STEPHEN J. MILLER
WINCHESTER THURSTON SCHOOL
PITTSBURGH, PA

NCTM ANNUAL MEETING – SAN DIEGO
SESSION 386 – FRIDAY, 5 APRIL 2019

INTRODUCTION

- Statistics has been recommended in the High School mathematics curriculum for many years
  - NCTM
  - Common Core
- Implementation of these standards has been inconsistent

INTRODUCTION

- Common Core gives statistics a more prominent role
- Many high school mathematics teachers will find themselves in the dual role of being both teachers of ***mathematics*** and teachers of ***statistics***
  - Statistics can be *new* and *uncomfortable* territory

INTRODUCTION

- Common Core Standards
  - Beyond mechanical and computational aspects
  - ***Conceptual understanding*** necessary for sound statistical reasoning

INTRODUCTION

- *Computational fluency* is relatively easy to ***teach*** and ***assess***
- Developing and assessing ***statistical reasoning*** is more challenging

MAIN PURPOSE OF THIS SESSION

- To provide teachers with four recommendations so they can move beyond computations in teaching and assessing statistics
  - With a of goal of increasing students' understanding of statistical concepts and reasoning

FOUR RECOMMENDATIONS

1. Look Beyond Computational Fluency
2. Ask Good Statistics Questions
3. Give students the opportunity to practice "Talking Statistics"
4. Provide authentic assessments and meaningful feedback

RECOMMENDATION 1: LOOK BEYOND COMPUTATIONAL FLUENCY

- The ability to compute is of little value *by itself*
- Being able to compute a standard deviation or calculate a confidence interval estimate of a population mean is not much use ***without knowing what these computations reveal about the population of interest***.

RECOMMENDATION 1: LOOK BEYOND COMPUTATIONAL FLUENCY

- Students need to demonstrate more than an ability to compute.
- They need to show understanding and be able to draw conclusions in context.

RECOMMENDATION 1: LOOK BEYOND COMPUTATIONAL FLUENCY

- Students must go beyond the mechanics and understand the ***logic of statistical procedures*** and the ***interpretation of results***

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- The key to being a great statistics teacher is to ask good questions.
- So what is a good question?
- One possible answer: A question that incorporates a ***conceptual component*** or an ***interpretation component*** – or both.

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Example 1: What is the standard deviation of the following 10 numbers?

2009	2015	2002	1979	2032
1991	2016	2030	2001	1990

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- That question is not very good
- Some might not even say this is a statistics question
- Why?
- No context, no interpretation – purely computational

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Example 2: The data below are lifetimes (in hours) for 10 light bulbs from a new brand your school is considering for use in the football stadium light fixtures:

2009	2015	2002	1979	2032
1991	2016	2030	2001	1990

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

2009	2015	2002	1979	2032
1991	2016	2030	2001	1990

- What is the value of the standard deviation of the 10 lifetimes?

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Better than the first example, but not much better
- Why?
- We now have a context, but the question is still purely computational – no interpretation
- Information provided by the context is irrelevant to what students are asked to do with the data

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Example 3: The data below are lifetimes (in hours) for 10 light bulbs from a new brand your school is considering for use in the football stadium light fixtures:

2009	2015	2002	1979	2032
1991	2016	2030	2001	1990

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

2009	2015	2002	1979	2032
1991	2016	2030	2001	1990

- (a) What is the value of the standard deviation of the 10 lifetimes?

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- (b) The standard deviation for the lifetimes of bulbs from the brand currently in use is 40 hours. What does the standard deviation that you computed for the sample of light bulbs from the new brand tell you about how this brand might compare with the old brand?

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- This question is better than the previous two questions
- Part (a) is still purely computational
- Part (b) asks students to go beyond computation to provide an interpretation of the standard deviation that makes sense in the context of the data

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Example 4: To make this question even better, let's add a part (c):
- (c) Replacing stadium lightbulbs is difficult and requires special equipment. Because of this, rather than replace individual bulbs as they burn out, the school plans to replace *all* the stadium light bulbs as soon as one burns out.

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- (c) (continued): The mean lifetime is 2000 hours for both the current brand and the new brand under consideration, and the cost of the two brands is the same. Would you recommend that the school stay with the current brand or change to the new brand? Explain your reasoning.

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Part (a) is still computational
- Part (b) asks students to go beyond computation to provide an *interpretation* of the standard deviation that makes sense in the context of the data.

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Part (c) enriches the question by asking students to answer a question of interest that requires them to demonstrate conceptual understanding of the standard deviation and what it measures.

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- Example 4 is even better because:
  - It has a relevant context
  - It requires interpretation
  - It requires that students demonstrate a level of conceptual understanding

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- When developing questions, **ask yourself these questions**:
  - (1) Does the question provide a meaningful context?
  - (2) Is the context relevant to the problem – that is, do students need to make use of the context?

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- (3) Does the question have an interpretation component?
- (4) Does the question have an aspect that requires students to demonstrate an understanding of the relevant concepts?

RECOMMENDATION 2: ASK GOOD STATISTICS QUESTIONS

- (5) If the context, interpretation or conceptual component is missing, is there a way to improve the question by incorporating one or more of these components?

RECOMMENDATION 3: GIVE STUDENTS THE OPPORTUNITY TO PRACTICE "TALKING STATISTICS"

- Data analysis has a ***communication component*** as well as a ***computational component***
- The abilities to ***interpret*** and ***communicate*** results in context are not innate – they are skills that develop slowly and with practice

RECOMMENDATION 3: GIVE STUDENTS THE OPPORTUNITY TO PRACTICE "TALKING STATISTICS"

- It might be ***painful*** to listen to or read through students' initial attempts at explaining why they selected a particular method or how they interpreted results from a data analysis…

RECOMMENDATION 3: GIVE STUDENTS THE OPPORTUNITY TO PRACTICE "TALKING STATISTICS"

- However…opportunities to practice, coupled with ***meaningful and constructive feedback***, are essential
- By providing these, teachers can guide students' development in "talking statistics"
- This takes patience, but it is rewarding

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- If statistical reasoning is to be valued, it is important that we assess both ***conceptual understanding*** and ***the ability to interpret results in a meaningful way***.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Assessing computational fluency is relatively easy because students' responses are
  - Correct
  - Mostly correct (with some minor arithmetic errors)
  - Incorrect
- These errors are usually easy to identify and address

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Assessing students' abilities to interpret the results of a data analysis or their understanding of important statistical concepts is ***more difficult***
- So how can you do it?

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Use a ***holistic rubric*** to classify students' responses
- We propose a 3-2-1-0 scale, but that can be modified
- Two rubrics:
  - Communication and Interpretation
  - Conceptual Understanding

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- For questions that address ***communication and interpretation***:
  - 3: The interpretation is appropriate, complete, and well communicated
  - 2: The interpretation is appropriate and complete, but is not communicated well

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- For questions that address ***communication and interpretation***:
  - 1: The interpretation includes appropriate statements that demonstrate some understanding, but is incomplete or not quite correct
  - 0: No interpretation provided, or it is incorrect and/or inappropriate

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- For questions that address ***conceptual understanding***:
  - 3: Demonstrates an understanding of the relevant concept that is clearly communicated
  - 2: Demonstrates some understanding of the relevant concept, but the understanding is not communicated clearly

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- For questions that address ***conceptual understanding***:
  - 1: Demonstrates only limited understanding of the relevant concept
  - 0: Demonstrates little or no understanding of the relevant concept

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Recall Example 4, part (c):
- Replacing stadium lightbulbs is difficult and requires special equipment. Because of this, rather than replace individual bulbs as they burn out, the school plans to replace *all* the stadium light bulbs as soon as one burns out.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- The mean lifetime is 2000 hours for both the current brand and the new brand under consideration, the standard deviation for the lifetimes of bulbs of the brand currently in use is 40 hours and is roughly 20 hours for the new brand, and the cost of the two brands is the same.
- Would you recommend that the school stay with the current brand or change to the new brand? Explain your reasoning.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Sample response #1:
  - I don't think it makes any difference which brand the school uses. The mean lifetime is the same for both brands, so the school may as well stay with the old brand. There is no reason to change to the new brand.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Sample response #1:
  - Earned a zero because it fails to recognize the role that variability plays in the decision
  - The response does not demonstrate any understanding of variability

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Sample response #2:
  - The standard deviation of the new brand is less than the standard deviation of the current brand. There appears to be less variability in the lifetimes of the new brand, so the lifetimes of the new brand will tend to be more similar. We should go with the new brand.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Sample response #2:
  - Would earn a 2 because the response recognizes that less variability is desirable, but fails to consider mean lifetime and does not explain why less variability is desirable.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Sample response #3 (includes these phrases):
  - Current brand is much more variable than the new brand, so the first bulb to burn out is likely to be quite a bit before 2000 hours.
  - New brand is less variable, so lifetimes cluster more tightly around the mean of 2000 hours, so the first bulb to burn out is likely to last closer to 2000 hours.

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

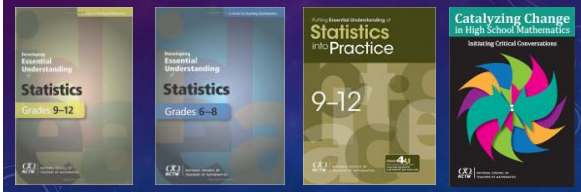
- Sample response #3:
  - Earned a 3 because it clearly recognizes the role that variability plays in the decision
  - The response demonstrates a strong understanding of variability

RECOMMENDATION 4: PROVIDE AUTHENTIC ASSESSMENTS AND MEANINGFUL FEEDBACK

- Providing meaningful feedback is challenging
- We need to think about how we can help them develop deeper understanding and express their thinking more clearly

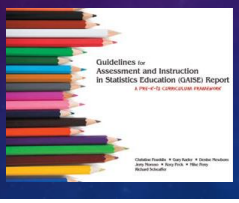
RECOMMENDED RESOURCES

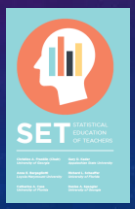
- From NCTM:



RECOMMENDED RESOURCES

- From the American Statistical Association (www.amstat.org):





RECOMMENDED RESOURCES

- From the American Statistical Association (www.amstat.org):

QUESTIONS AND CONTACT

- Questions?
- Feel free to contact me:
  - Stephen Miller
  - sm1016@gmail.com
  - Pittsburgh, PA