6-2019

Understanding Math Concepts in Data and Probability at Secondary Level: A Review of the Literature

Hyungsuk Maeng
hmaeng@stcloudstate.edu

Understanding Math Concepts in Data and Probability at Secondary Level: A Review of the Literature

by

Hyungsuk Maeng

A Starred Paper

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Curriculum and Instruction

May, 2019

Starred Paper Committee:
Ramon Serrano, Chairperson
Hseuh-I Lo
Jerry Wellik

Table of Contents

Chapter 1: Introduction

Background of the Study

When I was a middle school student, from my memory, it was not easy for me to understand math concepts in data and probability, and to develop problem-solving skills about those math concepts. Harper (2004) states that "since graphical displays of data are increasingly used in magazines, in newspapers, and on television to communicate relationships among numerical data, it is important to expose middle school students to various types of contexts where interpretation of graphical data is necessary" (p. 340). I wish I had been exposed to different types of contexts through math problems in which I could practice interpretation of graphical data.

To make it worse, most math problems with those concepts were produced based on real world applications. One of the authors states that "Youngsters and adults alike are confronted daily with situations involving statistical information" (Franklin & Mewborn, 2008, p. 14). I struggled with misconceptions that I had in the process of solving real world problems. I had to deal with and overcome those misconceptions so that I would sometimes ask a help of my math teacher and my friends. As time went by, I got used to and understood math concepts in data and probability. I was also able to develop problem solving skills about those concepts through solving many of math problems. It was a meaningful time for me to try to get true mastery of those math concepts.

Significance of the Study

This study focuses more on the secondary school students. They might face difficulties or misconceptions in exploring math concepts in data and probability. Authors state that: "Data

from the National Assessment of Educational Progress (NAEP) over the past 15 years indicate that middle school students have some difficulty finding the mean and median" (Zawojewski & Shaughnessy, 2000, p. 436). The lack of understanding of definition of necessary terminologies can affect academic achievement of secondary students on math concepts in data and probability.

How a teacher guides a student could also have an influence on students' performance. Authors state that: "we describe how one teacher organized instruction for middle-grades students to help them develop an understanding of the relationship between data type and the selection of measures of center to describe data" (Aisling, Leavy, Friel, & Mame, 2009, p. 345). The study also looks at how a teacher sets up a role and implement it has an impact on student learning process, especially when teachers are engaged in the student learning process.

Nowadays students can reach out to technology and utilize 21st century skills through a variety of electronic devices like I-pads, cell phones, laptop computers, and so on in studying data and probability math concepts. For example, 'TinkerPlots,' a software created for educational purposes can help students develop a better understanding of target math concepts about data sets. One of the authors states that: "Is there anything new to introduce in the way of graphical representations for middle school students? This article presents a new data representation tool called the hat plot, which is a featured tool of the data analysis software 'TinkerPlots Dynamic Data Exploration'" (Watson, Fitzallen, Wilson, & Creed, 2008, p. 4). Students can learn how to utilize those useful programs that can have huge impact on student learning in a positive way. 'TinkerPlots' software gives students considerable freedom to create graphs to tell the stories of their data sets. Because the hat plot representation will most likely be

new to middle school students, it is important to explore the concepts that the hat plot represents, how students use it, and how it links with students' intuitive notions of distribution." (Watson et al., 2008, p. 4). Through this program, students can approach data and probability math concepts more comfortably with fun and they might feel that they have initiative to operate the software in creating what they want to show about the given data sets.

Importance of the Study

High school students are supposed to go through math concepts in data and probability aligned to math standards that each state education department provides. Authors state that: "The importance of statistics education as an integral part of the mathematics curriculum was signaled by NCTM in its Curriculum and Evaluation Standards for School Mathematics" (Franklin & Mewborn, 2008, p. 10). NCTM can provide a guideline of what would be the necessary math concepts students are supposed to know and understand as their grade level expectations.

High school students are very often exposed to real world application with those math concepts day in and out. There might come a time for them to deal with financial plans like selecting better car insurance options or designing better retirement plans with knowledge and application that they obtain in exploring through the concepts of data and probability. Authors state that: "Making sense of data and dealing with uncertainty are skills essential to being a wise consumer, an enlightened citizen, and an effective worker or leader in our data-driven society" (Franklin & Mewborn, 2008, p. 10). By learning and utilizing the data and probability math concepts, students will be able to figure things out day in and out in a data-driven society.

If high school students might plan to go to college or graduate school, high school students will need to know how to collect and analyze data in their educational research areas. Based on a statement by De Groote, Hitchcock, and McGowan (2007), statement “quantitative administrative data are regularly collected on the number of reference questions asked and include the service point, type of patron, and query category (Table 1). Data collected from 1990 to 2005 were entered into a spreadsheet” (p. 27). In the academic research processes, students will find it important how they handle the data aligned to research purposes.

If high school students get a job in the field related to statistics after graduation, the students will also need to be aware of how to collect and organize data to find meaningful results for their profession. Based on a statement by Bansal and Srivastava (2018), the future of almost every sector whether it is business, IT, Medical Sciences, Forensics, and many more are revolves around the ability to make predictions and discover patterns in data” (p. 15). For those reasons, high school students may want to understand what those math concepts are and how students utilize them for real world applications, academic purposes, and career readiness in their lives.

As a math teacher, I want to help students get prepared for those situations by exploring those math concepts with students. Along with that, students might feel difficulty and have misconceptions when they study those concepts. I hope this research would be an opportunity for me to be able to find and think what needs of students would be in data and probability. As a math eudcator, I want to consider how I can help them meet the needs and get true mastery of those concepts. This is the reason why I am interested in this research.

Research Questions

Two research questions guided the development of this starred paper:

1. What are difficulties or misconceptions of math concepts in data and probability for secondary students?
2. How does technology have impact on student learning regarding math concepts in data and probability?

I critically reviewed research papers located under the following keywords: *Statistics, Categorical data, Numerical data, Experimental probability, Theoretical probability, and Thinker plots software*.

Chapter 1 includes the background of the study, research questions, focus of the paper, importance of the topic, strategies to explore the meaningful articles, and definitions of terms. In Chapter 2, I review recent research literature regarding my research topic and questions. Chapter 3 discusses the conclusions and implications of research reviewed in Chapter 2.

Strategies to Explore the Meaningful Articles

When it comes to exploring meaningful articles aligned to research questions, we would need some strategies for the effective and precise research process. Without those strategies, it could be difficult for us to find meaningful articles, which could lead to the consequence that the paper might not be able to support the research questions effectively. Therefore, I would like to introduce some strategies I have taken to find the articles through several steps.

First, I used search engines on the internet websites in the process of finding meaningful articles. It would be better for me to use search engines than to find the articles from books published in the library. This is the reason why I was able to reach out and find more articles

effectively through internet. Using search engines would also have me save much time to be spent for searching the articles. Specifically, I used search engine Google scholar, EBSCO, and the Article and Databases section with SCSU.

Second, when it comes to the process of finding articles through search engines, I did choose certain keywords relevant to my research questions. I had typed those words into each of the search engines and then articles related to those vocabularies came up. I looked at the titles of articles coming up, clicked the articles, and read whole articles. After that, I chose several meaningful articles for my research questions. I used those vocabularies to explore the articles: data and chance, middle school probability, high school statistics, learning elementary data and chance, 21st century skills for data and chance, electronic devices for learning probability, 21st century skills in math class, the relationship between 21st century skills and math class, software for middle school data and chance concepts, so on. By using search engines from websites and putting keywords into the engines, I was able to explore the meaningful articles for my research.

Definitions of Terms

1. Statistical Problem-Solving: Franklin and Mewborn (2008) state that students should be actively involved in the statistical problem-solving process: designing the questions to answer, collecting and representing the data, analyzing the data, and interpreting the data.
2. Categorical Data: Aisling et al. (2009) states that categorical variables, however, may be either nominal data or ordinal data. As nominal data, names are assigned to objects as labels based on some category definition, such as favorite kinds of pets (e.g., dog, cat, hamster, etc.), the data values are the nonnumeric group labels. There

is no natural ordering to these data. Ordinal data (i.e., ordinal categorical data values) can be ordered, but operations such as addition and subtraction are meaningless. These data may be ranked in some numerically meaningful way. For example, strongly disagree to strongly agree may be defined as 1 to 5 where the numbers represent: (1) strongly agree, (2) agree, (3) neutral, (4) disagree, or (5) strongly disagree.

3. Numerical Data: Aisley et al. (2009) states that numerical measurements may involve continuous data, such as height or temperature or discrete data, often occurring as counts, such as the number of children in a family, and can be readily ordered from smallest to largest and used with operations such as addition or subtraction.
4. Experimental Probability: Andrew (2009) states that Experimental Probability underlies more theoretical notions of probability, and participation in such activity helps a person understand fundamental principles used in chance problems.
5. Theoretical probability: Andrew (2009) states that multiple experimental approaches help students see that there could be an 'ideal' probability that may be calculated without actually performing an experiment. This is the notion of theoretical probability.
6. Thinker Plots Software: Watson et al. (2008) states that TinkerPlots software gives students considerable freedom to create graphs to tell the stories of their data sets.

Chapter 2: Review of the Literature

This paper examines how secondary students understand and study math concepts in data and probability in secondary school level. When the students study those concepts, they might face difficulties or misunderstandings in exploring math concepts in data and probability. Those difficulties or misconceptions can affect student understanding and achievement on those math concepts. The focus of this paper includes what difficulties or misunderstandings might be in studying data and probability, how students learn math concepts in data and probability with well-organized instruction and appropriate teacher's intervention, how software can help students understand math concepts in data and probability, and why students learn math concepts in data and probability.

Difficulties or Misconceptions in Studying Data and Probability Math Concepts

One of the authors shows difficulties students might have like "I'm just not sure how to help my students understand that the median is not a statistic that they use with categorical data. What can I do to help this make sense to them?" (Aisley et al., 2009, p. 345). Students might feel difficult in understanding meaning and its application of the median with categorical data.

Aisling et al. (2009) describes that students might feel difficulties of how to distinguish between categorical data and numerical data when as labels category labels are used. They also show "although the median is a relatively simple measure to identify, research studies and large-scale assessments of students' mathematical knowledge indicate that understanding the median is far more complex that we might realize. The students' struggles suggested a need to provide some problem or discussion that helps clarify the distinctions between categorical and numerical data types and their relationships to the median" (p. 346). Based on understanding the median

and the discussion about how to distinguish between categorical and numerical data types, students can obtain a better understanding about the math concept.

Zawojewski and Shaughnessy (2000) describe that middle school students may struggle with finding the mean and median with data from the National Assessment of Educational Progress (NAEP) for over the past 15 years. "Further, results indicate even greater problems in selecting and using the different statistics appropriately and that these difficulties persist into the high school years" (p. 436). Students also might face difficulties with problem-solving process. "When students engage in probability problem solving many unexpected situations can arise due to the counterintuitive nature of probability concepts. These situations can be difficult for students and challenging for teachers to analyze during teaching" (Koellner, Pittman, & Brendefur, 2015, p. 29). Students can face unexpected situations in the problem-solving process, which can make students feel difficult to find a way to figure out.

When students study math concepts in data and probability, they might also have misconceptions. "The misconception identified in this scenario is that students are identifying each path as being equally likely. LeCoutre (1992, as cited by Koellner et al., 2015) identified this phenomenon in his work as 'misunderstanding of equiprobability.' Students with an equiprobability bias assume all outcomes of an event are equally likely regardless of the compound events or junctures throughout the maze" (p. 30). If students have a misunderstanding of the target math concepts, it could lead to wrong answer. Understanding definition of the target math concepts can be the first step for students to take before they get into problem-solving process.

Andrew (2009) describes that students should use mathematical reasoning in problem-solving process, but many students do not do it most of the time. Van de Walle (2004, as cited by Andrew, 2009) states that: "many students have a naïve concept of probability in the beginning and that understanding probability is progressive." Some children believe, and event will happen "because it is my favorite color" or because "it did it that way last time" (p. 35). Based on understanding the target math concepts, students should know how to explain problem-solving process with mathematical reasoning.

Andrew (2009) describes that experiments cannot be performed all the time, nor are they time efficient. The more they do experiment with abstract measure of probability, the more students can lose their interest on experiment. Andrew also mentioned: "This decreased reliance on experimentation is superior to over-reliance on theoretical probability with little experimental backing" (p. 36). Abstract approach can make students feel difficulty, which can lead to the situation where students might lose their interest in studying the math concept.

Perner (1979) shows:

> the performance difference in the two types of tasks on the basis of their theory of children's acquisition of the concept of probability. According to their theory, young children have difficulty in comprehending part whole relationships. Every probability ratio is a part-whole relation, since it consists of relating the number of favorable cases (part) to the number of all cases (whole). A double-process task requires coordination of two such part-whole relations, one probability ratio for each process, whereas a single-process task contains three part-whole relationships. (p. 1121)

When it comes to probability ratio, students need to understand a part-whole relation. Otherwise, it can affect the next step in the problem-solving process and eventually lead to wrong answer.

Zawojewski and Shaughnessy (2000) describe difficulties students might have by quoting "Most students in the 7th and 11th grades appeared not to understand technical statistical terms such as mean, median, mode, and range. However, there is evidence that they could compute the mean when asked for the average" (Brown & Silver 1989, p. 28 as cited by Zawojewski & Shaughnessy, 2000). Judith and Michael also mentioned that students are confused about the meaning of the measures of central tendency, especially for eighth and twelfth grade students. Students need to comprehend the math terms of the target concepts otherwise they can get confused in the problem-solving process.

A Need of Teacher's Instruction on Student Learning

Watson et al. (2008) describes how students learn data and probability concepts saying: "The question for middle school mathematics teachers is whether students find the features of hat plots useful in telling the story within the data set being studied. 'Telling the story' in a hat plot is a representational transition to drawing an informal inference about the data being investigated" (p. 7). Students might have difficulties when they take an initial step in finding the features of hat plots. In the problem-solving process, teachers can give students a guidance that students might start the process with "Telling the story." It can make students feel easier to approach the target math concepts.

Aisling et al. (2009) describes how students can deal with their struggles with teacher instruction and engagement like:

> The students' struggles suggested a need to provide some problem or discussion that helps clarify the distinctions between categorical and numerical data types and their relationships to the median. Teacher decided to have his students answer the two questions posed in the problem, collecting data from their classmates before they worked on the data presented in the text. (p. 347)

Through teacher's appropriate intervention, students can be provided with directions in the problem-solving process.

Zawojewski and Shaughnessy (2000) suggest teachers to ask students several questions for reviewing students' performance:

> Do your students understand the procedures for finding the mean and median?
>
> Do your students make mathematical connections between statistics and other branches of mathematics? Do your students understand the terminology of mean and median?
>
> Do your students understand the relationship between the distribution of the data set and the selection of mean and median? (p. 439)

Teachers can ask students several meaningful questions where students can see the point regarding the target math concepts in the learning process. By preparing answers to those questions, students can build up mathematical reasoning skills, too.

Jacobs (1999) suggests teachers use written activities related to those used for the second study rather than to assess students' acquisition by doing oral questioning like interviewing, but this way might be tremendously time-consuming. Students need appropriate

questions that can be helpful for them to consider problem-solving strategies as the following: "Do all categories contain about the same number of shoes? Do some contain more shoes or fewer shoes than others?" (p. 12). Students need to work on math activities through written forms with helpful questions to come up with their own ways to figure out math problems.

Franklin and Mewborn (2008) suggest "how teachers engage student learning in designing four stages: Formulate questions, Collect and represent the data, Analyze the data, and Interpret the data." (p. 13). According to those four steps designed by a teacher, students can approach the target math concepts more efficiently and easily step by step.

McClain (1999) describes effective ways to find out where students are in studying target concepts like: "As a result of the need to clearly understand students' interpretation of, and reasoning about, the task, two of the group discussions and the subsequent whole class discussion were videotaped. Viewing the tape gave me the opportunity to analyze the students' ways of reasoning beyond what I observed as I monitored the different groups" (p. 375). As a part of teacher's observation, class activity can be recorded and analyzed, which can be a chance to see how students perform.

Andrew (2009) describes that when students understand the experimental probability related to certain events, it allows a student to get better understanding of what probability truly means saying:

> In particular, (1) performing experiments helps students differentiate between the elements of the sample space and a specific outcome of interest, the 'success' outcome, (2) performing an experiment a large number of times can demonstrate to students the likelihood of certain types of outcomes, and (3) matching experimental results with the

corresponding theoretical probabilities strengthens the meaning of theoretical probability. (p. 34)

Perner (1979) describes an experiment that can help students develop their understanding of probability. In one experiment, two boxes were given to children, and different proportions of red and white beads were filled in both boxes. Children were asked "which box one is more likely to blindly draw a bead of a given color ("double-process task"). The result is that children in the experiment provided more correct answers that those who were given only a single box ("single-process task"). He also described another experiment by using spinners like "A spinner consists of a disk with two unequally sized segments of different colors and a pointer that can be rotated on it. The probability that this pointer will stop at a particular color is determined by the angular length of the segment covered with the color" (p. 1121). Students can get better understanding about probability by performing experiments with teacher's appropriate instructions.

McClain (1999) shows that acquisition of knowledge with students could be done by monitoring students and making them as groups like: "As students began working in their groups, I walked around the classroom to try to monitor their activity and begin to understand how they were reasoning. Several groups began by finding the mean of the data set. One group asked for the calculators, and it could be argued that their request triggered a request from the other groups" (p. 376). As part of teacher's intervention, a teacher can move around to watch how students perform and process their mathematical reasoning. Students can directly get a help form the teacher moving around the classroom.

The Effectiveness of Utilizing Technology

Watson et al. (2008) describes that "when students represent mathematical ideas, lots of kinds of graphical forms are available for middle school students from mathematical methodology in ancient times to modern uses of technology like graphing calculators" (p. 4). By using graphing calculator or graphing calculator app like "DESMOS," students can approach the target math concepts with more graphical information, which can make students feel more interest and fun in studying.

> Some software has been invented for the use of mathematics like:
>
> Thinker-Plots software gives students considerable freedom to create graphs to tell the stories of their data sets. Because the hat plot representation will most likely be new to middle school students, it is important to explore the concepts that the hat plot represents, how students use it, and how it links with students' intuitive notions of distribution. Given the results of research showing that students have difficulties interpreting box-and-whisker plots, it was somewhat surprising to teachers how easily these middle school students were able to use the hat plots, often on their own initiative. After teachers initially introduced the software, they reported observing most students employing hat plots when constructing graphical representations from new data sets, possibly because of the ease with which hat plots in Tinker-Plots can be applied to a graph. (Watson et al., 2008, p. 5)

When students use Thinker-Plots to analyze data set and set up plots, they seem to feel interest in studying data math concepts even though they think the target math concepts are difficult to explore.

The Purpose of Learning Data and Probability

Harper (2004) describes a need of studying math concepts in data and probability in connection with real world application like: "Since graphical display of data are increasingly used in magazines, in newspapers, and on television to communicate relationships among numerical data, it is important to exposed middle school students to various types of contexts where interpretation of graphical data is necessary" (p. 340). Students can face the situation where they need to interpret graphical display of data in variety of circumstances day in and day out as part of real-world application.

Franklin and Mewborn (2008) describe that the chosen problems concentrate on statistics and data analysis, primarily regarding the use and misuse of statistics for real world applications. Harper (2004) expects students to utilize statistical concepts to analyze and communicate verified interpretations of the graphical and tabular forms of data from a variety of real-world contexts. Students are often exposed to real world contexts with math concepts in data and probability so that they may want to build up ability to analyze and interpret data that can be given in real world situation.

> Franklin and Mewborn (2008) describe a need of studying data and probability like: Youngsters and adults alike confronted daily with situations involving statistical information. Making sense of data and dealing with uncertainty are skills essential to being a wise consumer, an enlightened citizen, and an effective worker or leader in our data-driven society. The importance of statistics education as an integral part of the mathematics curriculum was signaled by NCTM in its Curriculum and Evaluation Standards for School Mathematics. (p. 10)

Curriculum and system of a society is designed based on data, and interpretation ability on data is important for students to understand how this data driven society goes and how they are being an appropriate citizen.

McClain (1999) describes a need for students to study with academic questions: "Do students first need to know how to construct various types of graphs before they can engage in an analysis of data, or can they learn how to construct various types of graphs by engaging in data analysis? Further, can they engage in data analysis before they have acquired a conceptual understanding of the multiple forms of data representation? (p. 374). Students may want to learn how to do data analysis step by step to utilize those skills for real world application, academic research process, job readiness, or so on.

Chapter 3: Conclusions and Summary

The purpose of this research paper is to analyze how secondary students can effectively learn data and probability math concepts realizing and overcoming difficulties or misconceptions. Background information of the topic was described in Chapter 1, and a review of the research literature was represented in Chapter 2. Implications from the research are discussed in Chapter 3.

Conclusions

Ten articles have been reviewed regarding math concepts in data and probability where secondary students might face difficulties or misconceptions in the problem-solving process with variety of different contexts of real-world application.

Difficulties that secondary students might have in the problem-solving process in data and probability can be overcome with appropriate teacher's instruction or intervention, and with proper usage of technology.

If teacher training opportunities can regularly be provided for teachers regarding teacher's instruction or technology usage, it would be a good chance for teachers to consider effective teaching instruction and the use of technology in math class. First, if teacher trainings for effective teaching instruction are offered, it would be a good opportunity for teachers to think how to design teaching instruction adequately depending on the level of students, how to intervene effectively when students work on individual or group work, and how to implement teaching instruction effectively. Second, when it comes to teacher trainings for technology, if we have a chance to look at what kinds of technology we can use in math class and how we utilize

online apps or software like Thinker-Plots in math class, it would be possible for teachers to be able to create more fun and effective classroom atmosphere.

Implication

Students are supposed to face a situation where they need to figure out things related to data and probability concepts in their real lives. In the process of making connection between the math concepts that they learn from math classes and real-world application, students might have some difficulties or confusion. In this learning process, teachers would be a good asset to help students understand the target math concepts and apply those concepts into real world problems. In other words, if appropriate teacher's intervention in the problem-solving process is implemented, difficulties or confusion that students might have could be lessened and students can get clear understanding for the target concept.

When it comes to data and probability math concepts, software program like Tinker Plots would be a good material for students to utilize. Through this program, students can approach to the target math concepts more easily with open mindset because students might feel this program as a game they can operate with initiative so that students could have a fun in using the program.

Summary

Appropriate teacher's intervention and well-organized instruction can help students develop a better understanding of the target math concepts in data and probability area so that it would be easier for secondary students to figure out some difficulties or misconceptions in the problem- solving process. Especially utilizing technology like "Tinker Plots" would be beneficial for secondary students to get more interest in the math concepts in data and probability. Secondary students are supposed to know why they need those math concepts for

college and job readiness and how the students make a connection between those math concepts and real-world problems.

References

Aisling, M., Leavy, S., Friel, N., & Mame, J. D. (2009). It's a fird! Can you compute a median of categorical data? *The National Council of Teachers of Mathematics, 14*(6).

Andrew, L. (2009). Experimental probability in elementary school. *Teaching Statistics. 31*(2), 34-36.

Bansal, A., & Svriastava, S. (2018). *International Journal of Recent Research Aspects, 5*(1), 15-18. ISSN: 2349-7688

De Groote, S. L., Hitchcock, K., & McGowan, R. (2007). *Journal of the Medical Library Association, 95*(1).

Franklin, C. A., & Mewborn, D. S. (2008). Statistics in the elementary grades: Exploring distributions of data. *Teaching Children Mathematics.*

Harper, S. R. (2004). Students' interpretation of misleading graphs. *The National Council of Teachers of Mathematics, 9*(6).

Jacobs, V. R. (1999). How do students think about statistical sampling before instructions? *The National Council of Teachers of Mathematics, 5(*4).

Koellner, K., Pittman, M., & Brendefur, J. L. (2015). Expect the unexpected when teaching probability. *Mathematics Teaching, 245*, 29-32.

McClain, K. (1999). Reflecting on students' understanding of data. *The National Council of Teachers of Mathematics, 4*(6).

Perner, J. (1979). Discrepant results in experimental studies of young children's understanding of probability. *Child Development. 50*(4), 1121. doi: 10.1111/1467-8624.ep7251734.

Watson, J. M., Fitzallen, N. E., Wilson, K. G., & Creed, J. F. (2008). The representational value of hats. *The National Council of Teachers of Mathematics, 14*(1).

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Mean and median: Are they really so easy? *The National Council of Teachers of Mathematics, 5*(7).