August 2017

A Study of Strengthening Secondary Mathematics Teachers' Knowledge of Statistics and Probability via Professional Development

Lina Devaul
University of Nevada, Las Vegas, linazangdevaul@gmail.com

A STUDY OF STRENGTHENING SECONDARY MATHEMATICS TEACHERS' KNOWLEDGE OF STATISTICS AND PROBABILITY VIA PROFESSIONAL DEVELOPMENT

By

Lina DeVaul

Bachelor of Science – Mathematics and Applied Mathematics
Liaocheng University
2007

Master of Science -- Curriculum and Teaching Methodology in Mathematics Education
East China Normal University
2010

A dissertation submitted in partial fulfillment
of the requirement for the

Doctor of Philosophy – Curriculum and Instruction

Department of Teaching and Learning
College of Education
The Graduate College

University of Nevada, Las Vegas
August 2017

UNLV | GRADUATE COLLEGE

Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

July 12, 2017

This dissertation prepared by

Lina DeVaul

entitled

A Study of Strengthening Secondary Mathematics Teachers' Knowledge of Statistics and Probability via Professional Development

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy – Curriculum and Instruction
Department of Teaching and Learning

Jeffrey Shih, Ph.D.
Examination Committee Co-Chair

Travis Olson, Ph.D.
Examination Committee Co-Chair

Jane McCarthy, Ed.D.
Examination Committee Member

LeAnn Putney, Ph.D.
Graduate College Faculty Representative

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

Abstract

A professional development program (PSPD) was implemented to improve in-service secondary mathematics teachers' content knowledge, pedagogical knowledge, and self-efficacy in teaching secondary school statistics and probability. Participants generated a teaching resource website at the conclusion of the PSPD program. Participants' content knowledge change and self-efficacy change were measured. After PSPD, three participants were selected to represent three types of change. Teachers' classroom instructions were video-taped and analyzed to explore the enactment of PSPD components, such as activities and concepts. Interviews were conducted to assess factors that facilitated teachers' change and the enactment of PSPD components. Preservice teachers who majored in secondary mathematics education were asked to evaluate the teaching resource website that PSPD participants generated. Their applications of content in this website and feedback to this website were analyzed.

Case study research method was adopted in this study. Findings revealed that PSPD participants' self-efficacy in teaching statistics and probability improved significantly. There was a moderate positive relationship between in-service teachers' statistical content knowledge and self-efficacy in teaching statistics and probability. PSPD participants' utilization of PD components may be influenced by school conditions, the characteristics of PD materials, and whether teachers perceive and accept PD content. Six out of 17 preservice teachers utilized materials in the teaching resource website in their lesson plans. Preservice teachers preferred teaching resources that were ready-to-use and easy to access.

Table of Contents

Approval page ii
Abstract iii
List of Tables vi
List of Figures vii
Chapter 1: Introduction 1
Statement of the Problem 1
Introduction of PSPD Program 2
Introduction of this Study 3
Research Questions 4
Importance of Probability and Statistics 4
In-service Teachers' Knowledge of Statistics and Probability 5
Pre-service Teachers' Knowledge of Statistics and Probability 6
Secondary School Statistics and Probability Professional Development 8
Teachers' Self-Efficacy in Teaching Statistics and Probability 11
Significance of this Study 13
Chapter 2: Literature Review 15
Introduction 15
Theoretical Framework 15
Theory of probabilistic intuitions improvement. 15
Theory of teacher content knowledge 18
Utilizing Content Training to Improve Stochastic Reasoning 22
Three types of improper probabilistic and statistical reasoning 22
Utilizing content knowledge training to overcome judgment biases. 26
Important Content Knowledge and Training in Secondary Probability and Statistics . 29
Statistical question 30
The meanings of mean 33
Randomness 36
Sample and sampling 43
Sampling distribution and statistical inference 47
Relationship Between Content Knowledge and Teacher Self-efficacy 53
Definitions 53
The role of teacher self-efficacy 54
Teacher content knowledge and teacher self-efficacy 55
Teacher self-efficacy to teach statistics 55
Chapter 3: Methodology 58
Research Questions 58
Design of the Study 58
Case study 58
Two phases 59
Data Sources and Analysis 62
LOCUS and SETS. 63
Classroom instruction videos. 64
Interview 65

Introduction of this Professional Development ... 66
Schedule of this Professional Development ... 68
The first day. ... 69
The second day. ... 70
The third day. ... 70
The fourth day. ... 70
The fifth day and sixth day. ... 70
Chapter 4: Results ... 71
Section I: Data Analysis of PD Participant Teachers' Knowledge and Self-efficacy .. 71
Analysis of participants' content knowledge change (LOCUS). ... 71
Paired samples t-test on Participants' Self-Efficacy Survey (SETS). ... 72
Correlation between teachers' content knowledge and self-efficacy. ... 75
Three types of cases ... 75
Section II: Three Cases ... 76
Case one: Tina ... 76
Case two: Donna ... 85
Case three: Cathy ... 93
Summary of three cases. ... 98
Section III: NPD math teachers' feedback ... 100
LOCUS test and SETS survey ... 100
NPD teachers' feedback on BLAST ... 103
Chapter 5: Discussion ... 106
Summary of Research Question One and Discussions ... 106
Content knowledge and pedagogical knowledge change ... 106
Teacher self-efficacy change ... 108
Summary of Research Question Two and Discussions ... 111
General Instruction. ... 111
Enact PD activities in classroom. ... 112
Summary of Research Question Three and Discussions ... 115
Implications ... 117
Limitations of this Study ... 118
Recommendations for Future Research ... 118
Appendix A: Activity One ... 120
Appendix B: Activity Two ... 125
Appendix C: Activity Three and Four ... 138
Appendix D: LOCUS ... 144
Appendix E: SETS ... 145
Appendix F: Interview ... 146
References ... 147
Curriculum Vitae ... 161

List of Tables

Table 1. Generation of PSPD Program 4
Table 2. Teachers' Change after PD Training 60
Table 3. Teacher Background and Teaching Experience Information 61
Table 4. Data Table 63
Table 5. SETS Items: GAISE Level A 73
Table 6. SETS Items: GAISE Level B 74
Table 7. Teachers' Change after PD Training 76
Table 8. Teacher Background and Teaching Experience Information 76
Table 9. MQI Instrument Coding System. 77
Table 10. Coding Result of Richness of the Mathematics 78
Table 11. Tina's Instruction Video Analysis Result 79
Table 12. Interview Tina 84
Table 13. Donna's Instruction Video Analysis Result 85
Table 14. Interview Donna 92
Table 15. Interview Cathy 97
Table 16. Cross Case Data Presentation 99
Table 17. SETS Items: Level A (α=0.05, one-tail) 101
Table 18. SETS Items: Level B (α=0.05, one-tail) 102
Table 19. NPD teachers' feedback on BLAST 104
Table 20. Teachers' change and factors influence self-efficacy 108

List of Figures

Figure 1. Shulman's Categories of Teacher Knowledge. 18
Figure 2. Mathematical Knowledge for Teaching 19
Figure 3. Statistical Knowledge for Teaching (SKT) (Groth, 2013). 21
Figure 4. Pottery Urn Problem (Bar-Hillel, 1973). 25
Figure 5. Football Problem.(Nisbett, Krante, Jepson, & Kunda, 1983, p.354) 27
Figure 6. Mean as a Balance Point.. 34
Figure 7. Mean as the Balance Point of the Probability Density Curve (Sukta, 2014). ... 34
Figure 8. Coin Tossing Problem. (Green,1991, p.321).. 38
Figure 9. Random Sequence Problem. Batanero, Gomez, Gea, & Contreras (2014). 42
Figure 10. Relationship among Population, Sample, and Statistical Inference. 48
Figure 11. Sampling Distribution (Chance & Rossman, 2006, p.3). 49
Figure 12. Sample Mean and Sampling Distribution.. 51
Figure 13. Probability of Event A... 80
Figure 14. Worksheet Utilized in Tina's Teaching Episode 81
Figure 15. Terminology Worksheet used in PD. 83
Figure 16. The Maze Game.. 89
Figure 17. Data Tables in Donna's Class.. 90
Figure 18. Box-plots Made by Team A 94
Figure 19. Content Knowledge and Pedagogical Knowledge Change 108
Figure 20. Teaching Practice Lead to Self-efficacy.. 109
Figure 21. Mastery in Content Knowledge and Pedagogies Lead to Self-efficacy 110
Figure 22. Deep Understanding of CK and PCK Lead to Self-efficacy......................... 111
Figure 23. Factors that Influence PD Components Enactment....................................... 115
Figure 24. NPD Teachers' Decisions on Utilizing BLAST Materials............................ 116
Figure 25. NPD Teachers' Enactment of BLAST Materials 117
Figure 26. Population, Sample, Population mean, Sample mean. 131
Figure 27. Means from 50 Samples. N=268. n=10. (Franklin & Kader, 2010)............. 133
Figure 28. Distribution of Sample Means 135

Chapter 1: Introduction

Statement of the Problem

Statistics and probability have been playing an increasingly important role in secondary school mathematics since 2010 when Common Core State Standards were published. However, reports revealed that secondary school mathematics teachers were unprepared for teaching statistics and probability (Lee & Hollebrands, 2008; Madden, 2008; Maker & Fielding-Wells, 2011). Therefore, professional development programs on statistics are needed (Shaughnessy, 2007; Franklin et al., 2015, p. 25). Studies showed that students' performance is related to teachers' knowledge and self-efficacy in teaching (Ashton & Webb, 1986; Ross, 1992; Hill, Rowan, & Ball, 2005). However, in the United States, few studies explore middle grade level statistics professional development (PD) participants' change in self-efficacy in teaching statistics or the relationship between teachers' content knowledge and self-efficacy in teaching statistics (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2015).

According to the sequence of *quality professional development---teacher learning---improved student learning* framework (Loucks-Horsley & Matsumoto, 1999), it is essential to make sure that PD improved teachers' pedagogies and classroom practice; however, few studies explore teachers' classroom instruction after teachers attended statistics professional development programs (e.g. Brendefur, Espinoza, & Pfiester, 2006; Haller, 1997). In addition, as an investment in teacher quality, professional development programs naturally expect a large-scope positive influence on teachers and schools, not limited to PD-participants. However, it is rare to see studies that investigated PD programs' influence on non-PD participant teachers. Instead, non-PD teachers were often used as a comparison to show PD effectiveness (e.g. Nishimura, 2014). This study aims at filling the above research gaps in statistics and probability

professional development. An explanation of details of the professional development program and the study follow in the next paragraphs.

Introduction of PSPD Program

A professional development (PD) program on statistics and probability played an essential role in this study. In this paper, this program was signified as PSPD (Probability and Statistics Professional Development). The researcher contributed to the design and delivery of the PSPD activities and discussions. This PD started in January 2016, and was completed on June 23, 2016. Twenty-nine in-service mathematics teachers (Grade 6 to Geometry) from a large urban school district in the southwest of the United States originally participated in this PD beginning in January 2016, and twenty-one participants remained to the end.

PSPD had two parts. In the first part, in-service mathematics teachers engaged in Common Core State Standards-Mathematics (CCSSM) aligned activity-based content and pedagogy training. In the second part, participants developed teaching materials, including lesson plans and assessments, which were integrated into a publicly accessible online teaching resource website called *Bring Learning and Standards Together* (BLAST). The BLAST website was treated as the 2nd generation of PSPD. The participants developed this content through collaboratively designing lesson plans. Thirty-three lesson plans were developed by twenty-one participants and uploaded online. In the end, the PSPD program distributed NCTM books *Developing Essential Understanding of Statistics (Grades 6-8 and Grades 9-12)* to participants as a PD close.

During this PD, pre-and post-tests of teachers' content knowledge in statistics and probability (LOCUS test) and pre-and post-survey of teachers' self-efficacy in teaching statistics and probability (SETS survey) were completed. Participants' performance in the LOCUS test

and reports in the SETS survey were used for selecting classroom instruction observation subjects in this research. Details of the rationality in selecting these three classroom observation subjects are available in Chapter Three: Methodology.

Introduction of this Study

This study was a follow-up exploration of the PSPD described above. This study aimed to explore: (a), the changes in secondary mathematics teachers' content knowledge, pedagogical knowledge, classroom instruction, and self-efficacy in teaching statistics and probability after participating in a six-month professional development (PD) program; (b), the factors that facilitated these changes; and (c), in what ways non-PD secondary mathematics teachers incorporated BLAST (website: *Bring Learning and Standards Together*) into their teaching.

To accomplish these explorations, this study adopted the case study research method. Two phases were designed. The first phase was called *case selection,* aimed at selecting representative participants to follow up. In the first phase, the researcher analyzed all PD participants' LOCUS test performance and SETS survey results, generated three types of teacher change, and then selected three teachers, one from each type to follow up. The second phase was called *case collection,* aimed at collecting data from each case selected. In the second phase, the researcher video-recorded these three teachers' classroom instructions to explore their teaching. The teachers were also interviewed about their feedback on the PD and how they prepared themselves to teach statistics and probability. The teaching resource website BLAST, which was generated by PD participants, was the 2^{nd} generation of this PD program (see table 1). In what ways NPD secondary mathematics teachers incorporate BLAST content (website: *Bring Learning and Standards Together*) into their teaching was explored. NPD teachers were mathematics teachers who did not participate this PD program. Most of the NPD teachers in this

study were college students and graduate students at a public university in the southwest part of the United States.

Table 1. *Generation of PSPD Program*

Generation of this PD program	
The 1st generation	Probability and Statistics Professional Development (PSPD) Program
The 2nd generation	The teaching resource website: Bring Learning and Standards Together (BLAST)

Research Questions

1. To what extent did components of the PD program change participants' content knowledge, pedagogical knowledge, and self-efficacy in teaching statistics and probability?
2. To what extent were the components of the PD program enacted in the classrooms at school?
3. In what ways did NPD secondary mathematics teachers incorporate BLAST (website: Bring Learning and Standards Together) into their teaching?

Importance of Probability and Statistics

Intensive data analysis and probability predictions have become increasingly essential in refining business, health care, science, public policy and other areas (Schenker, Davidian, & Rodriguez, 2013). In 2012, the federal government created the Big Data Research and Development Initiative, which received 200 million dollars in funding to help access and analyze intensive digital data to "accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning" (Kalil, 2012, pp.2). Citizens also

need probability and statistics knowledge to critically review numerical information published by social media critically to avoid being misled **(**Li, 2013)**.**

In regard to educational recommendations, the National Council of Teachers of Mathematics (NCTM) and the American Statistical Association (ASA) published standards and guidelines to stress the inclusion of and the requirements of statistics and probability in K-12 teaching and learning (NCTM, 2000; Franklin et al., 2007). Probability and statistics are also an essential part of Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) in which representing and interpreting data are highly emphasized (Cohen, 2012). Overall, probability and statistics are necessary knowledge for people to be wise citizens, make smart decisions, and to engage meaningfully in society.

In-service Teachers' Knowledge of Statistics and Probability

Probability and statistics are playing increasingly important roles within and outside of K-12 education. However, "practicing middle-school teachers are in need of professional development (on statistics and probability)" (Franklin et al., 2015, p. 25). Research showed that mathematics teachers were lacking essential content knowledge in teaching probability and statistics. For example, teachers were unfamiliar with the functions of randomness, sample size, distribution, sampling procedure, inference between sample and population, data distribution, measures of association, and statistical investigations (Ives, 2009; Swenson, 1998; Stohl, 2005; Lee & Hollebrands, 2008; Madden, 2008; Maker & Fielding-Wells, 2011; Batanero, Estepa, & Godino, 1997). Franklin et al. (2007) provided possible explanations for teachers' lack of statistical knowledge in *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) report:

> Statistics, however, is a relatively new subject for many teachers, who have not had an opportunity to develop sound knowledge of the principles and concepts underlying the practices of data analysis that they now are called upon to teach. These teachers do not clearly understand the difference between statistics and mathematics. They do not see the statistics curriculum for grades pre-K–12 as a cohesive and coherent curriculum strand. These teachers may not see how the overall statistics curriculum provides a developmental sequence of learning experiences. (p.5)

Another report, the *Statistical Education of Teachers (SET),* indicates that many in-service mathematics teachers did not get enough training in statistics during teacher preparation programs in college (Franklin et al., 2015). Generally speaking, mathematics teachers are lacking statistics and probability knowledge and teachers are not well prepared to teach statistics and probability in secondary schools.

In addition, because of mathematics teacher shortage, a significant number of in-service teachers in CCSD are from Alternative Routes to Licensure program (ARL). Teachers from these ARL programs did not get a professional training in teaching secondary school mathematics in college. Sufficient data is not available about these teachers' content knowledge and self-efficacy in teaching statistics yet. This study will use content knowledge test (LOCUS) and self-efficacy survey (SETS) to explore ARL secondary mathematics teachers' content knowledge proficiency and self-efficacy in teaching.

Pre-service Teachers' Knowledge of Statistics and Probability

Research also showed that pre-service mathematics teachers are lacking essential stochastic content knowledge. Carter (2005) surveyed 210 pre-service K-8 teachers' content knowledge in statistics and probability in a southern public university. Carter found that the

majority of these pre-service teachers were lacking content knowledge in randomness, Law of Large Numbers, event conjunction, central tendency, and data variability. These teachers also hold typical misunderstandings of probability, such as representativeness heuristic. It is noteworthy that many participants just finished taking one statistics course or one course that had statistics content involved, named Statistical Methods, Educational Statistics, Integrated Mathematics, or Problem Solving (Carter, 2005, p.20).

Ives (2011) interviewed three pre-service high school mathematics teachers to explore their understanding of random and probability. She found that all participants had difficulties in defining random in words, indicating that none of these pre-service teachers had a clear conceptual understanding of random. After taking an education course that had probability and statistics as a unit, all three teachers were able to provide a clearer definition of random, showing that their content knowledge improved after training. However, their self-efficacy in defining random remained low.

Carnell (1997) studied thirteen pre-service secondary grades mathematics teachers' conceptual understanding of conditional probability, finding that these pre-service teachers utilized independence properties improperly and applied probability calculation procedures in incorrect situations. Also, research on pre-service elementary school teachers showed that pre-service teachers were able to compute mean by following the athematic procedure; however, many of them did not understand mean as a balancing point of data, which is required by CCSS (Jacobbe, 2012; Leavy & O’Laughlin, 2006). In sum, prior research showed that pre-service teachers are lacking essential stochastic content knowledge also. Targeted training in statistics and probability for pre-service teachers is necessary.

Secondary School Statistics and Probability Professional Development

Secondary school mathematics teachers need suitable statistics and probability professional development programs (Franklin et al., 2015; Shaughnessy, 2007). Prior synthesis professional development programs carried out in the United States did not have an emphasis on middle school statistics and probability, such as *Middle Grades Teacher Enhancement Project* (Haller, 1997) and *Common Core High School Mathematics Leadership: Transforming Teachers' Content Knowledge and Leadership Skills for a New Era* (CCHSML) (Steele and McLeod, 2015). In these two projects, statistics and probability played a partial role. Different from these two programs, *Quantitative Literacy Project* (Scheaffer, 1988) is one public accessible synthesis professional development program that was absorbed in training of statistics and probability in middle school and high school, including data analysis (such as graphical displays and numerical summaries), probability (such as relative frequency, tree diagram, and Venn diagrams), simulation, and sampling (Scheaffer, 1990, p.47). However, this project was conducted twenty-eight years ago (in 1988) before CCSS (published in 2010) was utilized.

Additionally, secondary school mathematics teacher professional development programs that were tailored for statistics and probability education commonly covered specific topics with a narrow focus, such as technology applications (Hummerman & Rubin, 2004; Lee Hollebrands, 2008), comparing two distributions (Makar & Confrey, 2002), statistical data analysis (McClain, 2002), statistical inference (Lee & Mojica, 2007), variability (Hummerman & Rubin, 2004), mean (Bremigan, 2003), etc. Besides the narrow focus, it is noteworthy that the PD programs above were also conducted before CCSSM was published in 2010. As Heck, Weiss, and Thomas (2011) pointed out, "the current mathematics teaching work force will need opportunities to develop the knowledge and skills to enact CCSSM-aligned curriculum materials" (p. C-8), and

research in professional development programs aligned with CCSSM would provide “a fuller picture of the influence of CCSSM, as well as help the field understand the conditions under which these standards are and are not leading to the desired outcomes” (p. 25).

In brief, there is a need of professional development programs that support in-service secondary school teachers to teach CCSSM aligned statistics and probability content (Conference Board of the Mathematical Sciences, 2012, p.49). As METII Report (Conference Board of the Mathematical Sciences, 2012) pointed out:

> Although professional development experiences for middle grades teachers may take a variety of forms, the central focus should be providing opportunities to deepen and strengthen mathematical knowledge in the domains of the CCSS. Many current teachers prepared before the era of the Common Core State Standards will need opportunities to study and learn mathematics and statistics they have not previously been taught. (p. 49)

Based on PD theories, it is essential to involve teachers’ need and voice in PD designs (Sztajn, 2011). Most mathematics educational PD programs did not engage teachers in PD designing efficiently. For example, the most common teacher-involvement approach was to address technologies to be used in PD (Driskell, Bush, Ronau, Pugalee, & Rakes, 2016). To fill this gap and to adequately reflect and be well equipped to meet the needs of teachers, this study constructed a team that had two advanced placement (AP) statistics teachers and two mathematics education researchers as chief designers. Meanwhile, CCSSM aligned activities were carried out directly in this study so that teachers could go through teaching and learning pedagogies which constructed useful professional development experiences (Garet, Poter, Desimone, Birman, & Yoon, 2001; Lee, 2005).

After PD training, it is meaningful and important to explore the influence of PD, not only for PD evaluation but also for PD reflection and further research (Loucks-Horsely, Stiles, Mundry, Love, & Hewson, 2010). Mullins, Lepicki, and Glandon (2010) generalized four levels of PD program evaluation as follows: Level 1- Satisfaction (such as whether participants like the PD content or not). Level 2-Learning (such as the knowledge and skills that participants learned in the PD). Level 3- Behavior (such as participants' knowledge and skills learned in PD as applied to their classroom instruction). And Level 4 - Improvement (such as change in students' performance).

To examine the influence of PD programs, a few statistics and probability PD programs utilized pre-post-test, presentations, clinical interviews (Makar & Confrey, 2002), and survey of PD participants (Carbone, 1998) to examine participants' change. One PD program surveyed participants' students to evaluate PD project's outcome (Garfield & Ahlgren, 1994). Carbone (1998) and Edward (2008) used participants' self-exposure and self-evaluation to examine participants' classroom teaching behavior. Compared with self-reporting and other tools, classroom observation "occurs less frequently" (Mullins, Lepicki, & Glandon, 2010, p.15, line12) in PD program evaluation. In statistics and probability PD programs, few PD researchers conducted follow-up classroom observations to explore the influence of PD. For instance, Haller (1997) carried out classroom observation of four teachers after PD; however, her goals were to explore probability instruction only, not including statistics instruction. A literature review of high quality general mathematics PD programs found that, among twenty-seven PD studies, only one program, *Developing Mathematical Thinking Project* (Brendefur, Espinoza, & Pfiester, 2006), had probability and statistics in elementary school as a partial content and conducted classroom observations.

In sum, various methods were carried out to conduct PD program evaluation; self-reporting was a popular method to explore participants' change in teaching behavior (e.g. Carbone, 1998; Edward, 2008). However, Mullins, Lepicki, and Glandon (2010) pointed out that, although "self-reporting evaluations of behavior change are a legitimate and convenient method for conducting level 3 evaluations, a more valid method would be direct observation. Observations could be conducted by either program administrators or the training providers via onsite visits" (p. 15). Classroom observations will be conducted to fully explore participants' change in teaching behavior by the researcher in this study.

Teachers' Self-Efficacy in Teaching Statistics and Probability

Self-efficacy beliefs are "thoughts or ideas people hold about their abilities to perform those tasks necessary to achieve a desired outcome" (Hall & Vance, 2010, p.2). Self-efficacy studies have been conducted in various domains, including, but not limited to, mathematics self-efficacy (Hackett & Betz, 1989), statistics self-efficacy (Finney & Schraw, 2003), teachers' self-efficacy in teaching mathematics (Showalter, 2005), teachers' self-efficacy in teaching statistics (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2015), and developing teachers' self-efficacy through professional development (Wolf, Foster, & Birkenholz, 2010; Beauchamp, Klassen, Parsons, Durksen, & Taylor, 2014; Tschannen-Moran & Peggy McMaster, 2009).

Consistent with the definition of self-efficacy, *mathematics or statistics self-efficacy* refers to an individual's confidence in doing mathematics or statistics. The terminology *teachers' self-efficacy* (also known as *teacher efficacy* or *teacher self-efficacy*) refers to a self-judgment of a teacher's "capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated" (Tschannen-Moran, & Hoy, 2001, p.783). In other words, teachers' self-efficacy is "teachers' confidence in their ability to

promote students' learning" (Hoy, 2000, p.2). Prior studies suggest that teachers' self-efficacy is strongly related to students' achievement (Ashton & Webb, 1986; Ross, 1992) and students' motivation (Midgley, Feldlaufer, & Eccles, 1989), showing that it is meaningful to improve teachers' self-efficacy for the purpose of student outcomes. In the meantime, studies on teachers' self-efficacy in teaching mathematics indicates that mathematics anxiety, teacher content knowledge and pedagogical knowledge are highly related to teachers' self-efficacy (Gresham, 2008; Huinker & Madison, 1997; Swars, 2005; Iyer & Wang, 2013). However, there is no evidence to show this relationship between knowledge and belief in the domain of teachers' self-efficacy in teaching statistics (Harrell-Williams et al., 2015). There are two research gaps: (1), studies examining teachers' self-efficacy in teaching statistics are rare. In Australia, two studies interviewed or surveyed elementary/middle school teachers to explore teachers' confidence in teaching statistics (Begg & Edwards, 1999; Watson, 2001). However, these studies did not use a validity-proven systematic measurement. In addition, this type of research is uncommon in the United States except for a sole study conducted by Harrell-Williams et al. (Harrell-Williams et al., 2015); and (2), there are no available studies that show the relationship between teachers' knowledge and belief in teaching middle-school statistics (Harrell-Williams et al., 2015). To fill the first gap, Harrell-Williams et al. designed and utilized the Self-Efficacy to Teach Statistics (SETS) instrument to measure preservice middle-school mathematics teachers' confidence in teaching statistics in 2015. However, studies to examine the relationship between teachers' content knowledge and self-efficacy are still in need (Harrell-Williams et al., 2015). This study will address this deficiency in the literature by examining teachers' content knowledge and self-efficacy using validity-tested measurements, *Self-Efficacy to Teach Statistics* (SETS) and *Levels of Conceptual Understanding in Statistics Evaluation* (LOCUS).

Significance of this Study

First, as previously discussed, there is a research gap in PD studies that are aligned with CCSSM with specific emphasis on secondary statistics and probability (Rossman & Tabor, 2014, p.7; Conference Board of the Mathematical Sciences, 2012, p.49). This study will specifically focus on investigating, describing, and exploring effects of CCSSM Grade 6 to geometry statistics and probability professional development content training as follows: (1), on PD participants through their use, reflection on, and implementation of PD activities; and (2), on NPD teachers through their use, reflection on, and implementation of online materials generated through the PD project. This study will contribute to the field by exploring the factors that facilitate teachers' change in knowledge, self-efficacy, and instruction, as well as provide hypotheses of what aspects of PD experiences should be considered for future PD designs.

Second, there is a need of research on examining the relationship between middle school mathematics teachers' knowledge and self-efficacy in teaching statistics and probability (Harrell-Williams et al, 2015). This study examined teachers' content knowledge by utilizing the Levels of Conceptual Understanding in Statistics Test (LOCUS) (Jacobbe, Case, Whitaker, & Foti, 2014); this study also investigated teachers' self-efficacy via *Self-Efficacy to Teach Statistics* instrument (SETS) (Harrell-Williams et al., 2015). The validities of both instruments have been examined (Jacobbe, Case, Whitaker, & Foti, 2014; Harrell et al. 2009; Harrell-Williams et al., 2014). The relationship between teacher knowledge and self-efficacy in teaching statistics and probability was examined.

In general, mathematics teachers are lacking statistics and probability content knowledge and pedagogical knowledge (Ives, 2009; Swenson, 1998; Stohl, 2005; Madden, 2008). PD programs that aim at improving secondary school mathematics teachers' statistics and

probability knowledge are needed (American Statistical Association, 2007; SET, 2015). This study has the potential to contribute to mathematics teacher education at the practice level by exploring the factors that contribute to an effective statistics and probability PD program. This study also has potential to contribute to the mathematics education research by filling research gaps, such as exploring the relationship between teachers' statistical knowledge and self-efficacy in teaching statistics.

Chapter 2: Literature Review

Introduction

Several bodies of research literature were pertinent with building a foundation for this study: First, theories of mathematical knowledge for teaching (MKT) (Ball, Thames, & Phelps, 2008), statistical knowledge for teaching (SKT) (Groth, 2013), and theory of probabilistic intuitions improvement (Fischebein,1975). Second, targeted training that could facilitate improving people's stochastic reasoning by defeating inherent human judgment biases. Third, reviewing important concepts and typical misunderstandings in secondary probability and statistics content. Fourth, the relationship between content knowledge and self-efficacy. Please see literatures below for details.

Theoretical Framework

Two bodies of theories constructed the theoretical framework in this study: probabilistic intuitions improvement and teacher content knowledge. Theory of probabilistic intuitions improvement showed formal content knowledge training can help people develop probabilistic intuitions and hence make accurate judgments under uncertainty. In other words, knowledge training can lead to knowledge understanding. Theories of teacher content knowledge show that sufficient content knowledge of statistics and probability is essential for teacher quality and necessary for student achievement. In brief, mathematics teachers need professional subject content knowledge, which can be well developed by teacher content knowledge training.

Theory of probabilistic intuitions improvement. Piaget (1958) classified students' cognitive development into four stages: Sensorimotor (0-2 years old); preoperational (4-7 years old); concrete operational (8-11 years old); and formal operational (beyond 11 years old). Piaget and Inhelder (1951/1975) conducted studies on the development of the concept of chance in

children, particularly with respect to Piaget's cognitive development stages theory. They concluded that children could not develop the concept of probability as a group of formal ideas until the formal operational stage; when children reached the formal operational stage, children were able to conceive the nature of random and reason probabilistically when facing randomizing devices. For instance, in the formal operational stage, children could list all possible outcomes of a spinner, display the potential combinations of the sample space, and then use number comparisons and proportional reasoning to make probabilistic judgments. They also found that children in the formal operational stage have an intuitive understanding of the Law of Large Numbers. For instance, with increasing number of trials, children could recognize that the distribution of the results had a pattern (Jones &Thornton, 2006; Piaget & Inhelder, 1951,1975).

Different from Piaget, psychologist Fischbein (1975) combined probabilistic thinking development and probability instruction together. He studied how children responded in instructional settings and generated his theory of intuition, which "plays an essential part in the domain of probability, perhaps more conspicuously and strikingly than it does in other domains of mathematics" (Fischbein, 1975, p. 5). According to Fischbein, intuition is a cognitive belief that is "spontaneous, global, and self-evident to the believers" (p.117). For example, people accept the following claim intuitively without any doubt or proof: through a point outside a line one and only one parallel may be drawn to that line (Fischbein & Gazit, 1984).

Fischbein (1975) argued that probabilistic thinking could be trained because intuitions are adaptable, and intuitions could be influenced by systematic instruction. To explain his theory, Fischbein categorized two levels of intuitions: primary and secondary intuitions. He declared that primary intuition is based on the private experience of an individual with no systematic instruction needed. For instance, when a child was asked which color was most likely to come up

on a spinner, the child may choose a color just because he liked the color. Secondary intuition is formed mainly during instruction at school (Fischebein, 1975; Fischbein, Barbat, and Minzat, 1971). For instance, children aged 10 to 14 were found initially to be weak in intuitively estimating the number of permutations of 5 objects. After being trained in tree diagram, these children showed significant improvement in the same task, meaning that the techniques children learned during instruction can be transferred successfully (Fischebein, 1975; Jones & Thornton, 2006). Based on analysis of several teaching experiments, Fischbein (1975) concluded that:

(a) Even at the level of formal operations, combinatorial techniques are not spontaneously acquired. Instruction is necessary. (b) Even at the level of concrete operations, it is possible to induce children to assimilate combinatorial techniques quite readily with figurative aids. (p.115)

Fischbein concluded that the probabilistic intuitions cannot “benefit sufficiently from the development of operational schemas of thought” naturally (Fischbein, 1975, p. 73), showing the importance of external intervention to improve probabilistic intuitions. These psychological findings showed that people need proper training to build appropriate intuitions about uncertainty. These findings not only work for children but also work for adults. Research showed that adults, including mathematics teachers, have three types of improper probabilistic and statistical reasoning: representativeness heuristics, availability heuristics, adjustment and anchoring (Kahneman & Tversky, 1974). Studies showed that these improper reasoning are related to insufficient statistics and probability content knowledge, hence related to insufficient formal training (e.g. Bar-Hillel, 1979; Reaburn, 2008; Lehman & Nisbett, 1990; Fischbein & Gazit,1984). Therefore, Fischbein’s theory showed the necessity of formal training to improve teachers’ intuitions and judgement under uncertainty.

Theory of teacher content knowledge. Teacher's knowledge has been studied for years since Shulman (1987) designed a framework of teacher knowledge (Figure 1). Teacher knowledge is highly recommended for teacher quality and hence influences student achievement (Ma, 1999; Hill, Rowan, & Ball, 2005). In Shulman's framework, teacher content knowledge plays a role as a foundation because teacher content knowledge includes knowledge of the discipline and its systematizing structures (Grossman, Wilson, & Shulman, 1989), and teachers are not able to teach what they do not understand (Ball, Thames, & Phelps, 2008). Subsequently, teacher content knowledge has been identified as an essential component of effective teaching and learning mathematics (Ball, Thames, and Phelps, 2008).

Shulman's Categories of the Knowledge Base

- Content knowledge;
- General pedagogical knowledge, with special reference to those broad principles and strategies of classroom management and organization that appear to transcend subject matter;
- Curriculum knowledge, with particular grasp of the materials and programs that serve as "tools of the trade" for teachers;
- Pedagogical content knowledge, that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding;
- Knowledge of learners and their characteristics;
- Knowledge of educational contexts, ranging from the workings of the group or classroom, the governance and financing of school districts, to the character of communities and cultures; and
- Knowledge of educational ends, purposes, and values, and their philosophical and historical grounds.

Figure 1. Shulman's Categories of Teacher Knowledge.

Further, Ball, Thames, and Phelps (2008) developed the Mathematical Knowledge for Teaching (MKT) framework in which teacher content knowledge is comprised of three components: common content knowledge, horizon content knowledge, and specialized content

knowledge (Figure 2). Common content knowledge means the mathematical knowledge "of a kind used in a wide variety of settings—in other words, not unique to teaching" (Ball, Thames, & Phelps, 2008, p.399), such as how to do three-digit subtraction. Horizon content knowledge, which researchers have not determined that it belongs to subject matter knowledge yet, refers to teachers' knowledge of mathematics curriculum scope and sequence. For example, a teacher who teaches absolute deviation in grade six should be aware of and build a base for standard deviation in high school. Specialized content knowledge in mathematics is "mathematical knowledge beyond that expected of any well-educated adult but not yet requiring knowledge of students" (Ball, Thames, & Phelps, 2008, p.9). For instance, a well-trained mathematics teacher should know how to distinguish the takeaway model and the comparison model of subtraction while a non-teaching-professional does not need to understand this method (Ball, Thames, & Phelps, 2008).

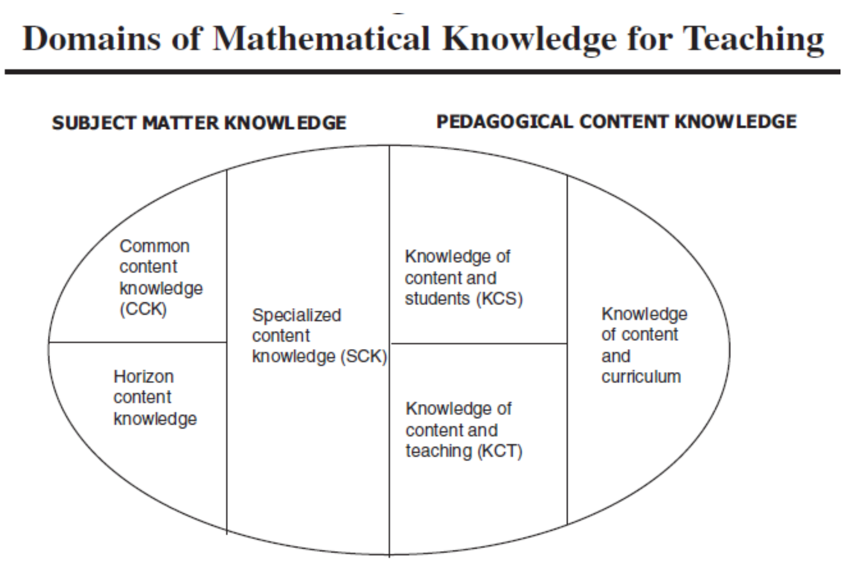


Figure 2. Mathematical Knowledge for Teaching

Although various knowledge frameworks for mathematics teachers were available, researchers argued that statistics is different from mathematical meaning (Franklin, et al., 2007). For example, in mathematics, context fogs mathematical structure; however, in statistics, context supplies. Therefore, specified knowledge frameworks for statistics teaching and learning are considered purposefully tailored for statistics education (Burgess, 2006).

Based on the framework of Mathematical Knowledge for Teaching (MKT), Groth (2013) developed the framework of Statistical Knowledge for Teaching (SKT) (Figure 3). In this SKT framework, Groth delineated common content knowledge in statistics, horizon knowledge in statistics and other knowledge domains. It is noteworthy that sufficient statistical content knowledge plays the role as a foundation (Batanero & Diza, 2010; Franklin et al., 2015). Different from the MKT framework, which is general and broad, Statistical Knowledge for Teaching (SKT) involves explicit pieces of statistical knowledge. SKT utilizes *Key Developmental Understandings* (KDT) to represent the "cognitive landmarks in learning subject matter…(KDUs) involve significant shifts in students' thinking that occur through reflection on a series of conceptually similar tasks" (Groth, 2008, p.127). For instance, in the domain of Common Content Knowledge in Statistics, "conceiving of theoretical probability as an anchor for predicting long-term behavior is a statistical KDU" (Groth, 2008, p.129). In the domain of Horizon Knowledge in Statistics, "conceiving of typical deviation (mean absolute deviation and standard deviation) from the mean as a measure of spread" (Groth, 2008, p.132) is one sample KDU. The correct understanding of standard deviation is built on the correct understanding of mean absolute deviation.

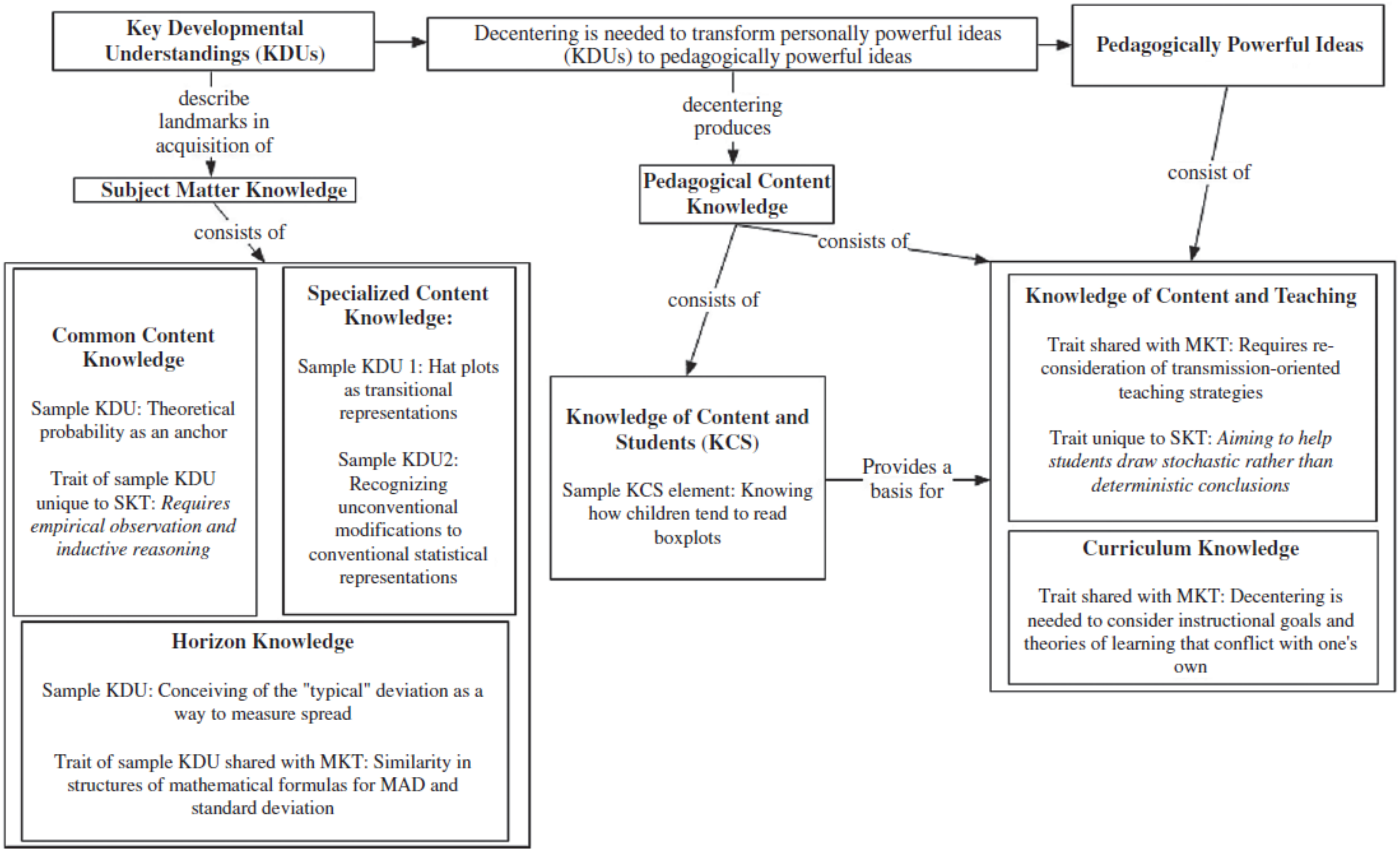


Figure 3. Statistical Knowledge for Teaching (SKT) (Groth, 2013).

In sum, theories of Mathematical Knowledge for Teaching (MKT) and Statistical Knowledge for Teaching (SKT) show the importance of content knowledge in statistics and probability for teachers' success in instruction. Theories of probabilistic intuitions improvement (Fischbein, 1975) indicate that teachers are regular human beings who have certain stages in stochastic reasoning development which can be upgraded by formal training. These two parts of theories build up the theoretical framework for this study.

Utilizing Content Training to Improve Stochastic Reasoning

Literature review in this section introduces three typical improper stochastic reasoning processes (*representativeness heuristics, availability heuristics, adjustment and anchoring*), and suggests that training in normative probability and statistics knowledge can improve people's stochastic thinking (Gigerenzer,1996; Larrick, 2004; Kirkebøen, 2009). The literature serves this study in two ways: (1) It provides experience in designing activities to help secondary school mathematics teachers overcome inherent human judgment biases, and (2) it provides a framework to examine data that is specific to answering research question No.4, which explores factors that facilitate teachers' change after the professional development phases. In brief, the literature indicates that improper heuristics are caused by insufficient knowledge of sample size, sampling distribution, statistical inference, and the Law of Large Numbers. Consequently, I argue that it is important to examine ways in which tailored content knowledge training might influence teachers' knowledge to facilitate positive adjustment of these improper heuristics.

Three types of improper probabilistic and statistical reasoning. This section examines three typical human judgment biases in psychological research, showing how people misunderstand certain statistical ideas in an improper approach. These misunderstandings can be attributed to insufficient statistical content knowledge (Fischbein, 1975).

***Representativeness heuristics*.** A heuristic is a mental shortcut that enables people who lack sufficient probabilistic and statistical thinking to solve problems or make judgments under uncertainty (Gilovich & Savitsky, 1996; Kahneman & Tversky, 1974). Heuristics are experience-based strategies that can cause serious misconceptions and bias, although sometimes heuristics may also result in acceptable judgements unintentionally. One key heuristic that Kahneman and Tversky generated is representativeness heuristics, which is broadly used not only in

psychological studies (such as decision making) but also in research about probability teaching and learning (Jones & Thornton, 2006).

When people use representativeness heuristics to estimate probability, "the subjective probability of an event, or a sample, is determined by the degree to which it (i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated" (Kahneman & Tversky, 1974, p.25). Representativeness is widely used in everyday life. For example, when you see Person A dressed well and wearing an expensive watch, and Person B wearing jeans and busy texting, you will think person A is more likely to be a CEO because A resembles what a CEO generally looks like. However, people easily made mistakes when they used this strategy to solve problems that involved knowledge of probability and statistics. For example, Kahneman and Tversky (1972) used the following questions to test 92 college students, 88% of who answered incorrectly.

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?
>
> (a) The larger hospital (21 students chose this answer).
>
> (b) The smaller hospital (21 students chose this answer).
>
> (c) About the same (that is, within 5 percent of each other) (53 students chose this answer) (p.1125).

The correct answer is (b); while 58% of the students chose the wrong answer (c). Kahneman and Tversky (1972) argued that people who chose (c) applied representativeness heuristics because they were fooled by the information containing the same statistic, namely that, each hospital recorded the *days on which more than 60 percent of the babies born were boys*. Hence, people who were carrying representativeness heuristics assumed "as the two events are described by the same statistic they will be equally representative of the general population" (Reaburn, 2008, p.145). That is, in people's representativeness heuristics judgment, sample sizes were not considered because sample size was not a statistic and therefore did not represent any characters of the population (Bar-Hillel, 1979). This representativeness heuristic happens widely with K-12 school students (Fischbein & Schnarch, 1997; Watson & Moritz, 2000), college students (Tversky & Kahneman, 1982) and pre-service teachers (Watson, 2000).

***Availability heuristics*.** People who hold availability heuristics judge the likelihood of an event by their personal experience with the event (Tversky & Kahneman, 1974). In other words, "people's probability estimates for an event are based on how easily examples of that event are recalled" (Garfield, 1995, p.28). One type of availability heuristics is called *Biases Due to the Retrievability of Instances* (Tversky & Kahneman, 1974). For example, people are likely to claim the traffic accident rate goes up after being witness to an accident. Similarly, people are likely to estimate the divorce rate in their community by counting the divorce cases of which they are personally aware. *Biases Due to the Retrievability of Instances* can likely be attributed to the consequence of unknowing basic statistical concepts, such as sample, population, and relative frequency. Another type of availability heuristics is called *Illusory Correlation.* People who apply *Illusory Correlation* perceive a nonexistent relationship between variables. For example, Ada holds a belief that people from a small town are very nice. One day, Ada sees a person who

is very nice, and she immediately assumes that the person grew up in a small town although there is no statistical evidence indicating a negative relationship between kindness and city population (Cherry, 2014). This type of availability heuristics is also likely due to the ignorance or misunderstanding of statistical correlation and sample size.

Adjustment and anchoring. People who hold adjustment heuristics make estimates by starting from an initial value (anchoring) and then adjusting the value to the final answer (adjustment). One classic misused adjustment heuristic is called *Biases in the Subjective Probability of Compound Events* (Bar-Hillel, 1973; Kahneman & Tversky, 1974). An example from Bar-Hillel's study (1973), *Pottery Urn Problem*, is presented in Figure 4.

> Pottery Urn Problem
>
> A gamble game provides two pottery urns and uses two gamble rules. Each urn corresponds to a rule. Urn No.1 has 30 marbles inside, among which 3 are red and 27 are white. The rule to play urn No.1 is: pick up one marble and if it is red, the player wins. Urn No. 2 also has 30 marbles, among which 15 are red and 15 are white. The rule to play urn No.2 is: pick up one marble each time with replacement to build a sequence and stop when a white marble is picked up. In the sequence, if there are no less than 4 red marbles, the player wins.
>
> Note: if the player choose Urn No.1, the probability of picking up a red marble is 0.1; if the player chooses Urn No.2, the probability of picking up a red marble each time is 0.5.
>
> Which urn should the player choose to have a better chance to win?

Figure 4. Pottery Urn Problem (Bar-Hillel, 1973).

In the Pottery Urn Problem, by using probability calculations, players can calculate the probability of winning urn No.2 to be $0.5^4 = 0.062$, lower than the probability of winning urn No.1, which is 0.1. However, Bar-Hillel found that people easily overestimated the probability of a compound event, and therefore chose the wrong urn. This adjustment heuristic is likely due to the ignorance or lack of knowledge of statistical content knowledge of compound events.

Utilizing content knowledge training to overcome judgment biases. As previously mentioned, people hold various heuristics and biases when making judgments under uncertainty, most of which are majorly attributed to the failure to know normative statistical content knowledge. Wilson and Brekke (1994) split reasons of human heuristics and biases into two groups: (1) "failure of rule knowledge and application" (p.118) and, (2) mental contamination. Mental contamination is defined as "the process whereby a person has an unwanted response because of mental processing that is unconscious or uncontrollable" (Wilson & Brekke, 1994, p.117). Wilson and Brekke (1994) believed that, in the first case, people made improper judgements because of knowledge shortage, such as the Law of Large Numbers, regression to the mean, causal, and contractual rules. To adjust people's heuristics that are caused by knowledge shortage, psychologists conducted various explorations.

First type of exploration: develop people's accurate reasoning through experience. Nisbett, Krante, Jepson and Kunda (1983) studied people's understanding of *Regression to the Mean* by using a *Football* problem (see Figure 5). In the *Football* problem, the second choice is the statistical reasoning, suggesting that the excellent performances in the tryout were outliers. Nisbett et al. (1983) found out that most of the participants who had athletic team experience in the past chose the statistical reasoning as an explanation, whereas most of the people without such experience chose other answers. Hence, they concluded that "greater expertise in a domain is associated with a greater tendency to reason statistically in that domain" (Nisbett et al., 1983, p.354), supporting the view that people can think statistically after they experienced a significant number of samples. In other words, people can build a probability model themselves based on their experience in that domain and figure out whether an event happened by chance. On the

contrary, a novice in these fields tended to use non-statistical reasoning to explain a certain event.

Football

Harold is the coach for a high school football team. One of his jobs is selecting new members of the varsity team. He says the following of his experience: "Every year we add 10-20 younger boys to the team on the basis of their performance at the try-out practice. Usually the staff and I are extremely excited about the potential of two or three of these kids—one who throws several brilliant passes or another who kicks several field goals from a remarkable distance. Unfortunately, most of these kids turn out to be only somewhat better than the rest." Why do you suppose that the coach usually has to revise downward his opinion of players that he originally thought were brilliant?

Choices: 1. Harold was probably mistaken in his initial opinion. In his eagerness to find new talent, he exaggerates the brilliance of the performances he sees at the try-out. 2. The brilliant performances at try-out are not typical of those boys' general abilities. They probably just made some plays at the try-out that were much better than usual for them. 3. The boys who did so well at try-out probably could coast through the season on their talent alone and don't put out the effort necessary to transform talent into consistently excellent performance. 4. The boys who did so well at try-out may find that their teammates are jealous. They may slack off so as not to arouse envy. 5. The boys who did so well at try-out are likely to be students with other interests. These interests would deflect them from putting all their energies into football.

Figure 5. Football Problem. (Nisbett, Krante, Jepson, & Kunda, 1983, p.354)

Although experience is shown to be helpful in certain situations (such as the Football problem), researchers also point out that experience is not associated with accurate judgment (Hammond, 1996) due to insufficient timely feedback (Kirkebøen, 2009). Experience is often inexact and may even mislead the user (Hogarth, 2001). People need instant, accurate feedback to reveal why their judgments had errors, and to learn how to improve judgements. Formal statistics training can provide people this feedback.

Second type of exploration: develop accurate reasoning via formal statistics training. Mevarech (1983) carried out an experiment on 139 freshmen who majored in education. Both

treatment and control groups were enrolled in an introductory statistics course. The control group was taught under traditional instruction (lecture and discussion); the treatment group was taught differently from the control group with "only respect to the provision of feedback and the implementation of corrective activities" (Mevarech, 1983, p.422). All participants were tested in the end, with questions covering both descriptive and inferential statistics. Mevarech found that the treatment group performed significantly better than the control group. His finding confirmed other researchers' finding that the majority of learners need external support to shift from a lower-level mental process to a higher level mental process (Russell, 1960). Mevarech (1983) also argued that it was the feedback--corrective activities that eradicated college learners' misunderstandings and helped learners improve.

Lehman and Nisbett (1990) did a longitudinal study by measuring 121 freshmen twice in statistical reasoning. The first measurement was in the first semester in their first year of college, and the second measurement was in the second semester in their fourth year of college. These 121 college students were majoring in four different fields: natural science, humanities, social science, and psychology. Learners' performance change showed that four years of formal discipline training in social science and psychology significantly improved learners' statistical reasoning. Lehman and Nisbett (1990) argued that this finding (1) confirmed that reasoning can be taught; (2) showed that formal statistics and probability training in psychology and social science can affect adult learners' statistical and methodological reasoning in everyday life (p.959).

Third type of exploration: strengthen specific statistics elements to improve statistical reasoning. When exploring interventions to improving adults' statistical reasoning, researchers found that specific statistics elements played essential roles in the development of statistical

reasoning. Fong, Krantz, and Nisbett (1986) designed four formal content instruction interventions to improve adults' conceptual understanding and everyday application of *the Law of Large Numbers*. They found that statistical content instruction significantly improved participants' understanding of statistical reasoning systematically and application of *the Law of Large Numbers* in everyday life. In addition, they found that real-world examples could strongly support participants' proper application of *the Law of Large Numbers*. Kirkebøen (2009) argued that people who do not know Bayesian reasoning likely have difficulties in making proper judgments when the problem-solving strategies involve Bayes' formula. Nisbett et al. (1983) found that the ignorance of *sample size* and *the Law of Large Numbers* lead to *representativeness heuristics.* Sedlmeier and Gigerenzer (1997) analyzed several empirical studies and generalized an argument: the misconception of *sample size* and *the Law of Large Numbe*r lead to the incorrect understanding of *the sampling distribution.* Garfield (1995) believed that the misunderstanding of *association* and *relative frequency* results in *availability heuristics.* Bar-Hillel (1973) conducted a study on eighteen high school seniors and fifty-seven college students and found that insufficient knowledge of *compound events* results in *adjustment and anchoring.* As Garfield (1995) pointed out, "It is important to learn some fundamentals of statistics to better understand and evaluate information in the world" (p. 26).

In sum, improper heuristic and biases are barriers for people to make accurate judgments. Formal training in statistics and probability knowledge can help strengthen people's content knowledge and hence improve people's statistical reasoning.

Important Content Knowledge and Training in Secondary Probability and Statistics

Literature review in this section introduces the definition, knowledge gap, misunderstanding, and training of important concepts in secondary school statistics and

probability in CCSSM, from grade six to geometry, including *statistical questions, mean, randomness, sample, sampling, sampling distribution*, and *statistical inference*. The literature contributes to this study in the following ways: (1) part of the literature inspired the researcher in designing activities in this PD. (2) Literatures in this section provide the researcher a knowledge base to analyze teachers' discussions and worksheets in PD, lesson plans and instructions after PD, and interview of teachers after PD. It is common to expect that professional development supports teachers' development. This literature can assist in showing the specific concepts that teachers may find essential to assist their knowledge and self-efficacy change. The following outlines content knowledge relevant to this study from CCSSM grade six to geometry.

Statistical question. The conceptual understanding and application of statistical question are required by CCSSM (in grade six) and GAISE Framework (e.g. distinguish and formulate statistical questions).

Definition of statistical question. A statistical question is a question that can be solved by collecting and analyzing data, within which people would expect variability. In other words, "the formulation of a statistics question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary" (American Statistical Association, 2005, p.11). In the definition, data variability is the key component. For example, "*How many official Brunei citizens are there in Brunei?"* is not a statistical question because the answer will be a deterministic answer and there will not be data variability. The process to answer this question may involve data collection if nobody knows the accurate number, however, this question is still not a statistical question because the data gathered do not vary, as the data are a single number.

Research on teachers' understandings of statistical questions. The ability to identify and compose statistical questions is an essential part of statistical literacy (Rumsey, 2002; American Statistical Association, 2007; SET, 2015). When facing a problem, a problem solver need to "clarify the problem at hand and formulate questions that can be answered with data" (SET, 2015, p.1). When formulating a statistical question, a problem solver needs to use precise language, which needs strict exercise (SET, 2015, p.86). Even statisticians, who have abundant training in statistics and probability content knowledge, may compose improper statistical questions. There is abuse of improper statistical questions among researchers in many different academic areas (Hand, 1994). Although there is not accessible research showing that mathematics teachers also lack the knowledge of statistical questions, reports did show that mathematics teachers often do not receive enough training on formulating statistical questions in teacher education programs (SET, 2015, p.35). Therefore, it is reasonable to deduce that in-service mathematics teachers are teaching or using *statistical questions* without enough preparation. Hence, it is necessary to pay attention to statistical questions training among mathematics teachers.

Training guidelines and examples on formulating statistical questions. Up to this time, published empirical studies on training teachers to formulate statistical questions are rare. Therefore, except the description of Halvorsen's (2010) paper, this section mainly focuses on guidelines and examples on training teachers to understand and formulate statistical questions. Literatures in this section guided the statistical question activity design in the PD program and provided the researcher with the knowledge base to analyze teachers' lesson plans, instruction, and interviews in this study.

Halvorsen (2010) described how introductory statistics courses were taught in Smith College since 1989. Each student in these courses was required to design and implement a statistics project in small groups. With the help from team members and the instructor, students learned how to "develop statistical questions out of broader research questions" (p.1). However, besides the description of the project procedure, not much information or experience on teaching and learning statistical questions was shared. Although not in details, Halvorsen's paper showed an approach to teach statistics questions --- let learners design, implement, and report statistics projects.

In a teacher training book written by Hopfensperger, Jacobbe, Lurie, and Moreno (2012), *Bridging the Gap between Common Core State Standards and Teaching Statistics*, specific standards for statistical questions are listed: "A well-written statistical question refers to a population of interest, a measurement of interest, and anticipates answers that vary" (p. 21). Hopfensperger et al. (2012) designed activities for teachers to distinguish statistical questions from non-statistical questions. For example, "Do plants grow better under colored lights?" is not a statistical question because the population is not clear. Meanwhile, "Do tomato plants grow taller under red light, blue light, or daylight?" is a well-written statistical question (p.25).

In the American Statistical Association's report, *The Statistical Education of Teachers (SET)*, the minimum requirements of middle school teachers' knowledge on statistical questions are "Distinguish between questions that require a statistical investigation and those that do not; translate a 'research' question into a question that can be answered with data and addressed through a statistical investigation" (SET, 2015, p.32). Different scenarios are provided in SET for teaching statistical questions. For example, "Do you think teachers at this school are giving too much homework?" is not a statistical question but a survey question; meanwhile, "How

many hours per week do students at this school spend on homework?” is a proper statistical question (SET, 2015, p. 87).

The activities in Hopfensperger et al.’s book and the statistical questions in SET followed CCSSM and GAISE Framework, highly fitting the goals of mathematics teachers’ professional development in secondary level statistical content knowledge. Therefore, the above activities were adopted and adapted in the activities in this professional development.

The meanings of mean. *Mean* is an important basic concept in statistics and is required by CCSSM from grade six.

Definition of mean. According to GAISE Framework, there are three levels of understanding *mean*. The three levels are outlined as, level A: mean as a fair share (p.30), level B: mean as a balance point (p.41), and level C: mean as a statistic that bridges sample and population (Franklin & Kader, 2010, p.3). The three levels of conceptual understanding of mean are also required by CCSSM (elementary and secondary). The computation of a fair share can be introduced to learners as early as grade three as a context for division. Learners normally have an intuitive concept of equal share and learn the calculation procedure without many conceptual barriers. For example, the problem *Share 10 books equally among 5 students, how many books will each student have?* can be easily solved by $10 \div 5 = 2$.

Mean as a balance point refers to the necessary conceptual understanding of mean as *a center of gravity* (Pitman, 1993, p.162) in data distribution. Data itself has no weight, however, when using manipulatives to represent data distribution, people can find out the mean by balancing the manipulatives on a fulcrum. The balance point is where the center of gravity is; the center of gravity is the mean. Please see figure 6 as an example.

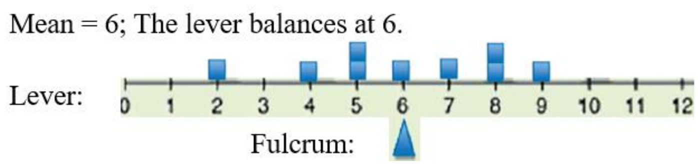


Figure 6. Mean as a Balance Point.

The understanding of *mean as a balance point* is important for advanced statistics concepts (such as probability density) because *mean* is the balance point of the probability density curve (Sukta, 2014). Please see figure 7 as an example (Sukta, 2014).

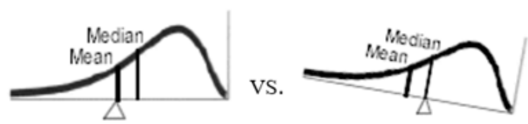


Figure 7. Mean as the Balance Point of the Probability Density Curve (Sukta, 2014).

Knowledge gap on mean. Although it is required to understand mean as a balance point by CCSSM, mathematics teachers are reported to be not well prepared. Leavy and O'Loughlin (2006) did a survey on 263 pre-service elementary teachers and found that only one-fourth of the participants had a conceptual understanding of mean while most of the participants confused procedural and conceptual understanding of mean. Jacobbe (2008) did a case study on three in-service elementary teachers and found that teachers did not possess knowledge of mean and median that was defined by GAISE Framework Level A and B. For example, two teachers were not sure about what the differences between mean and median were. Jacobbe argued that elementary teachers were not trained to teach by following GAISE, and therefore teachers need sustained professional development for catching up. Misconceptions about *mean* are even worse

among regular college students. Pollatsek, Lima and Well (1981) did a survey on 17 college students and found that computing *simple mean* was the only method that participants know. For example, the correct calculation for *A student got a 3.2 GPA in the first two semesters and got a 3.8 GPA in the next three semesters, what is the student's GPA for the five semesters in total?* should be:

$$mean = \frac{3.2\times 2 + 3.8\times 3}{5} = 3.56.$$

However, many participants got

$$mean = \frac{3.2 + 3.8}{2} = 3.5 \ \ or \ \ \frac{3.2 + 3.8}{5} = 1.4.$$

These mistakes showed that participants didn't understand the principles that led to mean calculation, and hence they applied the computation rule in a wrong approach. The above literature shows that there is a gap between what teachers should know about mean (level A and B) and what teachers know, which requires further training on *mean*. The GAISE level C of mean refers to statistical inference which will be discussed in the later sections because statistical inference is related to multiple statistical content knowledge, including, but not limited to, randomness, random sample, random sampling, and sampling distribution.

Training on mean. Franklin and Kader (2010) proposed models for teacher training which followed the GAISE Framework. They suggested several activities for developing teachers' understanding of *mean*. In level A---*mean* as a fair share, they suggested distributions comparison; in level B---*mean* as a balance point, they suggested distributions comparison, sum of the absolute deviations and the *mean* of the absolute deviations; in level C—*mean* as sample statistic that bridges sample and population, they suggested an activity that determined the mean word length.

These activities have also been introduced or used for classroom teaching (Berlin, 1989; Uccellini, 1996; Flores, 2008; Chance and Rossman, 2006), showing that the quality of these activities has been widely recognized although empirical evidence is still needed. Other researchers (Saldanha and Thompson, 2003; Sedlmeier and Gigerenzer, 1997) posed instruction suggestions that can help improve people's understanding of mean in level C: mean as a sample statistic that bridges sample and population, using statistical inference. Considering that statistical inference was complex to teach and learn, these strategies focused on proper training on sample, sampling, the Central Limit Theorem, and statistical inference. Please see following sections for details. TinkerPlots will be used as the data analysis technology in teaching and learning *mean* in this professional development program.

Randomness. Randomness is a starting root of probability and the base of statistical inference. Without randomness, people have no confidence in getting reasonable statistical inferences from the samples or experiments.

Definition of randomness and random. The original definition of randomness is similar to uncertainty, which means that the outcome of a case depends on luck or chance (Batanero, Green, & Serrano, 1998). Despite the academic debates about the official definition of randomness (Steinbring, 1991; Hellman,1978; Ekeland,1988), random phenomena can be understood as the opposite of deterministic phenomena. Random sequences can be comprehended as the opposite of deterministic sequence (Batanero et al., 1998; Batanero & Serrano, 1999), which means random phenomena and random sequences are unpredictable, without a pattern to be forecast in advance. For example, after you flipped a coin nine times, you are still not sure what the 10^{th} will be. However, random sequences done may have a pattern

afterward. For instance, after you flipped a coin ten times, the sequence may be HTHTHTHTHT although this type of random sequence looks rare.

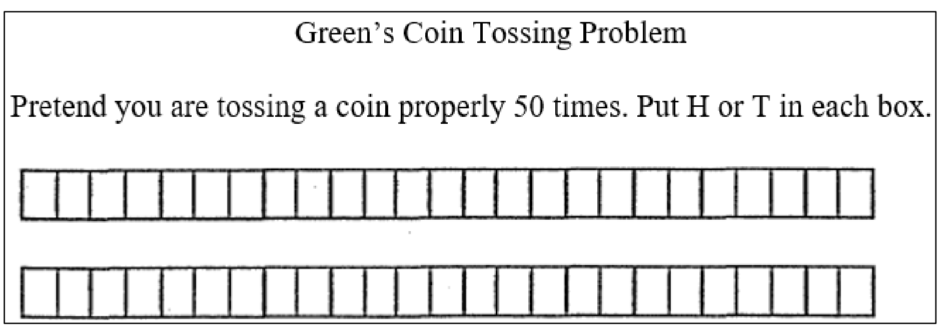
To be more specific, at the secondary education level, researchers Batanero, Green, and Serrano (1998) defined five vital characteristics of random phenomena that students should command:

> (1) In a given situation there is more than one possible result. (2) The actual result which will occur is unpredictable. (3) There is the possibility—at least in the imagination—of repeating the experiment (or observation) many times. (4) The sequence of results obtained through repetition lacks a pattern that the subject could control or predict. (5) In this apparent disorder, a multitude of global regularities can be discovered, the most obvious being the stabilization of the relative frequencies of each possible result. This global regularity is the basis that allows us to study random phenomena using the theory of probability. (p.122).

Let's flip a coin to make the above characteristics concrete. When a fair coin is thrown without cheating, the action has two possible results: head or tail (aligned with characteristic one). The result is unpredictable because each result may show either (aligned with characteristic two). The experiment can be repeated many times by throwing the same coin again and again (aligned with characteristic three). When people throw a fair coin 50 times, the sequence of the results does not have a pattern that the subject can predict or control. For example, if Bob throws a coin for 50 times and gets the first 40 results with a sequence as HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT HT (which looks like a pattern so far), Bob still cannot say that the following two results will definitely be HT because what will happen does not have a pattern that Bob can predict (aligned with characteristic four). On the contrary, if Bob

has the confidence to say so, Bob is cheating by using specific coin flipping skills or the coin was designed to show that pattern unfairly, making the outcome no longer random phenomena (a counter-example that aligns with characteristic four). Although it is impossible to predict the result, by calculating the relative frequencies of head and tail (for example, people can flip the coin for 10,000 times and count the heads to get the relative frequency of getting heads), people are able to generate that head and tail each has an approximate probability of 50 % (aligned with characteristic five). If people are able to understand the experiment of flipping a coin as the explanations above, it means that people truly get the meaning of randomness, random phenomena, and random sequence. However, research on students and teachers showed misconceptions on randomness.

Misconceptions on randomness. Green (1991) used the following problem to test 11 to 14 years old students' understanding of randomness. The problem is named *Green's Coin Tossing Problem* here for convenience (please see figure 8). Results showed that students didn't understand characteristics four and five in Batanero, Green, and Serrano's framework.


Green's Coin Tossing Problem

Pretend you are tossing a coin properly 50 times. Put H or T in each box.

Figure 8. Coin Tossing Problem. (Green,1991, p.321)

Results showed that most of the students put the same number of heads and tails in the boxes. Green (1991) argued that students were attempting to be “highly accurate in reflecting equal probability of Heads and Tails” (p.327). However, the equal probability happens reliably only in a large number of repeated experiments. Although there is a very slight chance of happening, 50 trials are just too small to show the theoretical probability ---50%. Students’ answers about the equal frequency of heads and tails revealed that they didn’t understand characteristic five, which requires a precondition of large sample size to show a distribution. It indicated that students’ insufficient knowledge of sample size interacted with their misunderstanding of randomness and random sequence.

Results also showed that the majority of students put down too many short runs and too few long runs. In real random sequences, like coin flipping, long runs have a higher probability to happen than people usually assume. For example, flipping a fair coin for 100 times, the expected frequency to get HHHH or TTTT is 3.13 and the expected frequency to get HHHHH or TTTTT is 1.56 (ABC Scientist, 2008). (Definition of “Run”: for example, in a sequence of HTHHT, there are four runs: H, T, HH, T, among which HH has the longest length of two). Green (1991) criticized students’ attempts to produce short runs without extreme long runs as being “too consistent to reflect random variation” (p.327). Students may think that after a few times of heads, there should be a tail to balance, and long runs shouldn’t happen unless the coin was unfair. However, in a real world, long runs happen from time to time, and too many short runs indicate non-randomness in Exploratory Data Analysis (EDA) (Filliben, 2013). This type of misunderstanding on random sequence is also called Gambler's Fallacy. For instance, many roulette gamblers believe that after a long run of black, the next should be red. However, each single outcome in a random sequence is independent from each other; therefore, there is not a

memory to record what happened or adjust what happens next to make a balance. Besides, students' misconceptions about randomness and random sequence were already found detrimental to their science learning. For instance, students believed that random processes were not reasonable enough to contribute the species evolution while designed processes were more adoptive (Garvin-Doxas & Klymkowsky, 2008). Students' misunderstanding of randomness (and hence results in errors in science learning) call for educators' attention in the teaching and learning of the concept randomness. Meanwhile, similar misconceptions happen with teachers.

Ives (2011) conducted a case study on three pre-service high school mathematics teachers' conceptions of randomness as well as the relationship between their conceptions about randomness and their pedagogy in teaching randomness. The study design included, 1) four task-based interviews; 2), one activity which was about assessing whether a dice was fair by using simulation; 3), one activity which was school students' probability homework analysis; and 4), a reflection of what was learned in the activities. Ives found that all the participants had difficulties in understanding randomness. In the initial interview, participants believed that the outcome sequence of random samples should be non-repetitive, which meant these pre-service teachers held a typical wrong view of random sequence---the less order, the more randomness. Later in the post-interview, all three participants gave up the prior wrong perspective and embraced another wrong one----they thought that all random phenomena fitted in the classical probability model and therefore randomness meant each outcome had an equal chance to be displayed.

In a classical probability model, there are limited outcomes and all outcomes have an equal chance of appearing, for example, flipping a coin or tossing a dice. However, a classical probability model is just one type of probability models. For example, there are the normal distribution model and the geometric probability model (image a spinner where 2/3 is black and

1/3 is red). Ives' findings showed that, first, pre-service mathematics teachers had a similar misunderstanding of randomness as students, which is the less order, the more randomness. Second, pre-service teachers lacked basic content knowledge of probability and statistics, for example, the distinct types of probability models. Findings in this study were consistent with what Begg and Edward (1999) found out among 34 pre-service elementary teachers in New Zealand. In that study, over 67% of the participants believed that "order or pattern would not be associated with random events" (Stohl, 2005, p.356). For example, participants thought that a coin flip of HTHTH was less likely than TTHTH while the two outcomes had the same probability to appear.

Schreiber (2014) studied novice mathematics teachers' conceptions about randomness in Israel. Many of the participants had been working as mathematics teachers from one year to three years. Two questionnaires that contained a random sequence problem and a random array problem were used as detection tools. Based on teachers' answers, Schreiber generated that first, novice mathematics teachers' opinions on randomness were highly related to "the deviation of order and coherence of sets of objects and the deviation of sets of objects from symmetries" (Schreiber, 2014, p.93), which meant that those novice teachers also held the view of *the less order, the more randomness*. Second, when judging which sequence was random, novice mathematics teachers tended to choose the sequences with disordered elements. For example, most of them believed that HHHH is not random while HHTH is random. In other words, novice teachers believed that random samples should have short runs to avoid self-cloning long runs, and the short runs shouldn't look the same.

In sum, findings showed that students and mathematics teachers shared similar misconceptions of randomness. In addition, the lack of fundamental stochastic content

knowledge contributed to these misunderstandings. The reality indicated a need for emphasizing the teaching of the concept randomness and related probability content knowledge in mathematics teacher education.

Training on randomness. Researchers in Spain designed a formative activity to assess and develop teachers' understanding of random sequence. The activity included three sessions which were one hour each and had a one-week interval between sessions. In the first session, prospective teachers finished an assessment questionnaire which was designed based on the work of Green (1983) to examine their current understanding of random (see figure 11). In the second session, prospective teachers explained and defended their answers among groups and with the researchers. Researchers used simulation activities to make participants realize their misconceptions and to help them overcome these misconceptions. In the third session, participants were shown Spanish children's responses to the same tasks and evaluated these responses in small groups (Batanero, Gomez, Gea, & Contreras, 2014).

Clara and Luisa were each told to toss a coin 150 times. One did it properly. The other just made it up. They put 0 for Heads and 1 for Tails.

Clara: 01011001100101011011010001110001101101010110010001
01010011100110101100101100101100100101110110011011
01010010110010101100010011010110011101110101100011

Luisa: 10011101111010011100100111001000111011111101010101
11100000001000101001000001000110001010000000011001
00000001111100001101010010010011111101001100011000

Question 1. Did Clara or Luisa make it up?

Question 2. How can you tell?

Figure 9. Random Sequence Problem. Batanero, Gomez, Gea, & Contreras (2014).

Researchers claimed that the three-hour intervention significantly improved pre-service teachers' probabilistic knowledge of randomness. Before the intervention, participants showed common difficulties in understanding randomness, for example, random sequences shouldn't have patterns and Gambler's Fallacy. By the end of the third session, most participants understood the simulation experiment and realized that (a) randomness cannot be identified as an absence of pattern, and (b) long runs are not rare in random sequence. Rossman (2008) suggested using "randomization tests for connecting the randomness used in collecting data to the inference to be drawn" (p.5), which was considered in this study.

Sample and sampling. Sample and sampling are required by CCSS from grade seven.

Definition of sample and sampling. A sample gives people some information about a population, which means that a sample cannot be perceived as an equal representative of the population (Rubin, Bruce, and Tenney, 1991; Smith, 2014). In practice, this means that data from samples enables people to get an approximate value of characteristics of the population; however, it is impossible to know precise values of characteristics of the population. For example, the portion of boys in one school (population) is 50%, however, in one classroom (sample), the portion may be 65% (Rubin et al., 1991). Sampling means picking up samples from the population. Random sampling is required for making statistical inference.

A proper sample should be a random sample, and an appropriate sampling should be a random sampling; that being random is the base in doing statistical analysis. However, it needs to be emphasized that random sampling does not guarantee a characteristic matching between the sample and population. Random sampling is required because "random sampling eliminates bias by giving all individuals an equal chance to be chosen" (Moore & McCabe, 2006, p.219). Even more critical, statisticians believe that "the mathematical theorems which justify most frequentist

statistical procedures apply only to random samples" (Smith, 2014, pp.3). Therefore, it is important to make sure that people understand random sample, random sampling, and can interpret data from samples correctly. Unfortunately, research showed that both students and teachers have misconceptions about sample, sampling and statistical inference.

Misconceptions on sample and sampling. Following is one of the items that Watson and Moritz (2000) used to analyze students' conceptions of sampling.

> ABOUT six in ten United States high school students say they could get a handgun if they wanted one; a third of them within an hour, a survey shows. The poll of 2,508 junior and senior high school students in Chicago found 15 percent had actually carried a handgun within the past 30 days, with 4 percent taking one to school. (a) Would you make any criticisms of the claims in this article? (b) If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona? Why or why not? (p.114).

Responses analysis showed that grade nine students did have a certain level of conceptual understanding of sampling correctly. For example, students believed that a bigger percentage of the population should be taken as a sample instead of a small one. However, only 6 percent of the students noticed the sampling fallacy—the sample was from Chicago only, which means the sample was not random (Batanero and Sanchez, 2005). In other words, students didn't appreciate the importance of random sample and random sampling.

Results also showed that students didn't get the correct cognition that population was not a solo homogeneous unit. Some students argued that the sample from Chicago was enough because the whole country would be just the same as Chicago (Watson & Moritz, 2000). This type of belief showed that students didn't understand that population has variability inside and

therefore random sampling is needed to give each individual an equal chance to be picked (Moore & McCabe, 2006).

Most of the misunderstandings in statistics, such as representativeness heuristics, the Gambler's Fallacy, misconception of randomness, are claimed to be related to sample (size), sampling, or how to make statistical inference correctly (Watson & Moritz, 2000). Researchers found that it was very hard to teach the concepts of sampling and statistical inference (Rubin, Bruce, and Tenney, 1991). Whether this is because the topic of sampling is more descriptive and less numerical than other concepts in mathematics curriculums (Watson, 2004), or because it is counter-intuitive (Borovcnik & Peard,1996), is unknown. Studies that are shown below indicate that teachers have misconceptions of these concepts themselves; that improving teacher knowledge may be an option for effective classroom teaching.

Lee and Hollebrand (2008) designed a project to prepare pre-service secondary mathematics teachers in teaching probability with technology-based simulations. The experimental group contained 18 pre-service teachers while the control group had 15. Researchers found that although these pre-service teachers knew how to use empirical experiments (which was a simulation) to generate a relative frequency, they all used small sample sizes to get the relative frequency and then estimated the theoretical probability without any doubt. It meant that these participants didn’t understand the importance of sample size and the prerequisites of making inference between sample and population. Participants also didn’t emphasize that students can practice repeated sampling to get a relative frequency (for example, flipping the same coin repeatedly), which may be because of their limited understanding of the procedure of random sampling. Batanero, Godino and Cañizares (2005) also used simulation as a tool to train pre-service teachers in overcoming probability misconceptions. After simulation

training, most of the participants still answered incorrectly in the hospital problem by applying representativeness heuristics because participants still believed that every sample should be representative of their parent population without considering the sample size. It indicated that even after simulation training, participants were still short of correct understanding of sample and statistical inference.

Watson (2001) designed a profiling instrument to detect in-service mathematics teachers' knowledge in teaching chance and data, including both content knowledge and pedagogical content knowledge. Fifteen grade five and grade six elementary school teachers and 28 secondary school mathematics teachers in Australia participated in the study. When being asked to choose a lesson to teach, elementary school teachers preferred to teach classical probability topics, for example, dicing. Secondary school teachers preferred to teach the general idea of chance with specific probability distributions, for instance, normal distribution. In general, teachers were found to be more familiar with the concept of average than the concept of sample. Teachers felt more comfortable in teaching average than teaching the concept sample (Watson, 2001).

This result may be a sign of these in-service teachers' inexperience with the concept of sample, which means these teachers were short of content knowledge of sample. Or the result may be because of these teachers' unawareness of the essential role of sample in learning probability and statistics, which means teachers lacked teacher content knowledge (Watson, 2001). Stohl (2005) pointed out that the teachers in Watson's study preferred to teach statistic concepts with computational procedures rather than the non-procedural concept of sample which is more descriptive and less numerical. Stohl argued that these mathematics teachers might have high self-belief on the concept of average because they knew how to do the calculation. To the

opposite, mathematics teachers had a little confidence in teaching the concept of sample because sample doesn't have a specific calculation procedure to follow, which is beyond these mathematics teachers' comfort zone (Stohl, 2005). More research is needed to show in-service teachers' conceptions of sample and sampling. As Stohl (2005) said, people think they understand sample; actually, they do not.

Sampling distribution and statistical inference. Informal inference is required by CCSS from grade seven; formal statistical inference is required in CCSS high school statistics standards; sampling distribution is the knowledge base for formal statistical inference.

Definitions. The purpose of statistical inference is to draw "conclusions about a population based on a sample or about a treatment effect based on random allocation of subjects" (Rossman & Chance, 2000). The following shows basic concepts in statistical inference. Parameter: a quantity that is computed from the population. For a huge/infinite population, the value of a parameter is usually unknown. Statistic: a quantity that is computed from a random sample. Sampling distribution: the distribution of the values of a statistic. There are many options for sample statistics, such as sample mean, sample median, and sample deviation, among which *sample mean* is a very popular statistic. CCSS emphasizes sample mean in statistical inference (HSS.IC.B.4); therefore, *sample mean* will be considered in this study. The relationship among sample statistics, population parameter, and statistical inference is as follows. First, people generate the value of sample statistics from samples. Next, people estimate the value of population parameter by carrying out statistical inference. Figure 10 shows the relationship among population, sample, random sampling, and statistical inference.

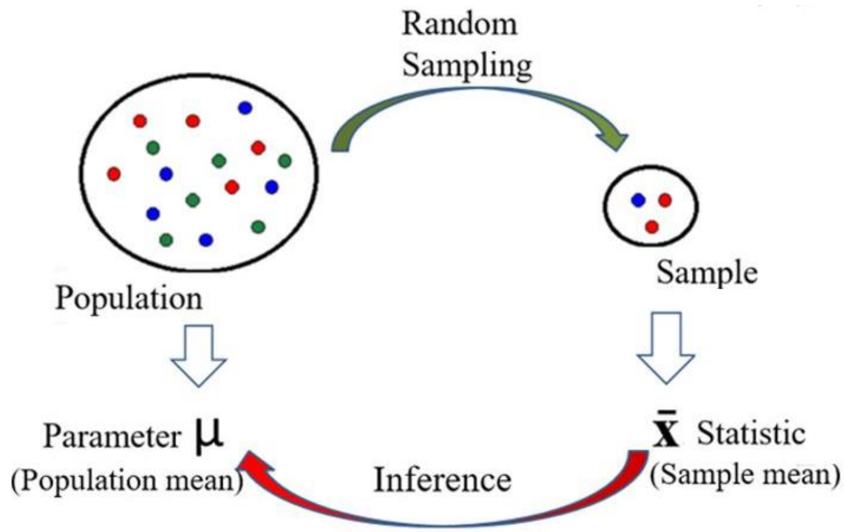


Figure 10. Relationship among Population, Sample, and Statistical Inference.

This figure shows the logic of statistical inference in a simple loop. However, it is the Central Limit Theorem, the Law of Large Numbers, and sampling distribution that support the logic with a solid theory base. The probability distribution of a statistic is called *a sampling distribution*. Let's use an activity that determines the mean length of words as an example to explain sampling distribution.

There are 268 words in the Gettysburg address. Teacher Ada wants to show a sampling distribution of word length (sample size is five) in this address. The first step, she picks up five words from the entire Gettysburg Address randomly and marks this group of words as Sample No. 1. The teacher Ada then calculates the mean length of words in Sample No.1 and labels the value as Sample Mean No.1. The second step, she picks up another group of five words randomly from the full Gettysburg Address and marks the second group as Sample No. 2. She also gets the mean length of words in Sample No.2 and labels the value as Sample Mean No.2. The third step, she keeps repeating the same procedure for 498 times more. In the end, she gets

500 samples and 500 sample means. She displays the distribution of the 500 sample means in a histogram graph, and this distribution is the sampling distribution (sample size is 5). Please see Figure 11 as the histogram graph. The mean length of all the 268 words, which is the population parameter, is 4.30 (in red).

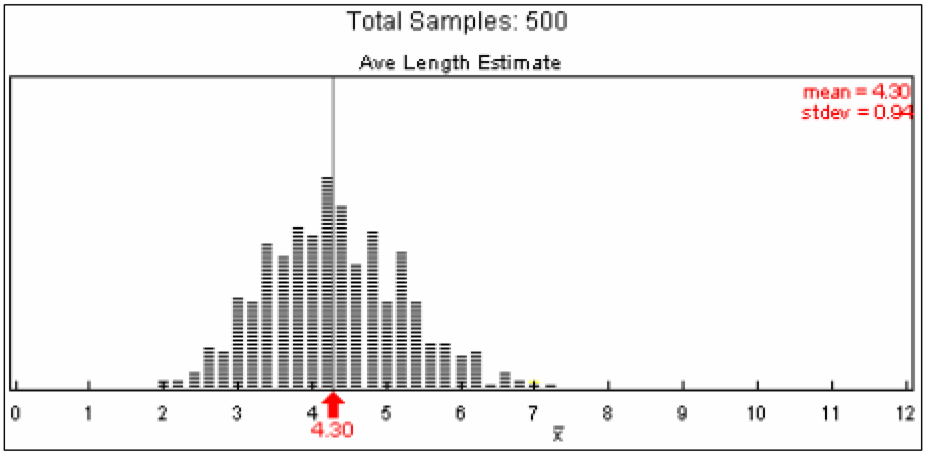


Figure 11. Sampling distribution (Chance & Rossman, 2006, p.3).

This sampling distribution resembles a normal distribution whose mean is the population mean-- 4.3. The Central Limit Theorem ensures that, when the sample size is large enough (generally ≥30), the distribution of sample mean resembles a normal distribution whose mean is the population mean. The Law of Large Numbers ensures that, when the sample size gets larger, the sample mean gets closer to the population mean. This type of definition is also called the Empirical Law of Large Numbers (Sedlmeier & Gigerenzer, 1997), which is the simplest approach to understand the Law of Large Numbers (Verhoeff, 1993). The statistical inference is based on sample size and random sampling. Without random sampling, a statistical inference cannot be made. The larger the sample size is, the closer the sample mean resembles the

population mean. Studies showed that people have misconceptions about statistical inference and sampling distribution.

Misconceptions. Saldanha and Thompson (2003) did an empirical study on high-school mathematics instruction and found that students had difficulties in understanding sampling distribution. They argued that, in practice, the statistical inference was usually based on one single random sample. Therefore, students didn't understand that sampling distribution was based on many random samples. Sedlmeier and Gigerenzer (1997) did a literature review on psychology studies that were related with people's misunderstanding of *the Law of Large Numbers.* They believed that people understood and could use *the Law of Large Numbers* intuitively to solve a problem that was based on relative frequency. For example, when flipping fair coins, people did hold the view that the more experiments were done, the more likely the relative frequency of getting tail was 50%. On the contrary, people did not understand the importance of sample size in the *Central Limit Theorem* or the data deviation from population mean in sampling distribution. For instance, Sedlmeier and Gigerenzer argued that the hospital problem that was used by Kahneman and Tversky in 1972 asked for people's understanding of two different sampling distributions instead of *the Law of Large Numbers*; one sampling distribution had a sample size of forty-five while the other had a sample size of fifteen. The sampling distribution that was based on a smaller sample size had a bigger deviation from the population mean; if people knew this property of sampling distribution, people would have answered this hospital problem correctly.

Training on sample, sampling and statistical inference**.** Saldanha and Thompson (2003) posed a new conception of sample, called Multiplicative Conception of Sample (MCS), to improve students' correct understanding of sample and hence improve the proper understanding

of sample mean, sampling distribution, and statistical inference. The key point in correctly understanding sample as an MCS is that although we usually use a single sample to make statistical inference, we always assume that this sample is one member of a sample family that has a sampling distribution as a whole. Please see figure 12 as an example to explain MCS.

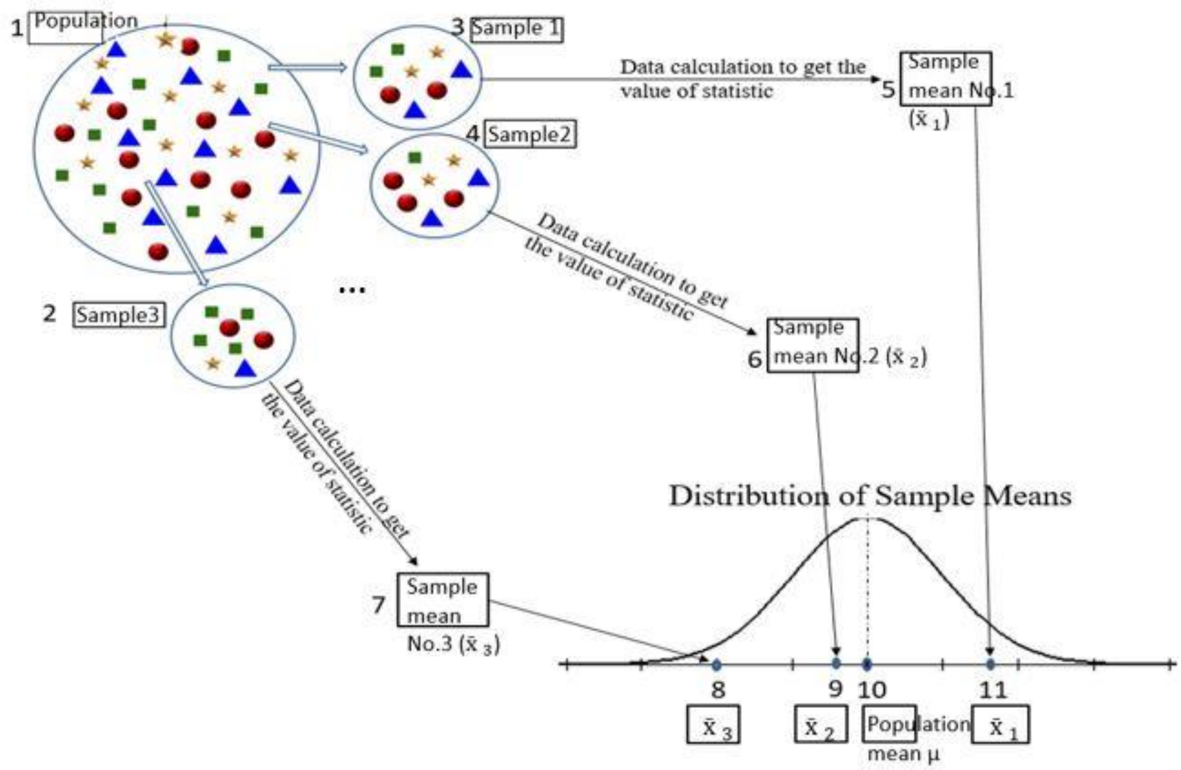


Figure 12. Sample mean and sampling distribution.

Meanwhile, Sedlmeier and Gigerenzer (1997) argued that the training on the Central Limit Theorem should help people understand the role that sample statistic play in statistical inference. Biehler (2001) contributed successful experience in improving statistical inference reasoning. He designed a research protocol aimed at helping students learn statistical inference reasoning by comparing specific distributions (e.g. boxplots). After analyzing the results, he

generated a four-stage student development route to upgrade students' thinking toward formal inference. The four stages were based on one statistical inference reasoning perspective---Exploratory Data Analysis (EDA). Unlike probabilistic thinking, which emphasizes constructing probability models (e.g. standard normal distribution), EDA seeks to understand the world by analyzing data, for example, constructing data distributions like boxplots. Student roles were described in each stage (Biehler, 2001; Pfannkuch, Budgett, Parsonage, 2004): 1. Be the EDA methods expert. For example, students were able to fine-tune the boxplots comparison. 2. Be the subject matter researcher and discoverer. For example, students could widen and exploit the context by bringing in more variables. 3. Be the critical theory builder. It required that students be able to do generalization based on distribution comparison. 4. Be the inferences statistics expert. In this stage, students need to do chance critique. Biehler claimed that these development stages could help guide instruction in inference reasoning, providing an option for relevant interventions. Watson and Moritz (1998) did an empirical study on comparing two data sets as the beginning of statistical inference in training elementary school students. They found that this instruction design improved students' understanding of statistical inference, and they believed that this activity could also be used in middle school instruction.

Other studies analyzed students' misconceptions of sample and sampling insightfully and provided meaningful suggestions for teacher education. Rubin, Bruce, and Tenney (1991) found that students hold a wrong view that good samples should stand for their parent population accurately, being free of error. However, even random sampling cannot guarantee flawlessness. It indicated that students didn't accept the notion of sample variability, but relied too much on sample representativeness. Sample variability means that samples from the same population may be different from each other and, therefore, not all the samples match their common parent

population. Sample representativeness is the view that a sample has similar features to its parent population. A balance of the two contrasting views can lead to a correct understanding of sample and inference while most of the students often over-respond to one side and ignore the other. These findings suggested specific components--- sample variability and representativeness---- that deserved symptomatic treatment in future researches and interventions. This professional development will take the above activities as sample interventions and utilize TinkerPlots as data analysis technology.

In sum, prior studies provided valuable suggestions for instructions and interventions for probability and statistics content knowledge training. Systematic activities-based content knowledge instruction, small-group work, data analysis based practices, and symptomatic treatments on specific items work together as a guide for the intervention design in this study.

Relationship Between Content Knowledge and Teacher Self-efficacy

This section of literature shows the influence of teacher self-efficacy on teachers' mathematics anxiety, students' belief in mathematics, student performance, and the relationship between teacher content knowledge, pedagogy knowledge, and teacher self-efficacy. This literature review contributes to this study by providing prior research findings and background in answering research question No. 3---"What is the correlation between teachers' content knowledge and self-efficacy in teaching middle-school statistics and probability?"

Definitions. Teacher content knowledge refers to the three domains of subject matter knowledge that Further et al. (2005) proposed: common content knowledge (such as how to calculate *mean*), horizon content knowledge (such as understand that *mean* refers to equal share in elementary school but refers to central limit in secondary school) and specialized content knowledge (such as understanding that *mean* is a balance point in a data set).

Mathematics/statistics self-efficacy means an individual's confidence in doing mathematics or statistics. Similarly, teacher self-efficacy refers to teachers' confidence in their capability to support student learning (Hoy, 2000, p.2). The following literature shows the role of teacher self-efficacy and the relationship between teacher self-efficacy and teacher knowledge.

The role of teacher self-efficacy. First, teacher self-efficacy is believed to be negatively correlated with teachers' mathematics anxiety. Gresham (2008) surveyed 156 pre-service primary teachers using two validity-proven measurement instruments and found that the lower the teacher self-efficacy was, the higher the mathematics anxiety was. Second, teacher self-efficacy is associated with teachers' choice of teaching strategies. Czernaik (1990) found that teachers with high self-efficacy were open to using various teaching methods, such as student-centered strategies and problem-solving inquiries. Other studies reported that teachers with high self-efficacy utilized non-traditional instructional strategies without server resistance, such as using manipulatives and teacher-facilitated learning (Hoy, Hoy, & Davis, 2009). Third, teacher self-efficacy influenced students' beliefs in mathematics considerably. Midgely, Feldlaufer, & Eccles (1989) followed 1,329 students for two years from elementary school to junior high school, uncovering that "students' expectancies, perceived performance, and perceived task difficulty" (p.247) strongly relied on teacher self-efficacy. For instance, students who had high efficacy teachers in elementary school but had low efficacy teachers in junior high school showed the lowest perceived academic performance and the highest perceptions of tasks complexity at the end.

In brief, past studies showed that teacher self-efficacy plays an indispensable role in influencing both teachers and students. Teacher self-efficacy is associated with teachers' classroom instruction (Czernaik,1990), is significantly correlated with teachers' mathematics

anxiety (Gresham, 2008), and can affect students' confidence on learning mathematics and solving mathematics tasks (Midgley, Feldlaufer, & Eccles, 1989). It is necessary to improve teachers' self-efficacy to benefit both teacher effectiveness and student learning (Bray-Clark & Bate, 2003).

Teacher content knowledge and teacher self-efficacy. Gresham (2008) measured 156 primary pre-service teachers' self-efficacy, finding that participants who had positive mathematics experience in the past held low mathematics anxiety and hence retained high self-efficacy in teaching mathematics (p.182). Huinker and Madison (1997) carried out a pre-and post- teacher self-efficacy survey on sixty-two pre-service elementary school teachers utilizing science and mathematics teaching methods courses as interventions. They found that the mathematics instruction courses consistently and significantly improved participants' self-efficacy in teaching mathematics (p.117). Swars (2005) investigated four pre-service primary teachers' self-efficacy in teaching mathematics after they finished a mathematics method course, finding that this content and pedagogy embedded course correlated with the substantial increase in teacher self-efficacy. Iyer and Wang (2013) examined 117 pre-service elementary teachers after they finished a mathematics method course, finding that participants who processed extra content knowledge tend to hold higher self-efficacy in teaching mathematics (p.10). In short, past studies indicated that teacher self-efficacy in teaching mathematics is correlated to teachers' content knowledge and pedagogy knowledge. Interventions that contain mathematics content and pedagogy knowledge have the potential to improve teacher self-efficacy in teaching mathematics.

Teacher self-efficacy to teach statistics. Very limited studies are available on teacher self-efficacy in teaching statistics as well as the relationship between stochastic knowledge and

teacher self-efficacy in teaching statistics and probability (Harrell-Williams et al., 2015). Only two studies in New Zealand and Australia are accessible to date. Begg and Edwards (1999) surveyed twenty-two in-service primary teachers and twelve pre-service primary teachers' beliefs in teaching statistics in New Zealand. They found that very few participants had formal statistics training, and all participants held a negative attitude towards statistics. Surprisingly, more than 80% of the participants held a certain confidence in teaching statistics, although most of them cannot explain several fundamental concepts correctly (p.6). Begg and Edwards argued that teachers' insufficient understanding of statistics would highly likely result in students' learning difficulty in statistics (p.8). In Australia, Watson (2001) surveyed forty-three primary and secondary in-service teachers' confidence in teaching data and chance by using a semi-opened questionnaire, finding that teachers had the highest confidence in teaching graphs and had the lowest confidence in teaching odds (p.319). It is noteworthy that, both studies were undertaken outside of the United States and utilized measurements/surveys that were not validity-proven.

To the author's knowledge, the only available study in the United States was conducted by Harrell-Williams, Sorto, Pierce, Lesser, and Murphy in 2015. Harrell-Williams et al. designed middle grades *Self-Efficacy to Teach Statistics (*SETS*)* as an instrument which aligned with CCSS and GAISE Framework and proved the validity of this instrument (Harrell-Williams et al., 2014, 2015). This measurement was implemented on 309 pre-service middle-grade mathematics teachers and found that participants had the lowest self-efficacy in teaching "making distributional comparisons across groups, using and interpreting measures of association, and developing a research question" which are all level B concepts in GAISE Framework (p.11). Harrell-Williams et al. (2015) believed that these low confidences were a consequence of

insufficient knowledge and training in statistics and probability. However, they were not able to measure teachers' content knowledge at the time because they did not have a proper instrument (p.12), making their assumptions hard to prove. Harrell-Williams et al. (2015) planned to measure teachers' content knowledge in statistics in the future and to explore the relationship between teacher content knowledge and teacher self-efficacy.

In short, studies on teachers' self-efficacy in teaching statistics are limited. Available studies from New Zealand, Australia, and pre-service mathematics teachers in the United States showed that teachers have low self-efficacy in teaching several fundamental concepts which may be because of insufficient content knowledge. Meanwhile, self-efficacy studies, which explore the in-service middle-school mathematics teachers' self-efficacy in teaching statistics in the United States, are in shortage. In addition, the relationship between teachers' statistics content knowledge and self-efficacy in teaching statistics is unknown or no-evidence-confirmed. This research gap has been observed by researchers and needs to be filled.

Chapter 3: Methodology

Research Questions

This study aimed at answering the following questions:

1 To what extent did components of the PD program change participants' content knowledge, pedagogical knowledge, and self-efficacy in teaching statistics and probability?

2 To what extent were the components of the PD program enacted in the classrooms at school?

3 In what ways did non-PD secondary mathematics teachers incorporate BLAST (website: Bring Learning and Standards Together) into their teaching?

Design of the Study

Case study. A multiple case study approach was adopted in this research. Merriam (1998) indicated that case study design fits education phenomena that are particularistic in a bounded system. This study was particularistic because it focused on stochastic learning and application by a group of secondary school in-service mathematics teachers who participated in a statistics and probability professional development program, while working in a large urban school district in the southwest of United States. The bounded system refers to the focal point of a study (Stake, 1995). Characteristics of the participants, professional development content, and the context of the secondary school setting distinguish this study from others, making it particularistic in a bounded condition. In brief, the characteristics of this study show that an ideal design would be a case study.

According to Yin (2003), depending on the research goals, a case study can be an exploratory case study, an explanatory case study, or a descriptive case study. This study

matches the features of an explanatory case study, which aims at exploring “possible cause-and-effect relationships.” (Yin, 2003, p. 7) This study investigates teachers’ change after professional development training, teachers’ applications of PD components, and the factors that facilitate teachers’ change and PD content enactment. In so doing, the research was intended to describe the potential relationship between teachers’ improvement and the professional development program. Therefore, this study has the characteristics of an explanatory case study.

According to Yin (2003), a case study is successful when multiple resources of evidence are collected. Also, Simons (2009) indicated that case studies are not limited to qualitative methods. Therefore, to fully answer the research questions, both qualitative data (interview, classroom observation, artifacts, worksheets, etc.) and quantitative data (pre-and post-measurements, such as LOCUS content knowledge test and SETS self-efficacy survey) were collected. This data, collectively, can provide an in-depth understanding of the problem with concrete details and abstract analysis. In addition, a multiple case study can provide a more robust conclusion with a stronger validity than a single case study (Yin, 2003). This study intends to collect diverse sources of data on each professional development participant (three participants in total), including content knowledge, pedagogical knowledge, self-efficacy belief, classroom instructions, etc., making this study match the characteristics of a multiple case study design. In brief, it is meaningful to carry out the multiple-case study design in this study with both quantitative and qualitative data.

Two phases. In this case study, two phases were designed: The first phase was called *case selection,* aimed at selecting representative participants from PSPD participants to follow up. The second phase was called *case collection,* aimed at collecting data from each case selected.

In the first phase, the researcher analyzed PD teachers' performance on pre-and-post content knowledge test (LOCUS) and teachers' pre-and-post self-efficacy responses (SETS) in teaching statistics and probability. Based on the quantitative data analysis findings, three PD teachers were selected as representatives to follow up, named Cathy, Donna, and Tina (pseudonyms). These three teachers were selected from twenty-one PD participants as representing three diverse types of PD participants:

1) Content knowledge and self-efficacy both increased;
2) Content knowledge increased but self-efficacy decreased;
3) Content knowledge decreased and self-efficacy increased;

It was reasonable to select these three diverse types of teachers' classroom instruction due to their perceptions of this PD program to generate factors that facilitated their change. Teacher information is available in Tables 2 and 3. Exploring these three teachers' classroom instruction and factors that facilitated their change allowed me to provide a proper representation of a large scope of PD participants and provide comprehensive details to answer the research question No.1 and No.2.

Table 2. *Teachers' Change after PD Training*

Teacher	Gender	Post and Pre content knowledge difference Full marks: 100 points	Post and Pre self-efficacy survey difference Full marks: 156 points
Cathy	Female	+6 points, increased	+3 points, increased
Donna	Female	+10 points, increased	-15 points, decreased
Tina	Female	-4 points, decreased	-2 points, decreased
Red: increase. Green: decrease. Note: more information is available on page 78.			

Table 3. *Teacher Background and Teaching Experience Information*

ID	First Language	Teaching Experience	Grade	Credential Subject	Courses Taken (Pedagogy in Teaching Statistics)
Cathy	English	10 years	Grade 6	All subjects	0
Donna	English	13 years	Algebra	Math	2
Tina	English	13 years	Geometry	Math	0

In the second phase, three PD participants were selected as follow-up subjects. Each teacher was interviewed as a case. The researcher visited three PD teachers' classrooms to observe and video-record their classroom instruction and conducted semi-open teacher interviews. Classroom observation and instruction video-coding was for exploring to what extent PD participant teachers' enacted PD components in their classrooms as well as teachers' change in content knowledge and pedagogy knowledge. Interviews were for exploring teachers' self-exposure of learning in this PD and opinions of the factors that facilitated their changes (if applicable) during/after this professional development. Each teacher was observed three times, around one hour per time. Teachers were interviewed after each observation.

Meanwhile, in the second phase, to what extent teachers utilized the 2nd generation of PSPD, the BLAST website, was explored. These teachers were labeled as NPD (non-PD) teachers because they did not participate PSPD. In fall of 2016 and spring of 2017, college and graduate students who took secondary mathematics education courses at a southwest public university were recruited as NPD participants. These NPD teachers included junior teachers who were in ARL programs and pre-service teachers who were in math teacher programs.

Studies on NPD teachers included five steps. In the first step, NPD teachers completed pre-test of content knowledge (LOCUS) and pre-survey of self-efficacy in teaching statistics (SETS). For the second step, NPD teachers designed pre-lesson-plans, focusing on secondary

statistics and probability. In the third step, NPD teachers read and discussed BLAST content to explore this teaching resources website and provided feedback on BLAST. For the fourth step, NPD teachers designed post-lesson-plans, focusing on the same statistics and probability topic as the pre-lesson-plans. In the fifth step, NPD teachers completed post-test of content knowledge (LOCUS) and post-survey of self-efficacy in teaching statistics (SETS).

Data Sources and Analysis

To show a clear roadmap of this study, a data table (Putney, 1997) provided below is followed by detailed explanations (Table 4).

Table 4. *Data Table*

Research Questions	Kind of Data to be Collected	Process of Analysis	Literature	Time of Collection
Question 1	(n =21) PD teachers' pre-and post-content knowledge test (LOCUS), pre-and post-self-efficacy in teaching statistics and probability (SETS), PD worksheets, mini in-class tests,	t-test, descriptive statistical analysis.	Harrell-Williams, Sorto, Pierce, Lesser, and Murphy, 2014,	2016 January, March, and June.
Question 2	three PD teachers' classroom instruction videos, lesson plans, and interviews	content analysis, examination of topics taught in class, errors and imprecision, Common Core-aligned student practices, and how classroom work is connected to mathematics (MQI)	Hill, Kapitula, & Umland, 2011; Mayring, 2000.	2017 spring
Question 3	(n=17) NPD teachers' pre-and post-content knowledge test (LOCUS), pre-and post-self-efficacy in teaching statistics and probability (SETS), pre-and post-lesson plans, feedback on reading the 2nd generation of PD (called BLAST)	t-test, descriptive statistical analysis, content analysis, examination of topics in lesson plans, errors and imprecision, Common Core-aligned student practices, and how classroom work is connected to mathematics	Harrell-Williams, Sorto, Pierce, Lesser, and Murphy, 2014, Hill, Kapitula, & Umland, 2011	2016 fall and 2017 spring

LOCUS and SETS. Levels of Conceptual Understanding in Statistics Evaluation (LOCUS) was used and will be used as pre-and post-tests to examine participants' statistical and probability content knowledge (LOCUS Committee, 2015). This evaluation, which has thirty questions, tests participants' content knowledge of middle school level statistics and probability,

including reading stem-and-leaf plots, choosing sampling methods, etc. The validity of LOCUS is justified by Jacobbe, Case, Whitaker, and Foti (2014).

Self-Efficacy to Teach Statistics (SETS) will be used as a survey to detect teachers' self-rating of confidence in teaching probability and statistics (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014). SETS is a six-point scale survey designed by Harrell-Williams et al. (2014) to explore middle school mathematics teachers' self-efficacy in teaching probability and statistics. There are twenty-six items in total. The total score of the SETS survey is 156 points. SETS includes two parts: participants' background information and participants' self-efficacy in teaching statistics and probability. For example, one of the questions is to rate participants' confidence in teaching *interpret measures of association*. The validity and reliability of SETS have been proven by Harrell-Williams et al. (2004). Data from LOCUS and SETS will be analyzed to clarify teachers' content knowledge and teachers' self-efficacy in teaching statistics and probability. Detailed background information about SETS and LOCUS is available in appendixes.

The analysis of PD teachers' performance on SETS and LOCUS was utilized in Phase I to show PD teachers' change and served as evidence for selecting three follow-up subjects in Phase II. The analysis of NPD teachers' performance on SETS and LOCUS will be utilized to show NPD teachers' change after they read BLAST. In addition, a comparison of PD teachers' performance and NPD teachers' performance on SETS and LOCUS will be conducted to explore the influence of this PD among these two separate groups.

Classroom instruction videos. In this study, three PD participant teachers were observed, each teacher representing one type of content knowledge and teacher self-efficacy change. Haller (1997) also observed four teachers' teaching probability (twice per teacher) after

they participated a statistics and probability PD training. Taking prior research experience and actual situation into account, it was reasonable to observe three teachers in this study.

Three PD participant teachers' classroom instruction was videotaped and analyzed by following video coding protocol Mathematical Quality of Instruction (Hill, Kapitula, & Umland, 2011). Mathematical Quality of Instruction (MQI) could be used in analyzing videotaped classroom instructions to provide details on teachers' performance of mathematics instruction. MQI measured five dimensions of instruction: (1) richness of the mathematics; (2) working with students and mathematics; (3) errors and imprecision; (4) Common Core-aligned student practices; and (5) classroom work is connected to mathematics (National Center for Teacher Effectiveness, 2016b). PSPD delivered training possessed the above instruction features positively; therefore, this MQI was also used to measure PD participants' classroom instruction.

Two sources of variation were considered for MQI application: raters, and lessons (Shih, Ing, & Tarr, 2015). Raters were video coders. Three raters analyzed instruction videos separately to examine inter rater reliability (IRR). Lessons referred to classroom instructions that were observed. To avoid teachers designing lesson plans intentionally to show a different performance than usual, it was meaningful to observe each teacher two to three times. Meanwhile, analysis of the relationship among raters, lessons, and coding reliability showed that "more than three lessons yields diminishing returns in terms of the reliability coefficient; instead, adding a second rater to each lesson increases the reliability coefficient markedly" (Shih, Ing, & Tarr, 2015, p.368). Therefore, each teacher was asked to be observed three times to ensure a three-hour-length instruction video record.

Interview. Open-ended semi-structured interviews were conducted with three PD teachers. These interviews were carried out after classroom instruction observations. Questions

were about teaching experience, content knowledge, pedagogy knowledge, self-instruction evaluation, comments on this professional development, and lesson preparations. These interviews were to detect the influence of this professional development on teachers' knowledge, self-efficacy, classroom practice, and factors that facilitate teachers' change (if applicable). Sample questions were: *Where did you learn pedagogies to teach statistics and probability content, please? What factors influence your self-efficacy in teaching statistics and probability the most, please? What do you expect to see in a statistics and probability PD?*

T-test, correlation, and descriptive statistical data analysis methods were used to analyze data from the LOCUS test and SETS survey. MQI video-analysis protocol was utilized to analyze classroom instruction videos. Three coders worked separately, using the same rubric, to code the classroom videos. Inter-Rater-Reliability was calculated to examine the reliability of video-coding (Trochim, 2006). Content analysis was utilized to analyze teachers' lesson plans and feedback on BLAST.

Introduction of this Professional Development

This study was a follow-up to PSPD. More details about PSPD designing, content, and participants selection were introduced here for reference. Several sources provided guidance for the content design and structure design of PSPD to promote changes in in-service secondary mathematics teachers' stochastic content knowledge, pedagogical knowledge, and self-efficacy in teaching middle-school level statistics and probability. Included were theories and experience on stochastic reasoning development (Fischbein,1975; Wilson & Brekke, 1994), theories and research on stochastic professional development (Sztajn, 2011; Scheaffer, 1988), statistics and probability standards in CCSSM, suggestions from GAISE Framework (American Statistical Association, 2007), and prior studies on strengthening human stochastic knowledge.

As shown in the section of theoretical framework, Fischbein's (1975) theory indicates that increase in content knowledge can contribute to improvement in stochastic reasoning. Prior studies that explored human stochastic reasoning and training showed that normative training on sample size, sampling distribution, randomness, and the Law of Large Numbers could support human stochastic reasoning development (Larrick, 2004; Kirkebøen, 2009; Nisbett, Krantz, Jepson & Kunda, 1983; Sedlmeier & Gigerenzer, 1997; Garfield,1995). Therefore, this PSPD took classic activities from previous studies, such as the Hospital Problem (Kahneman, & Tversky,1972), as important assessment materials in this professional development to improve participants' content knowledge and hence strengthen participants' stochastic reasoning.

According to Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (American Statistical Association, 2007), four components of stochastic problem solving should be included in mathematics teacher training: formulating statistical questions, collecting data, analyzing data and interpreting results. Meanwhile, these four components are embedded in CCSSM from grade six to geometry. For instance, *statistical question* is in 6.SP.A.1, *collecting data* is in 6.SP.A.2 and 7.SP.C.6, *analyzing data* is in 6.SP.B.5.C (such as calculating mean and median), and *interpreting results* is in 7.SP.A.1& 2 (such as statistical inference). In this PD program, activities covered all these four components for stochastic problem solving, content ranging from CCSSM grade six to geometry. GAISE Framework also suggests that there are three levels of understanding statistics content: Levels A, B, and C, from a superficial understanding to a profound statistical understanding. For instance, *mean* refers to equal share at level A, indicates a balance point at level B, and serves as a statistic at level C. These three levels of understanding of *mean* were embedded in the PD activities. Following CCSSM, activities in this professional development were designed based on classic activities, such as the *Gettysburg*

Address Activity (Chance, & Rossman, 2006), and activities from other professional resources, such as Engage NY (New York State Education Department, 2016), CPAMLS (Florida State University, 2013), *Bridging the Gap* (Hopfensperger et al., 2012), etc.

Besides content knowledge training, lesson planning was another essential component in this professional development. Participants were asked to design statistics and probability lesson plans and activities and shared these materials via *Bring Learning and Standards Together(BLAST*). Constructing lesson plans was not only a reinforcement of teachers' content and pedagogy knowledge learning in this professional development, but also an approach to extend the influence of this professional development outside of this training group.

This PD aimed at improving teachers' content knowledge and pedagogical knowledge in teaching statistics and probability. Therefore, PD content matched five implicit features: (1) PD instructions and activities were connected to mathematics, particularly to statistics and probability; (2) PD content was rich of mathematics; (3) PD content had trainer-participant interactions, such as discussions and activities; (4) PD content was precise at terminologies and procedures, for instance, carefully designed terminologies sheets and worksheets were handed out; (5) PD activities were CCSS aligned. The above five items were also expected in high quality mathematics instruction; for instance, in the framework of measuring Mathematical Quality of Instruction (MQI), the above five components were utilized to measure classroom instruction (Hill, Kapitula, & Umland, 2011). Therefore, PSPD possessed the above five features implicitly.

Schedule of this Professional Development

This PSPD was carried out in six days from January to June. Pre-test of statistics content knowledge (LOCUS) and pre-survey of self-efficacy in teaching statistics and probability

(SETS) were finished on the first day and the last day. From the first day to the fourth day, professional development training was implemented. From the fifth to the sixth day, PD teachers designed lesson plans and instruction materials to construct *Bring Learning and Standards Together (BLAST*) as an online teaching resource website. Following is a brief description of CCSSM standards covered and activities utilized in these six professional development days. A few PD activities are available in appendixes.

The first day. CCSSM standards in grade six and grade seven (7. SP. A.) were covered on the first day. Two main activities were carried out. Activity one: develop a profound understanding of formulating a statistical question. Activity two: develop a profound understanding of measure of center.

Activity one was approximately forty minutes, focusing on improving teachers' knowledge in distinguishing and formatting statistical questions. This activity was composed based on Hopfensperger et al.'s activity---Formulating a Statistical Question (2012). The trainer (who is also the researcher) introduced the standards of formulating statistical questions first, and then participants worked in groups to practice. A mini-test was provided at the end of activity one. Detailed schedule and materials for activity one are available at Appendix A.

Activity two was composed based on several studies, among which the Gettysburg Address Activity made the major contribution (Chance & Rossman, 2006), by aiming at improving teachers' understanding of the mean in three levels (equal share, balance point, and statistical inference). Several statistical concepts were emphasized, including random, random sampling, sampling distribution, and statistical inference. Trainers first carried out a brief content knowledge heads-up and then moved to hands-on activities with TinkerPlots software as a simulation tool and demonstration tool. Mini pre- and post-tests were provided to check

participants' knowledge change. Detailed schedule and materials for activity two are available at Appendix B.

The second day. CCSSM standards in grade seven and grade eight (7. SP.B., and 8. SP. A.) were covered on the second day. The researcher designed and delivered activity three (find the probability of compound events), activity four (design and use a simulation to generate frequencies for compound events), and mini pre-test on probability. Materials are available at Appendix C.

The third day. CCSSM standards in Algebra I (S.ID.A and B) were covered on the third day. The researcher designed and delivered activity five (two-way relative frequency table).

The fourth day. CCSSM standards in Geometry (S.CP. A and B) were covered on the fourth day. The researcher designed and delivered activity six (Venn diagram, tree diagram, and two-way table) and mini post-test on probability. Mini post-test was available at Appendix C. Due to possible copyright conflicts, activities in the third day and the fourth day were not attached in this dissertation.

The fifth day and sixth day. In groups, participants composed lesson plans and designed classroom activities corresponding to Common Core State Standards. These materials were uploaded to and available at an online mathematics professional development website: BLAST (Bring Learning and Standards Together). On the last day, participants finished the post-content-knowledge test (LOCUS) and the post-self-efficacy survey on teaching statistics and probability (SETS).

Chapter 4: Results

Corresponding to the two phases of this case study research, data analysis in this chapter was divided into three sections. Section I contained data analysis on professional development participants' statistics and probability content knowledge and self-efficacy in teaching statistics by using the LOCUS test and SETS survey. Section II contained content analysis of three PD teachers' classroom instruction videos/notes and interviews. Section III included NPD math teachers' feedback on the 2nd generation of PD—teaching material website *BLAST*.

Section I: Data Analysis of PD Participant Teachers' Knowledge and Self-efficacy

Analysis of participants' content knowledge change (LOCUS).

PD participants' knowledge was measured by pre- and post-LOCUS tests and SETS surveys. LOCUS, which is short for *Levels of Conceptual Understanding in Statistics*, was first published in 2014 (Jacobbe, Foti, et al., 2014). LOCUS offers three types of tests: 1) Beginning and Intermediate Statistical Literacy; 2) Intermediate and Advanced Statistical Literacy; 3) Beginning, Intermediate and Advanced Statistical Literacy. In this study, considering PD participants were secondary school math teachers, Intermediate and Advanced Statistical Literacy, which contains 30 questions, was adopted.

Questions in LOCUS test were designed purposely into four categories: formulating questions (5 in total), collecting data (7 in total), analyzing data (7 in total), and interpreting results (11 in total). These four categories are the four steps of statistical problem solving per GAISE Framework (Franklin et al., 2007). Q-Q plot was conducted and showed that the population distribution of pre-and post-LOCUS test score difference was not normal. Therefore, a repeated-measure t-test was not carried out (Gravetter & Wallnau, 2013, p.330). Descriptive statistical data analysis was conducted to explore participants' content knowledge change.

LOCUS full credit was 100 points. The mean performance in pre-LOCUS test was 61.25, the standard deviation was 15.69. The mean performance in post-LOCUS test was 63.00, and the standard deviation was 14.13. Among twenty-one post-and pre-LOCUS performance differences, the mean was 1.75, the median was 0, the mode was 0, the smallest value was -17, the largest value was 24, and the standard deviation was 10.26. By comparing the pre-and post-tests mean scores, participants' LOCUS performance had an increase of 1.75 points (out of 100). LOCUS performance was also utilized to select three participants to follow up. Details are available in the section of Three Types of Cases.

Paired samples t-test on Participants' Self-Efficacy Survey (SETS).

PD participants' self-efficacy was measured by pre- and post-SETS survey. SETS, which is short for *Self-Efficacy to Teach Statistics (SETS) in Middle School Survey*, was first published in 2014 (Harrell-Williams, et al., 2014). SETS survey utilizes a scale of 1 to 6, asking participates to rate their confidence in teaching 26 statistics and probability topics. For example, one item asked teachers to "identify the association between two variables from scatterplots" (Harrell-Williams, et al., 2015, p.14). For all Likert items, the introduction was:

> Using a scale of {1, 2, 3, 4, 5, 6} where 1= not at all confident, 2 = only a little confident, 3 = somewhat confident, 4 = confident, 5 = very confident, 6 = completely confident, please rate your confidence in *teaching middle school students* the skills necessary to complete the following tasks successfully (Harrell-Williams, et al., 2015).

Paired sample t-test was conducted to analyze PD teachers' self-efficacy change on teaching statistics and probability. One participant did not finish post-SETS survey. Therefore, sample size n was 20. Taking all SETS survey items into account, there was a statistical

significant difference between pre-SETS survey (M=4.9365, SD= 0.7515) and the post-SETS survey (M=4.3846, SD= 0.9064), t (19) = 4.1224, p =0.0003, significant level α=0.05.

These 26 statistics and probability items could be divided into two categories: GAISE level A and GAISE level B. In SETS survey, level A was labeled as *Reading the Data* while level B was labeled as *Reading Between the Data*. All 26 items are shown in Table 5 and 6. Paired sample t-test was conducted on each item. Twenty-four paired t-tests showed statistically significant differences between pre-and post-SETS survey while four t-tests results were not significant. Detailed information is available in Table 5 (Harrell-Williams et al., 2015).

Table 5. *SETS Items: GAISE Level A*

Item	Question Stem	Category	Result
1	Collect data to answer a posed statistical question in contexts of interest to middle school students.	Level A	significant
2	Recognize that there will be natural variability between observations for individuals.	Level A	significant
3	Select appropriate graphical displays and numerical summaries to compare individuals to each other and an individual to a group.	Level A	significant
4	Create dot plot, stem and leaf plot, and tables (using counts) for describing distributions.	Level A	significant
5	Use dot plot, stem and leaf plot, and tables (using counts) for describing distributions.	Level A	significant
6	Create boxplots for summarizing distributions.	Level A	significant
7	Use boxplots, median, and range for describing distributions.	Level A	significant
8	Identify the association between two variables from scatterplots.	Level A	Not significant
9	Generalize a statistical result from a small group to a larger group such as the whole class.	Level A	significant
10	Recognize that statistical results may be different in another class or group.	Level A	Not significant
11	Recognize the limitation of making inference (i.e. generalization) from a classroom dataset to any population beyond the classroom.	Level A	significant
Note: Level A: Reading the Data. Items cited from Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics in middle grades. *Journal of Statistics Education, 23* (1), 1-20.			

Table 6. *SETS Items: GAISE Level B*

Item	Question Stem	Category	Result
12	Distinguish between a question based on data that vary and a question based on a deterministic model (for example, specific values of rate and time determines a particular value for distance in the model d = r × t).	Level B	significant
13	Identify what variables to measure and how to measure them in order to address the question posed.	Level B	Not significant
14	Describe numerically the variability between individuals within the same group.	Level B	significant
15	Create histograms for summarizing distributions.	Level B	significant
16	Use histograms for comparing distributions.	Level B	Not significant
17	Compute interquartile range and five-number summaries for summarizing distributions.	Level B	significant
18	Use interquartile range, five-number summaries, and boxplots for comparing distributions.	Level B	significant
19	Recognize the role of sampling error when making conclusions based on a random sample taken from a population.	Level B	significant
20	Describe numerically the strength of association between two variables using linear models.	Level B	significant
21	Explain the differences between two or more groups with respect to center, spread (for example, variability), and shape.	Level B	significant
22	Recognize that a sample may or may not be representative of a larger population.	Level B	significant
23	Interpret measures of association.	Level B	significant
24	Distinguish between an observational study and a designed experiment.	Level B	significant
25	Distinguish between "association" and "cause and effect".	Level B	significant
26	Recognize sampling variability in summary statistics such as the sample mean and the sample proportion.	Level B	significant
Note: Level B: Reading Between the Data. Items cited from Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics in middle grades. *Journal of Statistics Education, 23* (1), 1-20.			

Correlation between teachers' content knowledge and self-efficacy.

Pearson r was calculated to examine the linear relationship between teachers' LOCUS performance and SETS survey results. $r = 0.5404$. There is a moderate positive relationship between teachers' statistical content knowledge and self-efficacy in teaching statistics and probability.

Three types of cases

Based on the quantitative data analysis findings, four different types of changes in content knowledge and self-efficacy in teaching statistics were generated.

1) Content knowledge and self-efficacy both increased (eight participants in total);
2) Content knowledge increased but self-efficacy decreased (only one -- Donna);
3) Content knowledge and self-efficacy both decreased (only one -- Tina);
4) Content knowledge decreased/remained the same but self-efficacy increased (seven participants in total).

It would have been ideal to select one representative from each category. However, based on voluntary participation, three PD participants, Cathy, Donna, and Tina (pseudonyms), were selected to represent the first three types of PD participants. Donna and Tina were chosen because they were the only representative in their own categories. Cathy was selected to represent the first category because her content knowledge change (LOCUS) was close to the average change (mean=1.75, s.d.=10.26) and her self-efficacy change (SETS) was close to the average change (mean=15.85, s.d.=19.11). Tables 7 and 8 show the performance analysis and background information of these three teachers. The researcher observed three teachers' classroom instruction and interviewed all three teachers. Each teacher was treated as a separate case. More data is available in Section II.

Table 7. *Teachers' Change after PD Training*

Teacher	Gender	Post and Pre (LOCUS) content knowledge difference Full marks: 100 points	Post and Pre (SETS) self-efficacy survey difference Full marks: 156 points
Cathy	Female	+6, increased	+3, increased
Donna	Female	+10, increased	-15, decreased
Tina	Female	-4, decreased	-2, decreased

Table 8. *Teacher Background and Teaching Experience Information*

ID	First Language	Teaching Experience	Grade	Credential Subject	Courses Taken (Pedagogy in Teaching Statistics)
Cathy	English	10 years	Grade 6	All subjects	0
Donna	English	13 years	Algebra	Math	2
Tina	English	13 years	Geometry	Math	0

Section II: Three Cases

To evaluate the extent to which components of the PD program were enacted in the classrooms, three teachers' classroom instruction videos were analyzed, and teachers were interviewed. Three aspects were considered: 1), the extent to which teachers' instruction meet the five MQI dimensions that PSPD possessed positively; 2), the extent to which teachers' instruction utilized or was inspired by PSPD components; 3), how teachers designed lessons and teachers' feedback on PSPD.

Case one: Tina

Case descriptions.

Tina has been teaching secondary mathematics in public schools for 13 years. In fall 2017, Tina taught Geometry, which covered probability. There were 32 students in Tina's Geometry class. Her instruction was activity-based with problems and practice on worksheets.

The researcher video-recorded Tina teaching probability three times, for 75 minutes each. These instructions covered different probability standards. Video analysis, examples of teaching episodes, and teacher interview analysis are provided below.

Video Analysis- MQI.

MQI framework required each video segment to be less than 7.5 minutes. Therefore, each 75-minutes long video was split into 10 segments. In total, there were 30 video clips of Tina's instruction. MQI Instrument contains five main dimensions. Please see Table 9 for each dimension and corresponding levels.

Table 9. *MQI Instrument Coding System.*

MQI Instrument Dimensions	Levels			
(1) Classroom work is connected to mathematics	Yes		No	
(2) Richness of the mathematics	Not present	Low	Medium	High
(3) Working with Students and Mathematics	Not present	Low	Medium	High
(4) Errors and imprecision	Not present	Low	Medium	High
(5) Common Core-aligned student practices	Not present	Low	Medium	High

Three coders, the researcher, Leo, and Sara, analyzed Tina's instruction videos independently by utilizing MQI framework. Leo held a Ph.D. degree in statistics and has been teaching statistics and probability for more than three years. Sara is a Ph.D. candidate, majoring in higher education. Sara also holds a master's degree in higher education and leadership. Both Leo and Sara were familiar with the statistics and probability content that was taught in these instruction videos.

Intra-class correlation coefficient (ICC) was calculated to assess inter-rater reliability (IRR) among three independent coders. Data in this study was ordinal, therefore ICC was an

acceptable approach to measure IRR (Halgren, 2012). The ICC value for average measure was .895, which indicated good agreements among three raters. The ICC value for single measure was .740, which indicated that the reliability of a single rater was moderate. Three coders analyzed 10 video segments out of 50 video segments, which was 20% of the total video segments. The researcher coded 40 out of 50 video segments as a solo coder. Considering ICC value for single measure was .740, it was acceptable for the researcher to analyze videos solely. An example of three coders' analysis of *Richness of the Mathematics* is shown in Table 10.

Table 10. *Coding Result of Richness of the Mathematics*

Video Segment	Coder 1 (Researcher)	Coder 2 (Leo)	Coder 3 (Sara)
1	Not present	Low	Not present
2	Low	low	Medium
3	High	High	High
4	High	High	High
5	High	High	Medium
6	Medium	High	Medium
7	Low	Not present	Medium
8	High	High	High
9	High	Medium	Medium
10	High	High	High

In total, 30 video clips of Tina were analyzed. The summary was shown in Table 11 (see next page).

Table 11. *Tina's Instruction Video Analysis Result*

MQI Dimensions	Levels and Proportion				Sum
(1) Classroom work is connected to mathematics	Yes 97%		No 3%		100%
(2) Richness of the mathematics	Not present 0%	Low 3%	Medium 3%	High 94%	100%
(3) Working with Students and Mathematics	Not present 0%	Low 0%	Medium 17%	High 83%	100%
(4) Errors and imprecision *	Not present 100%	Low 0%	Medium 0%	High 0%	100%
(5) Common Core-aligned student practices	Not present 0%	Low 7%	Medium 3%	High 90%	100%
Note: There were 30 video segments in total. *: there were two instances of terminology nonstandard-usage.					

This PSPD executed the five MQI dimensions above implicitly. Video analysis elaborated that Tina's instruction was highly connected to math, highly rich in mathematics, highly CCSS student practice aligned, and medium in interaction with students and mathematics. Her instruction had very few mathematical errors while there were minor spelling errors. In general, Tina's instruction reached the requirements of being high-quality instruction according to the five dimensions in MQI framework.

Two instances of nonstandard-usage of terminologies in Tina's instruction caught the researcher's attention. When interpreting probabilities with percentage, Tina used 100 instead of 100% two times. For example, "if you are certain about it, what is the probability? 100, right?". Tina may be influenced by the term 50/50 in daily life or Tina just simplified the standard vocabulary without being aware of it. Another nonstandard-usage of terminologies was set and elements. For example, when calculating the probability of event A happening (please see figure 13), Tina expressed it as following-- "the region of A is 1, 2, 4 and 6". Tina utilized *region* instead of *elements* or *elements in set A.* Both Tina and students ignored these light nonstandard-

usage of terminologies and communicated about probability without difficulties. Obviously, students understood the content. However, it may cause extra explanations when students try to discuss statistics with or interpret probabilities to people outside of this classroom.

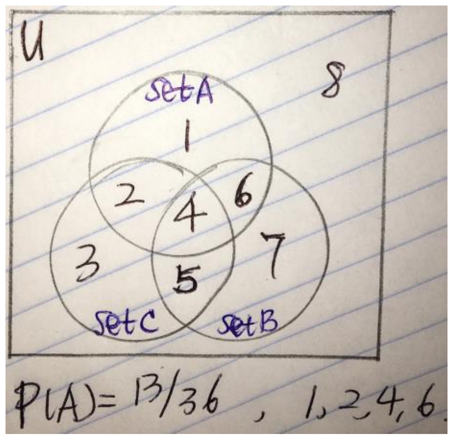


Figure 13. Probability of Event A

Teaching episode.

This episode was chosen from Tina’s video segment 3. Instructional goals in this episode were to introduce the concepts of *probability*, *sample space*, *uniform probability*, and *non-uniform probability*. Tina’s instruction was worksheet based, which was adapted from geometrycommoncore.com. One part of this worksheet is available below.

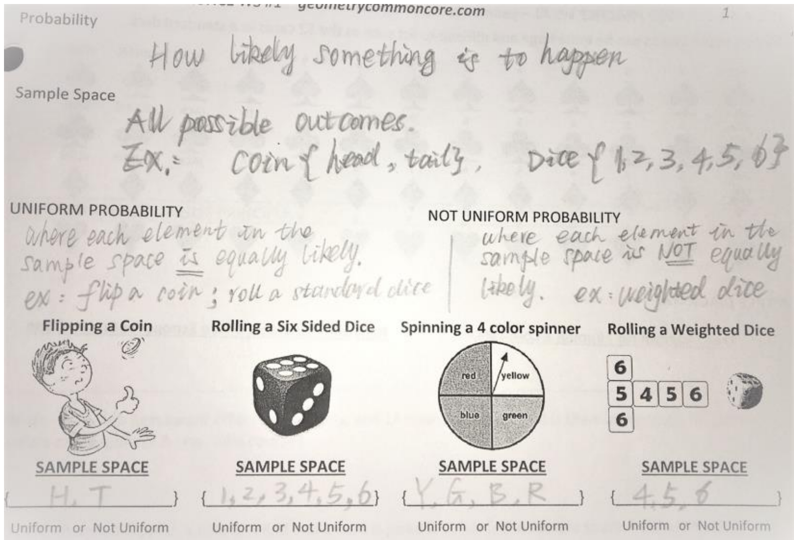


Figure 14. Worksheet Utilized in Tina’s Teaching Episode

Tina: I think what we are reading here are what you should already know. But … (teacher paused), let’s have a general picture of the knowledge first. So, what is probability? (teacher paused). Don’t be shy. Read after your paper.

Students: (answered in a low voice) probability is how likely something is to happen.

Teacher: We assign numbers to probabilities. If something cannot happen, what number should we use to represent the probability? Such as, how likely I am a dinosaur? 0? Right?

Students: (laughing and answered in a low voice) yes.

Teacher: Very sad, but it is true. What is the probability that I am a math teacher?

Students: (answered in a low voice) 100%.

Teacher: Can you change the percentage into a number?

Students: (answered together) 1.

Teacher: yes, 1, so, it is certain. It is certain that I am a math teacher. (Teacher wrote down *0-impossible* on the left and *1-certain* on right). What is the probability in between? If something is very likely to happen, where is the number should be?

Students: (answered in a low voice) should be close to 1.

Teacher: yes, it should be close to 1. Okay, I know we talked about sample space when something is happening. For example, for flipping a coin, the possible outcomes will just the two: head or tail, right? Assuming that the coin is a fair coin. However, coin can be not fair. Some coins have both heads or tails on both sides. Die is the same thing. You throw a die; you would examine it at first. (Teacher showed a die to the class). It should have one of each number on one side. How many sides are here?

Students: (answered in a low voice) 6.

Teacher: So, these are the potential outcomes (teacher pointed at the sample: {1, 2, 3, 4, 5, 6} on the worksheet). (Episode ended).

Students were required to finish the worksheet the day before this class started. This video segment showed that Tina's instruction was activity based with discussions between teacher and students. The same probability content was covered in the PD. However, Tina did not utilize the materials or worksheets from this PD. During the interview, Tina explained that PD opened her eyes to multiple various teaching methods and materials; however, she still preferred the teaching materials that she already used. Please see figure 15 as the Terminology Worksheet that was used in PSPD. Activity sheet was available in Appendix C.

CCSS 7. SP. C. 5 – 8

Terminology worksheet

Chance experiment/experiment: an experiment is an act or process of observation that leads to a single outcome that cannot be predicted with certainty. In other words, a chance experiment must have more than 1 possible outcomes and people cannot predict which outcome will occur.

Outcome: Each possibility of an experiment is called an outcome. Or, an outcome is the result of a single observation of an experiment.

- Eg.1. Experiment: Flip a coin once. Outcomes: H, T.
- Eg.2. Experiment: Flip a coin twice. Outcomes: HH, HT, TH, TT.
- Eg.3: Experiment: Toss a die twice. Outcomes:

		Second throw					
		1	2	3	4	5	6
First throw	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Sample space: The set of ALL the possible outcomes is called the sample space of the experiment. Sample space is usually denoted by capital S.

- Eg. 5. Tossing an empty paper cup to see how it lands. Three possible outcomes: right side up, upside down, on its side. Sample space: {right side up, upside down, on its side}.
- Eg. 3. Experiment: Toss a die twice. Sample space= the set of all possible outcomes= {(1,1), (1,2), …, (6,6)}

Figure 15. Terminology Worksheet used in PD.

Interview.

An interview was conducted to explore to what extent Tina believed that this PD helped her in teaching statistics and probability as well as Tina’s feedback to PSPD. Tina’s answers to key questions are shown in Table 12.

Table 12. *Interview Tina*

Key Questions	Tina's Answer
Where did you learn pedagogies to teach statistics and probability content?	Professional developments usually help a little bit; however, I learned most by teaching myself, such as searching online for materials and teaching the course.
Did the PD help you in getting/understanding stochastic content knowledge?	No. I understand the content knowledge already.
Did the PD help you in improving pedagogies in teaching stochastic content?	Yes. I learned a little bit. I would like to know how other people teach this content.
Did you use materials from this PD in your teaching?	No. I already have materials that I like.
Do you utilize the GAISE investigation cycle in your own classroom?	No.
What is your favorite part of this PD?	Collaborating with other professionals.
What factors influence your self-efficacy in teaching statistics and probability the most?	Experience. This is the second time that I teach geometry that has statistics and probability inside. I am better now than I taught it the first time. I believe I can be good at it after I teach it for three times.
What do you expect to see in a statistics and probability PD?	Connections among content and a picture of overall statistics and probability content in secondary school.

In Tina's opinion, PSPD helped her get fresh ideas of how to teach statistics and probability. However, she preferred to utilize the materials that she was familiar with already. Her instruction also reflected her point of view. Tina might trust the materials that she knew would work more than the new materials learned from a short-spanned PD. As previous research showed, teachers seem reluctant to adopt new practices unless they are sure they can make these new ideas work (Lortie, 1975).

Case two: Donna

Case descriptions.

Donna has been teaching secondary mathematics for more than 15 years. In fall 2017, Donna taught Algebra, which covered summarizing, representing, and interpreting data. There were 31 students in Donna's class. Her instruction was activity-based with problems and practice on worksheets. The researcher video-recorded Donna teaching statistics three times, which was around 45 minutes each. Video analysis, examples of teaching episodes, and teacher interview analysis are provided below.

Video Analysis-MQI.

MQI framework requires each video segment to be less than 7.5 minutes long. Therefore, Donna's instruction videos were broken up into 20 video clips, 2 minutes to 7 minutes long each. Please see table 13 for MQI video analysis results.

Table 13. *Donna's Instruction Video Analysis Result*

MQI Dimensions	Levels and Proportion				Sum
(1) Classroom work is connected to mathematics	Yes 100%		No 0%		100 %
(2) Richness of the mathematics	Not present 0%	Low 0%	Medium 15%	High 85%	100 %
(3) Working with Students and Mathematics	Not present 0%	Low 0%	Medium 5%	High 95%	100 %
(4) Errors and imprecision	Not present 100%	Low 0%	Medium 0%	High 0%	100 %
(5) Common Core-aligned student practices	Not present 0%	Low 5%	Medium 5%	High 90%	100 %
Note: There were 20 clips in total.					

Video analysis disclosed that for the most part Donna's instruction was highly connected to math, highly rich in mathematics, highly CCSS student practice aligned, and high in interaction with students and mathematics. Her instruction had very few mathematical errors, with minor spelling errors. In general, according to MQI framework, Donna's instruction fulfilled the requirements of being high-quality math instruction. Different from Tina, Donna used materials highly similar to PSPD for her instruction. For example, Donna used activities that were based on GAISE statistical investigation cycle and the *Estimating Ages of Famous People Activity* that was used in the PD. Details can be found in teaching episodes.

Teaching episode 1: Activities that were based on GAISE statistical investigation cycle

This episode was chosen from the second observation of Donna. In this class, Donna's instructional goals were aligned with statistics standards HSS.ID.A.2: Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

The activity that Donna utilized was called *An A-Maze-ing Comparison* (Maltlloure, Richardson, & Rogness, 2012). Donna obtained this activity via Regional Professional Development Program (RPDP) official website: http://rpdp.net/. According to the lesson plan description, this activity was a GAISE level C activity. Meanwhile, this activity followed all four components of GAISE statistical problem-solving procedure: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question.

Donna asked students to work in groups to answer this statistical question: *Does the mean time in seconds to complete a maze significantly differ between males and females*? Students were required to finish the maze game (see figure 16), record time spent as data, analyze data,

make graphs, and eventually answer the question. During most time of the class, Donna walked around the classroom and facilitated each team to complete the entire activity. Students worked in pairs, playing maze and counting time in turn. Please see below as the teaching episode.

Donna: Okay. When we are tracking the data, how should we put the data in the chart?

Students: all the data should be in seconds.

Donna: Good. So, make sure that you convert your time into all seconds. (Donna wrote on the board) For each minute, add 60 seconds. What else should we do with data? When we list the data on the table, how should we organize the table then?

Students: have male or females on one column and the time on the other.

Donna: Okay. What else could be good to simply this piece of data?

Students: Check who goes first (when checking gender).

Donna: Okay. Why is that important?

Students: Because the second person may memorize the maze and be faster than the first person.

Donna: Very good. Check your vocabulary sheet, these things can be called? What is your vocab word?

Students: (In low voice) Bias.

Donna: Okay. Bias can play a role here because the second person can be more experienced, because the second person counted the time and watched the first person stressed through it. Any methods we could use to avoid this bias?

Student A: Just let both people do the maze at the same time and each person just marks down their own time.

Student B: I started the maze from the top and she started from the bottom.

Donna: Okay, very good. That will be something to do to compare time spent (fairly). Now, how else do you think we can (use to) organize the table?

Students: (low voice) put gender in one column and put the time in the other column.

Donna: Okay…do we need to put F and M mixed in one column, just like the table on the paper sheet (see figure 16), or we could use other formats, split the F and M, put them into two columns? Can we have three columns in total? Can we have more than three columns?

Students: (low voice) Yes. It may help calculate the mean. We should put ages there too.

Donna: Okay, that's kind of what I am trying to get you think about when you are organizing the data. The table (which has two columns) on the paper sheet does not show much (information of the data). You need to examine your piece of data and think about *is there something that may impact my (data) organization?*

Students and Donna discussed about multiple situations that may impact data organizations, such as same maze vs. different mazes.

Donna: Okay. Because you are going to plotting every single person's time, and because histograms allow you to compare data, using frequencies. Histogram can show the frequency of each groups of time. We could also use box-plot and dot-plot. (episode ended.)

In this episode, Donna led students in exploring different approaches to analyze data, such as tables and histograms. However, students paid attention to bias at the beginning. Donna took advantage of this opportunity, explained bias, and then moved the topic to data presentations. Data presentation was one important part of the GAISE framework. This episode, together with MQI measurement results, showed that Donna's classroom teaching matches the

requirements of being high-quality mathematics instruction. Meanwhile, this episode showed that Donna taught students how to analyze data and present data, which were two essential parts of the GAISE framework, which was included in PSPD. Please see figure 16 as the maze-game. Please see figure 17 as the data presentation in class.

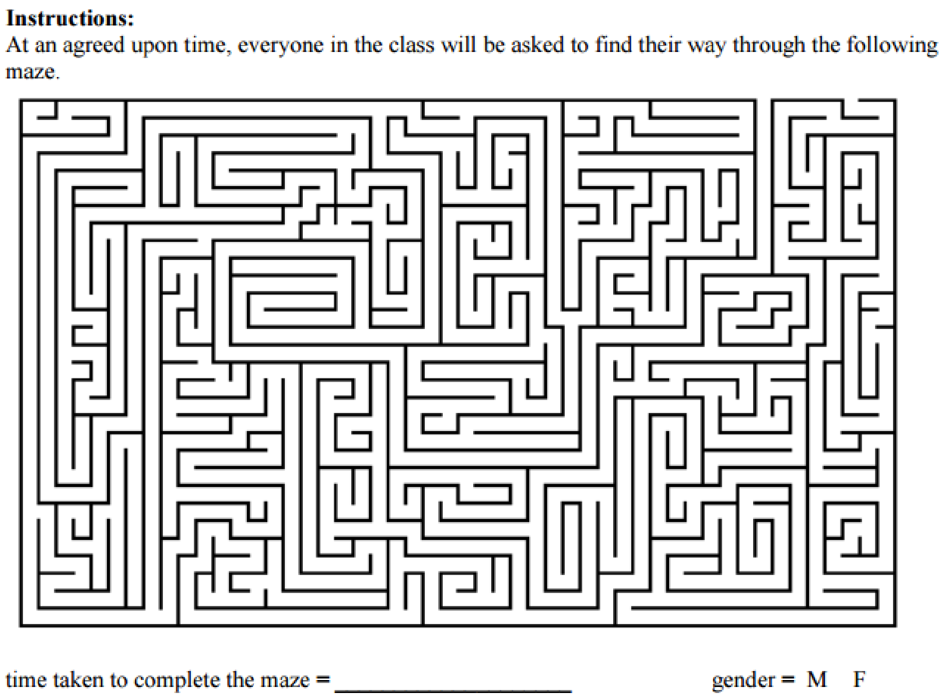


Figure 16. The Maze Game. (Maltlloure, Richardson, & Rogness, 2012).

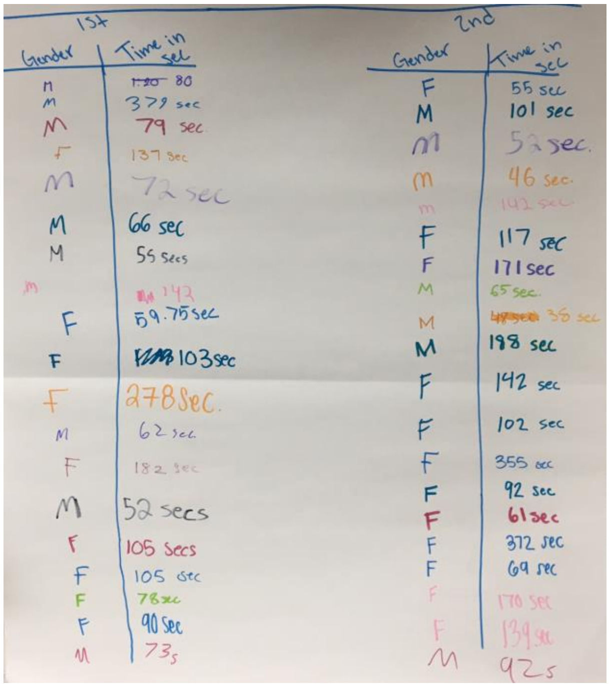


Figure 17. Data tables in Donna’s class

Teaching episode 2: A PD activity that Donna enacted in her classroom

This episode was chosen from the third observation of Donna. In this class, Donna focused on explaining human bias in making judgments, especially in making inferences and justifying conclusions. She utilized *Estimating Ages of Famous People Activity* (Frederick & Roberts, 2016) as an initial activity leading into human’s probabilistic intuitions and misjudgments caused by lack of stochastic knowledge, such as the *Sally Clark Case*. After the Hollywood game, Donna showed two videos explaining how media interpreted statistics in a certain approach to mislead the audience and how misunderstandings of sample size could lead to wrong judgments. *Estimating Ages of Famous People Activity* was also utilized in PSPD. Following paragraphs showed how Donna enacted this PD activity in her classroom.

Donna showed Hollywood celebrities' pictures online one by one and asked students to guess each celebrity's age. This activity was available at https://goo.gl/Yq4bZD . Most students made wrong judgments and were very surprised of the big gaps between their guesses and celebrities' real ages. In the end, after showing the age activity and two other videos, Donna summarized the activity with detailed explanations on human bias and stochastic knowledge. Please see below as the episode.

> Donna: Okay. In the famous people age activity, I did ask you to judge a book by the cover, named ages. If you didn't know the person, you are making judgments just by whether they look younger or they look elder. That was how you made judgments. What other things could come to play with that?
>
> Students: (in low voice) If we know better about them ahead.
>
> Donna: Yes, if you have prior knowledge of them. The more you know, the less likely you made mistakes. If you are a big fan, you may even know when they were born. So, there are personal influence here; there are human bias here. Bias can make wrong judgments. (In the other two videos) these are examples that show how statistics can be misinterpreted or misrepresented. There are a lot of pieces of data that you can interpret from. You must be careful when you look at statistics; we must be clear and smart about what we can generate from the data. (Episode ended.)

Estimating Ages of Famous People Activity was utilized in the PD to align with the statistics standards HSS. 8.SP.A.1: Investigate patterns of association in bivariate data. Donna utilized the same activity in her instruction; however, she used it to explain bias instead of investigating the relationship between guessed age and real age (there should be a linear relationship). Donna's action showed that she agreed that at least some PD activities supported

her in teaching; she enacted PD activities that she approved in her classroom instruction. Please see Donna's interview for her opinions and feedback on this PD.

Interview.

Donna's answers to key questions are shown in Table 14.

Table 14. *Interview Donna*

Key Questions	Donna's Answer
Where did you learn pedagogies to teach statistics and probability content?	A little bit from college; a lot from professional developments.
Did the PD help you in getting/understanding stochastic content knowledge?	No. I understand the content knowledge already.
Did the PD help you in improving pedagogies in teaching stochastic content?	Yes. I learned several activities that I can use in my instruction. The technology in PD was also inspiring, but I have no time to have students use it in class.
Did you use PD materials for teaching?	Yes. I already used two activities and plan to use them again. One is the "height and arm spam" activity; the other is the "Estimating Ages of Famous People Activity".
Do you utilize the GAISE investigation cycle in your own classroom?	Yes. Last year, I used it once for teaching statistical investigations.
What is your favorite part of this PD?	I learned from different types of activities. Also, I got a general idea of the sequence of secondary level statistics and probability.
What factors influence your self-efficacy in teaching statistics and probability the most?	Deep understanding of the content knowledge is the base; mastery in activities.
What do you expect to see in a statistics and probability PD?	Effective pedagogies, including but not limited to videos, engaging activities, and worksheets.

In Donna's opinion, this PSPD inspired her by offering diverse activities that can be utilized in her own classroom. Donna also adopted GAISE framework implicitly in the maze activity. Her point of view of this PD was aligned with her instruction actions. According to Lortie (1975), teachers usually adopt new practices when they are sure these new ideas will work. That Donna enacted PD activities in her own classroom instruction revealed that Donna believed these PD activities were meaningful pedagogies. This finding supported Fullan and Miles (1992)' conclusion that teachers expected PDs to offer specific ideas that were practical for day-to-day instruction. As Guskey (2002) argued, PDs that failed to offer practical ideas were very unlikely to succeed.

Case three: Cathy

Case descriptions.

Cathy has been teaching middle school mathematics in public schools for nine years. In May of 2017, which was the end of spring term, Cathy taught probability and statistics for advanced students in an accelerated program. There were 16 students in Cathy's class. Cathy uploaded PowerPoints and video to Google classroom 24 hours before classes. Students were required to watch the video at home and complete the practice in groups in class. Therefore, there was no direct lecturing in Cathy's classroom. During the whole class, students worked in groups and had access to Google classroom via Chromebooks. Cathy walked around to provide directions to each group. Due to IRB regulations, students cannot be shown in videos. Therefore, videos of Cathy's instruction videos were not available. MQI video analysis did not apply to Cathy's case. Researcher took audio-recordings to record discussions between Cathy and students. Teaching episodes and interview of Cathy are available in the next paragraphs.

Teaching episode.

This teaching episode was chosen from the third observation of Cathy. In this class, Cathy's instruction goals were aligned with the standards 6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots. Given three different data sets of test scores, students were required to display data in box plots. Students also needed to compose statements of data sets based on box plots, such as range, quartiles, outliers, etc. Figure 18 showed the work done by student team A. Cathy stopped in front of team A and asked students to explain.

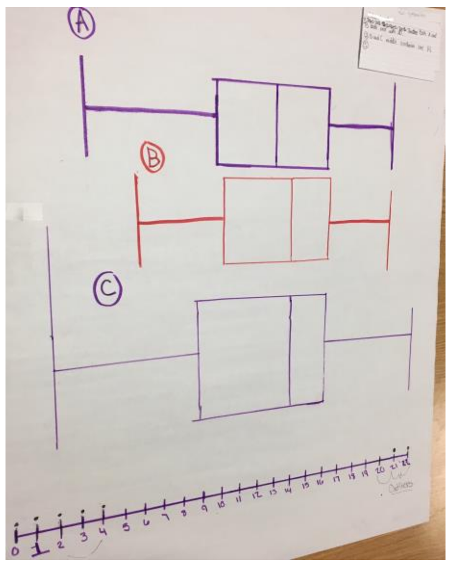


Figure 18. Box-plots made by team A

Cathy: Talk to me about your box. Based on your box, what can you tell me about the actual test scores? (students were quiet) what controls where your box is?

Student I: The middle, or the median (student paused) the second quartile.

Cathy: What does the median tell you?

Student II: (Pointing at Box B and C) These two have the same median, 14. (Pointing at Box A) this median, 13, is smaller than the other two.

Cathy: So, you know what the middles are. What does the middle tell you about the upper half of the scores?

Cathy paused. Students discussed and whispered.

Cathy: What does 13 represent? If you cut something down in the middle, what does the middle tell you about the upper half?

Student III: Half of the scores are higher? than 13?

Cathy: Yes, what does the 13 tell you vs. the 14 tell you?

Student III: (Pointing at Box B and C) half of the scores here are bigger than 13. Here, (pointing at Box A) half of the scores here are bigger than 14. (Episode ended.)

Although Cathy's instruction could not be evaluated by using MQI framework, this teaching episode could still prove that Cathy's classroom teaching matched requirements of being high quality instruction. To be specific, Cathy's instruction was related to math, was highly rich in math, was medium in working with students, had no errors, and was highly aligned with Common Core student practices. Different from Tina and Donna, Cathy utilized Google classroom via Chromebook so that students could revisit knowledge whenever needed. Meanwhile, Cathy did not provide direct answers to students; rather, she asked several questions to guide students, instead of lecturing. Cathy taught as a facilitator instead of a lecturer. In other words, her instruction in the classroom was heuristic, meaning that she supported students to explore and discover knowledge; she did not tell students the correct answers directly. While Cathy did not utilize any activities from this PD, content in her class was covered by PD activities.

Interview.

An interview was conducted to see to what extent Cathy believed that this PD helped her in teaching statistics and probability. Also, the interview was conducted to collect teachers' feedback on this PD. Cathy's answers to key questions were shown in table 15.

Table 15. *Interview Cathy*

Key Questions	Cathy's Answer
Where did you learn pedagogies to teach statistics and probability content?	I learned from a lot of professional developments, such as the PSPD. I took hundreds of hours of PDs. I also learned a lot from master teachers and colleagues. My business background helped me also because I can connect math with practice.
Did the PD help you in getting/understanding stochastic content knowledge?	A little bit. I understand most of the content knowledge already. I took a lot of business courses in college that had a lot of statistics inside.
Did the PD help you in improving pedagogies in teaching stochastic content?	Yes, the PD gave me good ideas, but I have to rethink about it to use it in my own classroom.
Do you use PD materials for teaching?	No. Most of my regular students are two years below their grade. These activities and ideas from this PD cannot be used for my regular students. Also, BLAST is not available in Curriculum Engine; it is not convent for me to revisit these PD materials.
Did you utilize the GAISE investigation cycle in your own classroom?	Yes. I did not show students this cycle, but that is what we do in class. I like students to understand where the information come from, what's the numbers look like, what's the numbers interpret. If students do not collect data by themselves, they do not even know what these data mean. When students compose their own statistical questions, they understand what they information are collecting. They do the survey, interview, analyze the data, and discover the finding.
What is your favorite part of this PD?	I always pay attention to content in lower grades so that I can build my teaching on it. From this PD, I got an idea of the sequence of secondary level statistics and probability, which was very helpful. I also communicated with other PD participants and get information about their struggling and experience.
What factors influence your self-efficacy in teaching statistics and probability the most?	Content knowledge, connections among contents, and teaching methods to teach these content.
What do you expect to see in a statistics and probability PD?	Sequence. I hope PDs can show the sequence of statistics and probability from elementary school to high school because math learning is brick by brick, like a scaffolding. I would like to see the connections among contents.

In Cathy's opinion, activities from PSPD were not attractive because these activities were too advanced for her regular students. Although she analyzed these activities carefully during PD, she did not revisit these activities. Also, Cathy usually hunted for activities in Curriculum Engine, which is a school district supported online teaching resource website; however, BLAST (the 2nd generation of PSPD) was no longer available in Curriculum Engine. This default judgment of PSPD activities and inconvenience to reach BLAST may lead to ignorance of usable PSPD activities when teaching advanced students at the end of semester.

For PSPD, Cathy most appreciated that 1), this PD showed all the statistics and probability content from Grade 6 to Geometry in sequence, providing her a connection of contents spanning different levels; 2), she was inspired by communicating and discussing her struggles and experiences with other PD teachers. Like Donna, Cathy also agreed with GAISE framework and adopted the four steps in GAISE framework in her classroom, implicitly.

Different from Donna's classroom and Tina's classroom, students in Cathy's classroom worked in small groups and Cathy facilitated students by asking questions to stimulate mathematical thinking, making her classroom student-led instead of teacher-led. Cathy had an open mind to new pedagogies and she was an active participant in all kinds of PDs, as she said, "I have participated in hundreds of hours of PDs…I don't have to". Cathy admitted that numerous PDs she participated previously helped her build a student-led classroom. Cathy was inspired by PD content; however, Cathy did not utilize the activities from PSPD due to students' low-level. Just as Lortie (1975) said, teachers usually adopt new practices when they are sure these new ideas will work.

Summary of three cases.

A cross-case data analysis is shown in table 16.

Table 16. *Cross Case Data Presentation*

	Tina	Donna	Cathy
Type	Content knowledge and self-efficacy decreased.	Content knowledge increased and self-efficacy decreased.	Content knowledge and self-efficacy increased.
MQI	High-quality math instruction with two nonstandard-usage of terminologies.	High-quality math instruction.	High-quality math instruction (flipped classroom).
PSPD content enactment	No enactment of PSPD activities.	Utilized two PSPD activities; utilized GAISE based activity; utilized GAISE framework once implicitly.	No enactment of PSPD activities; utilized GAISE framework once implicitly.
Interview: content knowledge and pedagogy sources	1. Teaching practice improved content knowledge and pedagogy. 2. PDs did not play an important role in building pedagogies. 3. Self-training was the major source of pedagogies.	1. PD was the major source of content knowledge. 2. PDs and online sources were main sources of pedagogies.	1. PDs and college courses were main sources of content knowledge. 2. PDs and online sources were main sources of pedagogies.
Interview: Enactment of PD components.	Tina did not use PD content. Tina already had preferred teaching materials.	Donna utilized PD content. Donna said, "I learned several activities that I can use in my instruction."	Cathy did not use PD activities. Cathy utilized GAISE investigation cycle once. Cathy said, "most of my regular students are two years below their grade."

These three teachers represented three diverse types of participants' change after PSPD. Their similarities were that they all provided high-quality mathematics instruction, and their instruction was all activity-based. Their main differences were 1), how much they attributed PDs in strengthening their content knowledge and pedagogy knowledge. Tina did not give PDs much credit; Donna believed that PDs and online sources supported her the most; and among the three,

Cathy was the one that gave PDs the most credit and showed enthusiasm in participating all kinds of PDs, including PSPD. As Cathy said, "I learned from a lot of professional developments, such as the PSPD. I take hundreds of hours of PDs, which is not required"; 2) how they enacted PD components in classroom teaching. Tina did not use PSPD components because she already had preferred materials. Donna utilized two activities from PSPD that could fit in her teaching. Cathy did not use PSPD content because her students were low-level.

Section III: NPD math teachers' feedback

The term NPD stands for *non-PD participants*. Data in this section was utilized to answer the third research question: *In what ways do non-PD secondary mathematics teachers incorporate BLAST (website: Bring Learning and Standards Together) into their teaching?* To answer this question, NPD teachers' pre-and post-content knowledge test (LOCUS), pre-and post-self-efficacy in teaching statistics and probability (SETS), pre-and post-lesson plans, feedback on reading the 2nd generation of PD (called BLAST) were collected. All NPD teachers were college students taking math education courses at a university in the southwest United States.

LOCUS test and SETS survey

LOCUS test. Q-Q plot was conducted and showed that the population distribution of pre- and post-LOCUS test score difference was not normal. Therefore, a repeated-measure t-test was not carried out to examine non-PD teachers' content knowledge change (Gravetter & Wallnau, 2013, p.330). Descriptive statistical data analysis was conducted to determine participants' content knowledge change.

The full marks for the LOCUS test was 100 points. The mean performance in pre-LOCUS test was 66.65, the standard deviation was 14.59. The mean performance in post-

LOCUS test was 68.88, and the standard deviation was 18.50. Among seventeen post-and pre-LOCUS performance differences, the mean was 2.76, the median was 0, the mode was 10, and the standard deviation was 19.07. By comparing the pre-and post-tests mean scores, participants' LOCUS performance had an increase of 2.76 points (out of 100).

A paired-samples t-test was conducted to compare NPD teachers' self-efficacy change (SETS) before and after studying and discussing content in BLAST. There was a significant difference in the SETS self-ranking scores for before-BLAST (M=4.070, SD=0.878) and after-BLAST (M=4.937, SD=0.682), t (16) =3.923, p = 0.001.

Table 17. *SETS Items: Level A (α=0.05, one-tail)*

Item	Question Stem	Category	Result
1	Collect data to answer a posed statistical question in contexts of interest to middle school students.	Level A	significant
2	Recognize that there will be natural variability between observations for individuals.	Level A	significant
3	Select appropriate graphical displays and numerical summaries to compare individuals to each other and an individual to a group.	Level A	significant
4	Create dotplot, stem and leaf plot, and tables (using counts) for describing distributions.	Level A	significant
5	Use dotplot, stem and leaf plot, and tables (using counts) for describing distributions.	Level A	significant
6	Create boxplots for summarizing distributions.	Level A	significant
7	Use boxplots, median, and range for describing distributions.	Level A	significant
8	Identify the association between two variables from scatterplots.	Level A	Not significant
9	Generalize a statistical result from a small group to a larger group such as the whole class.	Level A	significant
10	Recognize that statistical results may be different in another class or group.	Level A	Not significant
11	Recognize the limitation of making inference (i.e. generalization) from a classroom dataset to any population beyond the classroom.	Level A	Not significant
Note: Level A: Reading the Data. Items cited from Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics in middle grades. *Journal of Statistics Education, 23* (1), 1-20.			

Table 18. *SETS Items: Level B (α=0.05, one-tail)*

Item	Question Stem	Category	Result
12	Distinguish between a question based on data that vary and a question based on a deterministic model (for example, specific values of rate and time determines a particular value for distance in the model $d = r \times t$).	Level B	significant
13	Identify what variables to measure and how to measure them in order to address the question posed.	Level B	significant
14	Describe numerically the variability between individuals within the same group.	Level B	significant
15	Create histograms for summarizing distributions.	Level B	significant
16	Use histograms for comparing distributions.	Level B	significant
17	Compute interquartile range and five-number summaries for summarizing distributions.	Level B	significant
18	Use interquartile range, five-number summaries, and boxplots for comparing distributions.	Level B	significant
19	Recognize the role of sampling error when making conclusions based on a random sample taken from a population.	Level B	significant
20	Describe numerically the strength of association between two variables using linear models.	Level B	Not significant
21	Explain the differences between two or more groups with respect to center, spread (for example, variability), and shape.	Level B	significant
22	Recognize that a sample may or may not be representative of a larger population.	Level B	Not significant
23	Interpret measures of association.	Level B	Not significant
24	Distinguish between an observational study and a designed experiment.	Level B	significant
25	Distinguish between “association” and “cause and effect”.	Level B	significant
26	Recognize sampling variability in summary statistics such as the sample mean and the sample proportion.	Level B	significant
Note: Level B: Reading Between the Data. Items cited from Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics in middle grades. *Journal of Statistics Education, 23* (1), 1-20.			

A Pearson product-moment correlation coefficient was computed to assess the relationship between NPD teachers' content knowledge (LOCUS pre-test) and self-efficacy in teaching statistics (SET pre-survey). There was no correlation between the two variables, $r = -0.021$, $n = 17$, $p > 0.01$. There was also no correlation between LOCUS post-test performance and SETS post-survey, $r = -0.093$, $n = 17$, $p > 0.1$.

NPD teachers went through a brief period reading and discussion about content in BLAST. This training was 2.5 hours per week for two weeks. These results showed that NPD teachers had improved self-efficacy in teaching statistics and probability significantly after this short training; however, their improvement in statistical content knowledge was not significant. NPD teachers received detailed explanation of statistics and probability standards and went through activities that were usable in classroom via BLAST; it was understandable that NPD teachers improved their confidence in teaching. This training had a brief time span; therefore, it was not surprising to see the non-significant increase in content knowledge.

NPD teachers' feedback on BLAST

NPD teachers were asked to design lesson plans before and after reading and discussing content in BLAST. It was not mandatory to use BLAST content in lesson plans. These pre-and post-lesson plans were collected as another source of data to support or testify against NPD teachers' feedback on BLAST. NPD teachers' feedback on BLAST is available in table 18. According to this feedback, 53% teachers agreed that BLAST website strengthened their content knowledge; 59% teachers agreed that BLAST website improved their pedagogical knowledge; 47% teachers believed that BLAST helped improved their self-efficacy in teaching statistics and probability although there was a significant increase of self-efficacy in line with SETS survey; 65% teachers agreed that BLAST could support them in preparing instructions while only 47.1%

teachers would like to visit BLAST in the future; this gap of 17.9% may because of the unsatisfying BLAST platform—47% teachers disagreed that BLAST website was well-designed or easily accessible. Details are available in Table 19.

Table 19. *NPD teachers' feedback on BLAST*

Item	Extremely Disagree	Moderately Disagree	Slightly Disagree	Neutral	Slightly Agree	Moderately Agree	Extremely Agree	Sum
Strengthen Content knowledge	3	3	2	1	3	6		17
Strengthen Pedagogical knowledge	2			5	4	5	1	17
Strengthen self-efficacy	2	1	3	3	4	3	1	17
Help prepare lessons	1	1	2	2	5	5	1	17
Use BLAST in the future	1	1	2	5	4	3	1	17
Website well designed	2	1	4	1	4	4	1	17
	NO		YES with action			YES without action		
Utilized in post LP	8		6			3		17

53% teachers (nine in total) decided to use BLAST materials or were inspired by BLAST materials when revising post-lesson plans. Three typical reasons were generated: (1), activities were inspiring (4 out of 9). For example, one NPD teacher said she added two-way table activity in her post-lesson-plan. (2), detailed explanations of students' understanding were supportive (1 out of 9). In this teacher's opinion, "*BLAST does a good job in emulating or projecting common mistaken-steps by students in extrapolating paths of cognitive failure that do not bridge a*

student's understanding". (3), questioning strategies were inspiring (1 out of 9). This teacher decided to replace assessment sheets with assessment questions in her post-lesson plan.

To be specific, in post-lesson plans, two teachers utilized a two-way table activity for teaching conditional probability; one teacher revised and adopted *Human Box Plot* activity as part of her lesson plan; one teacher adopted the *Barbie Bungee Activity* for investigating patterns of association in bivariate data; one teacher revised student assessment from paper-pencil to oral questioning; and one teacher added one section in lesson plan to assess students' misunderstandings of mean. In total, six NPD teachers upgraded lesson plans via inspiration from BLAST.

Among the 53% teachers (nine in total), three teachers responded that they would like to use BLAST materials in post-lesson plans. They generally admitted that BLAST provided different teaching strategies that were inspiring, and they would adopt content in BLAST when needed. However, the revisions in their post-lesson plans were minor and could not relate to BLAST.

The remaining 47% teachers decided not to use BLAST materials in their lesson plans. Four typical reasons were generated: (1), they used other well-designed materials (2 out of 8). (2), CCSS standards in pre-lesson plan were not covered in BLAST (3 out of 8). (3), BLAST did not provide ready-to-use-worksheets (2 out of 8). (4), pre-lesson plan did not need to be modified (1 out of 8).

Chapter 5: Discussion

This chapter presented discussions based on data collected. Three sections of discussions were composed to answer three research questions, being followed by sections of application, research limitations, and suggestions for further research.

Summary of Research Question One and Discussions

To what extent did components of the PD program change participants' content knowledge, pedagogical knowledge, and self-efficacy in teaching statistics and probability?

Content knowledge and pedagogical knowledge change

A tailored six-day-long PD on statistics and probability was carried out for in-service secondary mathematics teachers. Pre-and post-content knowledge test (LOCUS) showed a very small increase on teachers' content knowledge; however, this increase was not statistically significant. For each GAISE investigation cycle category (formative questions, collecting data, data analysis, and interpreting data), data could not show statistically significant differences either. These findings matched follow-up teacher interviews. Tina and Donna felt that this PD did not improve their statistical and probability content knowledge; Cathy felt that this PD increased her content knowledge a little bit. This PD covered seventeen standards from Grade 6 to Geometry; due to the short-span of this PD, it was not surprising to see a non-significant result.

Although all three teachers thought that this PD did not improve their content knowledge, Donna's content knowledge performance increased 10 points (out of 100 points) and Cathy's content knowledge performance increased 6 points (out of 100 points). This increase may be explained by their attitude about PD materials. Donna utilized PD materials several times in her classroom and Cathy avoided these activities considering her students' low-level. These

decisions cannot be made unless they both actively participated in PD activities, analyzed these activities carefully and seriously thought about taking these activities into practice. This finding supported Desimone (2009)'s argument that it made a PD program effective if teachers took an active role in the work. In other words, Donna and Cathy did study these activities in depth during PD and took an active role. Although they did not feel that they learned any additional content knowledge, these PD materials still strengthened or upgraded their existing content knowledge without being noticed.

Teachers' attribution of pedagogical knowledge source may have a positive relationship with their active participation in this PD. Both Donna and Cathy believed that multiple PDs in which they participated previously had helped them considerably in real classroom teaching. It would be reasonable to link their positive attitude of PD and their active learning in PD together; possibly, their active learning in PD resulted in teachers' content knowledge increase. On the contrary, different from Donna and Cathy, Tina did not give PD much credit in her construction of pedagogies. When being asked about where she learned pedagogies all along, Tina did not put professional developments as an important resource. As Tina said, "Professional developments usually help a little bit. However, I learned most by teaching myself." Tina's attitude of PD may result in not-sufficient effort devoted to PD content, which may be one part of the reasons that caused the 4 points of decrease in Tina's content knowledge performance. In sum, a possible explanation of PD participants' content knowledge improvement and pedagogical knowledge change was shown in figure 19.

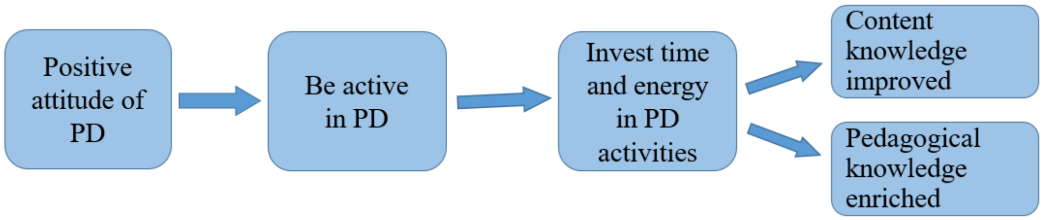


Figure 19. Content knowledge and pedagogical knowledge change

Teacher self-efficacy change

After this PD, participants' teacher self-efficacy improved significantly in general. For example, teachers improved self-efficacy in teaching content that was related with random and random samples (item 19, GAISE level B). This general improvement supported previous research finding that professional development that included content knowledge and pedagogy training could improve teachers' self-efficacy in teaching mathematics (Ingvarson et al., 2005; Watson, 2006).

However, Tina, Donna, and Cathy represented diverse types of teacher self-efficacy change. Tina's content knowledge and teacher self-efficacy both decreased. Donna's content knowledge increased while teacher self-efficacy decreased, and Cathy's content knowledge and teacher self-efficacy both increased. Please see Table 20 for details.

Table 20. *Teachers' change and factors influence self-efficacy*

	Tina	Donna	Cathy
Content knowledge	-4, decreased	+10, increased	+6, increased
Teacher self-efficacy	-2, decreased	-15, decreased	+3, increased
Factors influence self-efficacy (based on interview)	Experience (number of times teaching the same course).	Deep understanding of the content knowledge; be familiar with activities.	Content knowledge, connections among contents, and teaching methods.
Note: LOCUS full marks: 100 points; SETS full points: 156 points.			

According to interviews, these three teachers held different opinions on what factors influenced their self-efficacy in teaching statistics and probability. Tina believed that teaching practice strengthened her teacher self-efficacy in teaching statistics, without giving PDs credit. This answer was consistent with her attitude of PDs; as she mentioned, she learned most of her teaching methods via self-training. This short-span PD did not give her opportunity to practice teaching activities. Meanwhile, PD activities and assessments challenged participants' existing content knowledge and pedagogies. Considering this context, it was understandable that Tina's self-efficacy decreased after this PD. This finding supported findings from prior studies that showed strong correlations between teacher self-efficacy and instructional practices (Day, 2016). This finding was also consistent with results from prior studies that showed math teachers expected "performance experience" (Bandura, 1997) to improve self-efficacy in teaching (Boyd, Foster, Smith, & Boyd, 2014). Tina's opinions could be explained as figure 20.

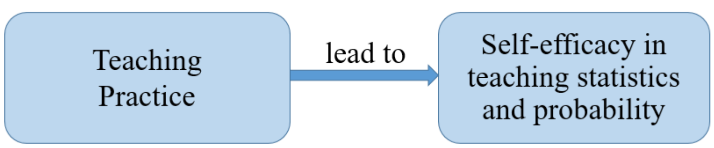


Figure 20. Teaching practice lead to self-efficacy

In Donna's opinion, to improve self-efficacy in teaching statistics and probability, she needed to be familiar with activities or teaching materials. In her classroom, Donna enacted two PD activities. However, Donna was not familiar with these activities. One evidence was that Donna could not recall the name of *Estimating Ages of Famous People Activity* and asked friends for this resource two days before she carried it out. She did not recall when or where she saw this

activity either, although she kept a memory that this Hollywood celebrity age activity was engaging and revealed human misjudgments efficiently. Considering Donna's opinions of how self-efficacy in teaching could improve, it was understandable that her self-efficacy dropped since she was not familiar with PD activities.

Donna's content knowledge increased while her self-efficacy decreased; this finding was opposite to prior studies that showed teachers' self-efficacy in teaching was highly related to teachers' self-efficacy in math (Ünlü & Ertekin, 2013). However, since Donna was not familiar with pedagogies provided in PD and her self-efficacy in teaching dropped after PD, this finding endorsed theories of teachers' pedagogical content knowledge (Ball, Thames, & Phelps, 2008; Groth, 2013). Donna's case could be explained in figure 21.

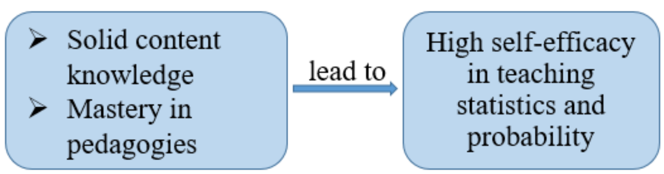


Figure 21. Mastery in content knowledge and pedagogies lead to self-efficacy

In Cathy's opinion, it could increase her self-efficacy in teaching statistics to have a deep understanding of content knowledge, connections among content, and teaching methods to teach this content. After this PD, Cathy's LOCUS test performance increased, which could be interpreted as she had a better understanding of content knowledge. In this PD, Cathy also reviewed the sequence of statistics and probability in middle school level (as she expected), providing her the connection among content. In other words, PD strengthened her content knowledge and pedagogical content knowledge, which was what she desired in order to improve

teacher self-efficacy. Considering these aspects, it was understandable that Cathy's teacher self-efficacy increased at the end of PD, along with her content knowledge. This finding was consistent with prior study that teacher content knowledge and pedagogical knowledge are highly related to teacher self-efficacy (Gresham, 2008; Huinker & Madison, 1997; Swars, 2005; Iyer & Wang, 2013).

Meanwhile, Cathy held a positive attitude of PDs; as she said, she adopted teaching and learning theories and activities from PD to upgrade instruction. She participated in hundreds of hours of PDs, which were way more than required. Her positive attitude of PD may also assist in her improvement in content knowledge and self-efficacy in teaching. Cathy's case could be explained as figure 22.

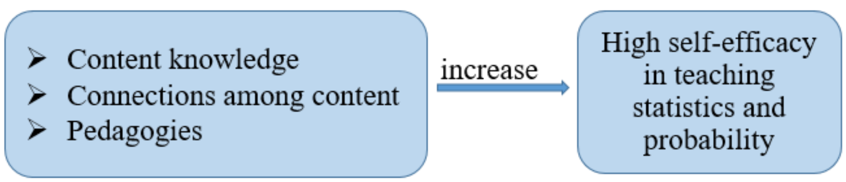


Figure 22. Deep understanding of CK and PCK lead to self-efficacy

Summary of Research Question Two and Discussions

To what extent are the components of the PD program enacted in the classrooms at school?

General Instruction.

This section discussed Tina and Donna's classroom instruction by using MQI framework. PSPD aimed at delivering high-quality CCSSM aligned activities to improve teachers' content knowledge and pedagogical content knowledge. This PD followed MQI framework implicitly:

PD activities were connected to mathematics, class was rich in mathematics, instructor worked with participants and mathematics, there were very few or no instruction errors, and PD content was aligned with CCSSM practice.

Instructional video analysis revealed that Tina and Donna's instruction possessed high-quality per MQI measurement scales. Due to the absence of contrast, this finding cannot be interpreted as PD improved teachers' instruction quality or helped teachers keep high quality instruction. However, this finding showed that instruction in teachers' classroom and training in this PD followed the same guidelines; teachers and PD trainers put effort in meeting the same standards. Especially, in PSPD training, teachers were not taught or required to use MQI framework as instruction standards. It was teachers' independent judgments that led them to fulfill MQI framework of high-quality instruction standards.

That teachers shared the same instruction standards with PD and teachers provided high quality classroom instruction provided a meaningful background for the analysis of PD activities application in teachers' teaching. Under this context, when discussing the reasons why teachers utilized or discarded PD activities, researchers had the confidence to argue that these teachers' judgments were professional and expert. Findings and discussions based on data from these professional teachers are enlightening since these teachers were experienced teachers who followed high standards and offered high-quality instruction.

Enact PD activities in classroom

Tina did not have high expectations of learning teaching pedagogies in PDs, as she said, she learned most pedagogies by self-training. Tina's feedback on PD activities were positive---this PD opened her eyes by showing her different approaches how statistics and probability were taught. However, when she taught experiment outcome, sample space, and probability, which

were covered in PSPD, Tina did not utilize PD materials or worksheets; as she mentioned, she already had teaching materials that she preferred. The materials that she utilized were one part of a series of activities instead of several single activities; she used this series of activities for an entire chapter.

Tina did not address whether she made comparison between PD activities and chosen activities before she decided what to use. It was logical to deduce that she trusted materials that she knew would work more than new materials learning from a short-spanned PD since Tina said she preferred the materials with which she was familiar. As previous researches showed, teachers seem reluctant to adopt new practices unless they are sure they can make these new ideas work (Lortie, 1975). It was understandable that teachers did not want to take extra time and effort to be familiar with new teaching strategies or activities.

Different from Tina, Donna looked forward to PDs that could present effective teaching pedagogies, such as PSPD. Donna utilized parts of PD activities in her classroom instruction. She also utilized GAISE framework based activities. This finding supported Fullan and Miles (1992)' conclusion that teachers expected PDs to offer specific ideas that were practical for day-to-day instruction. As Guskey (2002) argued, PDs that failed to offer practical ideas were unlikely to succeed (p.382). This finding was also aligned with prior findings that showed "professional development focused on specific instructional practices increases teachers' use of those practices in the classroom" (Desimone, Porter, Garet, Yoon, & Birman, 2002, p.81).

According to Lortie (1975), teachers usually adopt new practices when they are sure these new ideas will work. That Donna enacted PD activities in her own classroom instruction revealed that Donna affirmed these PD activities were meaningful, practical pedagogies. According to researchers' observation of PD, Donna was highly active and fully involved in each

activity. This might be because Donna expected to learn from PD and explored each possible usable activity and therefore she was motivated to be active in PD. If this link between self-motivation and PD-involvement was true for Donna, then Donna's case aligned with prior studies which showed highly motivated teachers were more likely to engage in PD and implement PD content in their classrooms (Cave & Mulloy, 2010; Schieb & Karabenick, 2011).

Cathy's flipped classroom did not utilize PD activities; she did not revisit PD materials because she felt that her regular students' achievement levels were too low to be exposed to these activities. However, she still led regular students through the four steps in GAISE investigation cycle to give students an idea of how statistical investigation should be. Cathy was very positive in attending PDs to improve her teaching. She liked to bring innovative ideas to her classroom when applicable. Her case, again, aligned with prior findings and arguments that teachers preferred to adopt pedagogies that fit their need or could improve students' performance (Lortie,1975; Guskey, 2002). In general, from these three cases, it was possible to generate that PD quality, teachers' self-motivation to learn from PD, and teachers' school conditions served as three main aspects that influence teachers' implementation of PD activities. To be specific, among these three cases, PD quality referred to whether PD provided practical materials, school conditions included student level (case of Cathy) and quality and localization of other competitive teaching materials (case of Tina). Please see figure 23 as an illustration.

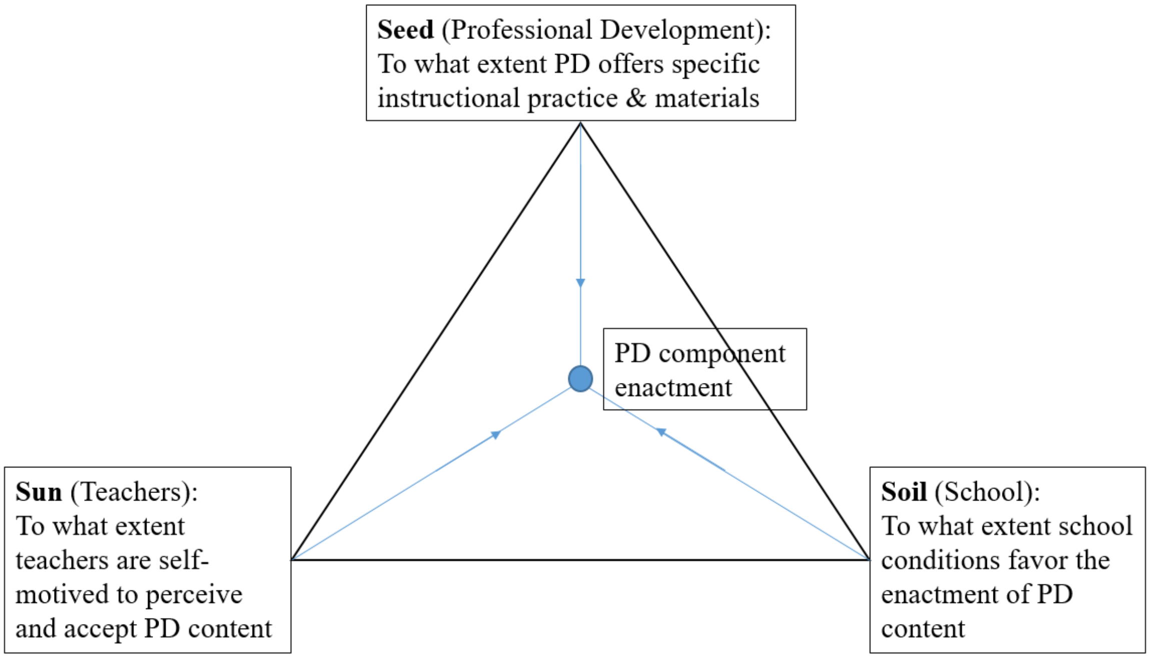


Figure 23. Factors that influence PD components enactment

Summary of Research Question Three and Discussions

In what ways do non-PD secondary mathematics teachers incorporate BLAST (website: Bring Learning and Standards Together) into their teaching?

NPD teachers' content knowledge increased slightly (2.76 out of 100) after reading and analyzing instruction materials in BLAST. However, this increase could not be proven to be statistically significant. NPD teachers' self-efficacy improved significantly in general. No correlation was found between teachers' content knowledge and self-efficacy in teaching statistics and probability. Among seventeen NPD teachers, three teachers (18%) agreed that BLAST provided meaningful materials, however, they did not utilize BLAST content into their lesson plans; six teachers (35%) utilized BLAST materials in their lesson plans; eight teachers

(47%) did not utilize BLAST materials mainly because BLAST website did not cover the standards that NPD teachers were interested in. Please see figure 24 for the distribution.

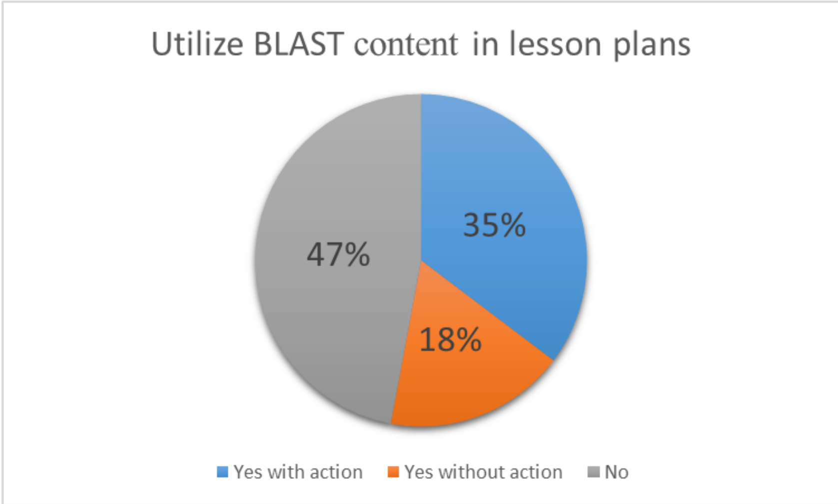


Figure 24. NPD teachers' decisions on utilizing BLAST materials

In sum, six out of seventeen NPD teachers utilized BLAST materials in lesson plans and three enactment types were generated: (1), utilized activities; (2), utilized students' misconceptions; and (3), utilized questioning strategies. Please see figure 25 as how BLAST materials were used. Teachers' feedback on BLAST provided meaningful insights on how to spread the influence of BLAST.

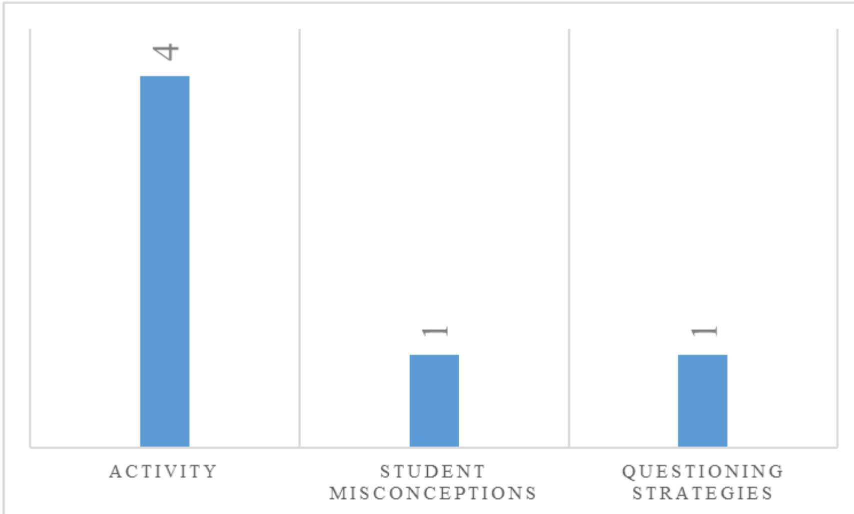


Figure 25. NPD teachers' enactment of BLAST materials

According to NPD teachers' feedback on BLAST, there were several factors that could be considered to improve the influence of BLAST. First, keep high quality of activities and supply ready-to-use worksheets; second, provide detailed information of students' typical misconceptions; third, make a list of questioning-examples instead of general questioning strategies; fourth, ensure coverage of all statistics and probability standards; and, fifth, make website design user-friendly. In addition, this website should have been easy to find. However, BLAST was not available in Curriculum Engine, which was the most popular platform that local teachers use. With these inputs, BLAST would receive more attention and keep improving. In sum, NPD teachers expected BLAST website should be easy to access and the materials inside should be practical without extensive prep-time needed.

Implications

The research findings augmented current knowledge of secondary in-service and pre-service math teachers' understandings of statistics and probability and teacher self-efficacy in

teaching secondary level statistics and probability. By analyzing three cases, this study provided insight into the factors that influence teachers' enactment of PD content into classroom instruction. Additional cognizance included experience, feedback, and suggestions on how to increase PD influence after a face-to-face training by utilizing an online website.

Limitations of this Study

Participants were chosen by convenience and the sample size was small, therefore, samples in this study could not be used to represent teacher population in the same school district or beyond this school district. This study was also limited to teachers who completed this specific PSPD or read and analyzed materials in BLAST. Findings in this study cannot be generalized beyond the sample and the specific PD.

Recommendations for Future Research

Three main directions for extending future research are recommended based on the findings of this study. The first direction would be to continue examining in-service and pre-service secondary mathematics teachers' self-efficacy in teaching statistics and probability. The second direction would be to focus on factors that influence teachers' self-efficacy in teaching statistics and probability. In this study, a moderate correlation between teacher content knowledge and teacher self-efficacy was found among PD participants who were in-service math teachers. However, no correlation was found among NPD participants who were pre-service or junior math teachers. More research is needed to explore reasons that led to the different findings of correlations. Two teachers in this study emphasized that pedagogical content knowledge played a key role in building their teacher self-efficacy. Further empirical research is needed to explore or prove the relationship between pedagogical content knowledge and teacher self-efficacy. The third direction would be to focus on designing statistics and probability PDs that

are aligned with CCSS and ready-to-use, focusing on specific instructional practices. As prior studies revealed, in-service math teachers need PDs to increase their knowledge in teaching statistics and probability (Franklin et al., 2015) with PDs that are practical to increase teachers' use of those PD materials in real classrooms (Desimone, Porter, Garet, Yoon, & Birman, 2002, p.81).

Appendix A: Activity One

Develop a profound understanding of Formulating a Statistical Question

Part 1: (5 minutes) Pre-assessment and Introduction of related Common Core State Standards---6.SP1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers (CCSS, 2010, p.45).

Part 2: Introduction of content.

(10 minutes) Participants will be shown 10 different questions (see *Intervention One, Sheet One*) and be asked to identify whether each question is a statistical question or not. Instructor will go over each question with participants. In this activity, instructors aim at supporting participants understanding regarding the rubric of statistical questions: “A well-written statistical questions refers to a population of interest, a measurement of interest, and anticipates answers that vary.” (Hopfensperger et al., 2012, p.20).

Assist standards: “First, statistical questions address issues related to a group and not an individual. Second, statistical questions do not anticipate deterministic answers, but answers based on data that have variability” (Kader, Jacobbe, Wilson, & Zbiek, 2013, p.94). “The formulation of a statistics question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary…The anticipation of variability is the basis for understanding the statistics question distinction (Franklin, et al., 2007, p.11).

Part 3: Practice.

(20 minutes) Participants will be divided into groups of five with lists of questions. For each question, participants work together to decide whether if it is a statistical question or not. If the question is a statistical question, participants need to specify the population, measurement

and variability; if the question is not a statistical question, participants need to explain why and amend the question into a statistical question. Please see *Intervention One*, *Sheet Two* for the worksheet which is designed based on *Bring the Gap* (Hopfensperger et al., 2012) and GAISE (Franklin, et al., 2007). Instructors will then explain answers and review the rubric of statistical questions again.

Part 4: post-assessment.

(5 minutes) Participants will be individually given two questions to determine whether the questions are statistical questions or not with reasons, individually.

Activity One

Are these questions statistical questions?

1. How tall is Alan Green who is in our class (Classroom 2, Grade 6, Blue Valley high school)?
2. How tall are the students in our class (Classroom 2, Grade 6, Blue Valley high school)?
3. Does our math teacher Bob White like Common Core State Standards?
4. Do people like Common Core State Standards?
5. Of all the teachers in Blue Valley high school, do people like Common Core State Standards?
6. How many words are there in this sentence “happy birthday to you”?
7. How many words are in the sentences in this book--History of Mathematics?
8. Will a Lucky Bamboo placed by the window grow taller than a Lucky Bamboo placed away from the window?
9. What do the teachers in Blue Valley high school prefer as their favorite toppings on a pizza?
10. Who was the oldest U.S. president when inaugurated?

Note: Questions were designed based on activities in *Bridging the Gap between Common Core State Standards and Teaching Statistics (*Hopfensperger, Jacobbe, Lurie, & Moreno, 2012)

Activity One (Answer)

Question: Are these questions statistical questions?

Rubric: A well-written statistical questions refers to a population of interest, a measurement of interest, and **anticipates answers that vary.**

1. How tall is Alan Green who is in Blue Valley high school?

   No. Population: there is only one subject, there is no population. Measurement variable: height. Variability: There is no variability.

2. How tall are the students in Blue Valley high school?

   Yes. Population: students in Blue Valley high school. Measurement variable: height. Variability: there can be. We expect several different heights.

3. Does Bob White like Common Core State Standards?

   No. Population: there is only one subject, there is no population. Measurement variable: like or dislike Common Core State Standards. Variability: There is no variability.

4. Do people like Common Core State Standards?

   No. Population: the population is not clear enough; it is too general.

5. Of all the teachers in Blue Valley high school, do teachers like Common Core State Standards?

   Yes. Population: teachers in Blue Valley high school. Measurement variable: like or dislike CCSS. Variability: there can be. We expect several different attitudes.

6. How many words are there in this sentence “happy birthday to you”?

   No. Population: there is only one sentence, there is no population. Measurement variable: length of the sentence. Variability: There is no variability.

7. How many words are in the sentences in this book--History of Mathematics?

   Yes. Population: sentences in this book. Measurement variable: length of the sentences (by counting words). Variability: there can be. We expect several different lengths.

8. Will a Lucky Bamboo placed by the window grow taller than a Lucky Bamboo placed away from the window?

Yes. Population: 2 Lucky Bamboo plants. Measurement variable: plant heights. Variability: there can be. We can expect different heights. (For example, a plant is placed on the window sill. A second plant is planted in a pot that is placed away from the window sill. After six weeks, the change in height for each is measured and recorded).

9. What do the teachers in Blue Valley high school prefer as their favorite toppings on a pizza?

   Yes. Population: teachers in Blue Valley high school. Measurement variable: favorite toppings. Variability: there can be. We expect several different toppings.

10. Who was the oldest U.S. president when inaugurated?

    No. Population: there is only one subject, there is no population. Measurement variable: age of president. Variability: There is no variability; the answer is deterministic.

Note: Answers were designed based on activities in *Bridging the Gap between Common Core State Standards and Teaching Statistics* (Hopfensperger, Jacobbe, Lurie, & Moreno, 2012)

Appendix B: Activity Two

Activity Two---Schedule

Part 1: (15 minutes). Pre-assessment and Introduction of related Common Core State Standards---6.SP. Introduction. The mean measures center in the sense that it is the value that each data point would take on if the total of the data values were redistributed equally (mean = fair/equal share), and also in the sense that it is a balance point (the total distance of the data values above the mean is equal to the total distance of the data values below the mean) (CCSS, p.39; California Department of Education, 2015, p.41).

Part 2: (5 minutes). Introduction of content.

There are three levels of understanding "mean".

Level A: mean = fair share value (Elementary school, Grade 6+).

Level B: mean is the "balance point" of the data distribution (Grade 6+).

Level C: students make the distinction between the mean of a population and the mean of a sample, and are introduced to the notion of a sampling distribution and using information from a sample to make inference about the population (Grade 7+). This sequence follows a hypothetical learning trajectory in learning statistics (Franklin & Kader, 2010).

Part 3: (90 minutes) Activities.

(10 minutes) Level A activity: Mean as fair share. Instructor will go through this activity with all participants.

(25 minutes) Level B activity: balance point.

(55 minutes) Level C activity: Sample Mean vs. Population Mean.

Part 4: (10 minutes). Post--assessment.

Activity Two -- GAISE Level A Activity

Activity A: Mean as Fair Share (10 minutes)

(10 minutes) Activity A: Teacher David has 9 students. He has 54 chocolates to distribute among the 9 students. Following is how the chocolates are distributed.

Name	Amount of chocolates
Alan	2
Bob	4
Chris	5
David	5
Eva	6
Frank	7
George	8
Halen	8
Ivy	9

Question 1: Is it a fair share? Answer: No.

Question 2: What is the mean amount of chocolate that each student got? Answer: 6.

Question 3: To be fair, how many chocolates each person should get? In other words, what is the value of fair share? Answer: 6.

Question 4: Are fair share and mean equal? Answer: They are equal.

This example above is a discrete data example. This conclusion also works on continuous data. For example, we need to distribute 112.5 pounds of sugar to 9 people. What is the value of fair share? 112.5/9 =12.5.

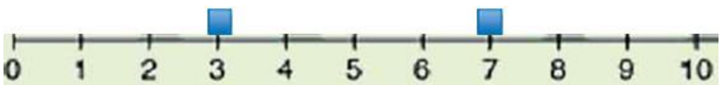
Conclusion: "(mean) is the value that each data point would take on if the total of the data values were redistributed equally (CCSS, p.39)".

Activity Two -- GAISE Level B Activity

Activity B: Mean as a balance point. (25 minutes in total)

(5 minutes) Activity B1:

Teacher Ada wants to know how much time her two kids, Cathy and Bob, spent on TV last week. Cathy said she spent 3 hours and Bob said he spent 7 hours. Ada represents these data on a lever as a dot plot.

Lever: 

Fulcrum: 

Question 1: Where should Ada put the fulcrum to make the lever a balance? Or, in other words, where does these data balance on a lever?

Answer: 5 hours.

Question 2: What is the mean of these data?

Answer: 5 hours.

Question 3: What is the distance from “3 hour” to the balance point? What is the distance from “7 hours” to the balance point? Are they equal?

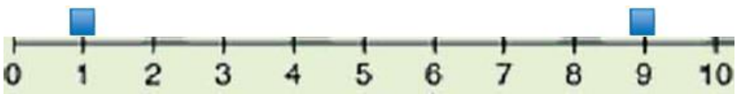
Answer: 2 hours; 2 hours; yes.

Conclusion:

1) A dot plot of the original data balance at its mean.
2) The distance between the data value bigger than the mean to mean is equal to the distance between the data value smaller than mean to mean.

(5 minutes) Activity B2:

Later, Cathy changed her answer and said she actually spent 1 hour on TV instead of 3 hours. Bob also remembered that he actually spent 9 hours on TV instead of 7 hours.

Lever: 

Fulcrum: 

Question 4: where does these new data balance on the lever?   Answer: 5 hours.

Question 5: What is the mean of these new data? Is it the same as the prior data?

Answer: 5 hours, yes.

Question 6: What is the distance from “1 hour” to the balance point? What is the distance from “9 hours” to the balance point? Are they equal? Answer: 5 – 1=4. 9 – 5=4. They are equal.

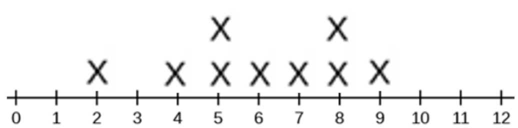
Conclusion:

1) A dot plot of the original data balance at its mean.
2) If in the process of moving data, the total movement of the data to the left of the balance point equals the total movement of the data to the right of the balance point, then the balance point does not change.
3) The sum of the distances from the balance point to points left of the balance points equals the sum of the distances from the balance point to points right of the balance point.

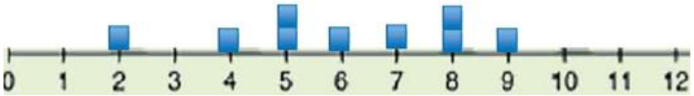
(5 minutes) Activity B3:

Look back into Activity A.

Question 7: Please draw dot-plot distributions for team 1.

Answer: Team 1: 

(10 minutes) Activity B4: Teacher Ada saw the data in Activity A1. She believed that she can put data from team 1 on a lever and make a balance too. She used cubes and a ruler and got an incomplete lever. Please see below.

Lever: 

Fulcrum: 

Question 8: What is the mean for this group of data?

Answer: 6.

Question 9: Where should she put the fulcrum to make a balance?

Answer: 6.

Question 10: What is the sum of the distances from the balance point left of the balance points?

Answer: 8.   6-2=4, 6-4=2, 6-5=1, 6-5=1.

Question 11: What is the sum of the distances from the balance point right of the balance points?

Answer: 8.   7-6=1, 8-6=2, 8-6=2, 9-6=3.

Question 12: Answer for questions 10 and 11 are the same, what does it mean?

Answer: The sum of the distances from the balance point to points left of the balance points equals the sum of the distances from the balance point to points right of the balance point.

Conclusion:

1) A dot plot of the original data balance at its mean.
2) If in the process of moving data, the total movement of the data to the left of the balance point equals the total movement of the data to the right of the balance point, then the balance point does not change.
3) The sum of the distances from the balance point to points left of the balance points equals the sum of the distances from the balance point to points right of the balance point.

Activity Two -- GAISE Level C Activity

Sample Mean vs. Population Mean (55 minutes in total)

(10 minutes) **Activity C1**: Instructor will introduce the relationship among four components: Population, sample, parameters, statistics, and inference.

- A **population** is any large collection of objects or individuals, such as Americans, students, or trees about which information is desired.
- A **sample** is a representative group drawn from the population.
- A measurable characteristic of a population, such as a mean or standard deviation, is called a **parameter;** but a measurable characteristic of a sample is called a **statistic**.
- The mean of a population is denoted by the symbol **μ**; but the mean of a sample is denoted by the symbol $\bar{x}$.
- Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. (CCSS, Grade 7)
- Statistics (is) a process for making **inferences** about population parameters based on a random sample from that population (CCSS--HSS.IC.A.1). Statistical inference means drawing conclusions about the population based on sample data.

Please see following as the relationship:

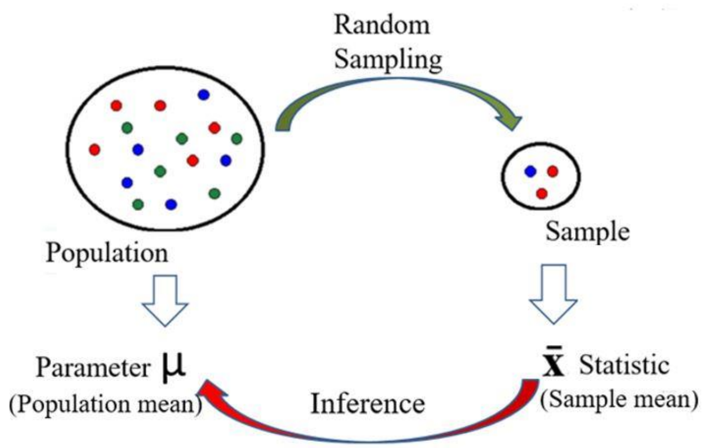


Figure 26. Population, Sample, Population mean, Sample mean.

.

(45 minutes) Activity C2: Why use random sampling?

(10 minutes) Task 1: Discussion.

Each participant will get a copy of the Gettysburg Address. There are 268 words in this address. Teacher Ada is trying to figure out the mean length of the words in the address. However, she is not sure about how to pick a sample and how big a sample should be. So, she picked up 10 words from the Gettysburg Address that she considers to be representative of the varying lengths of the words and determine the sample mean words length. The sample mean that she got is 4. Then, she made a conclusion that the population mean is 4. Do you think her conclusion is correct? Why?

Answer.

She cannot make that conclusion because her sample is not a random sample. She picked up the sample with bias. "Generalizations about a population from a sample are valid only if the sample is representative of that population…Understand that random sampling tends to produce representative samples and support valid inferences" (CCSS, Grade 7, SP.A.1).

However, why does "random sampling tend to produce representative samples and support valid inferences"? We will see the reasons by working with TinkerPlots.

(10 minutes) Task 2: Explore Sampling Distribution with TinkerPlots. Compare random-sample-mean-distribution with self-select-sample-mean-distribution.

Task 2-1**:** Please use the data in TinkerPlots sheet 1 and generate a sampling distribution. Sheet 1 contains 50 sample means. Sample size is 10. Each sample was chosen by following this procedure: researcher chose 10 words from the Gettysburg Address that the researcher considers to be representative of the varying lengths of the words and determined the sample mean words length. This procedure was repeated 50 times. Therefore, 50 samples were chosen and 50 sample means were determined. Please use TinkerPlots to show the sampling distribution. This distribution can be named "Self-Selection-Sample-Mean-Distribution".

Task 2-2: Please use the data in TinkerPlots sheet 2 and generate a sampling distribution. Sheet 2 contains 50 sample means. Sample size is 10. Each sample was chosen by following this procedure: researcher chose 10 words from the Gettysburg Address *randomly* and determined the sample mean words length. This procedure was repeated 50 times. Therefore, 50 samples were selected and 50 sample means were determined. Please use TinkerPlots to show the sampling distribution. This distribution will be called "Random-Selection-Sample-Mean-Distribution".

Task 2-3: Discussion. Gettysburg Address is short so researchers just counted the length of each word and got the <u>population mean: 4.3.</u>

Please put the two distributions together.

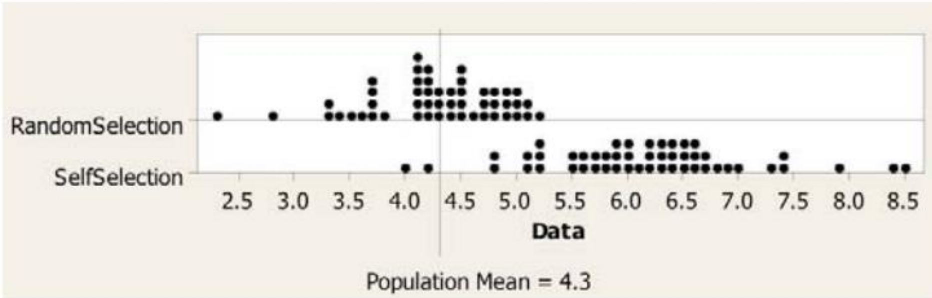


Figure 27. Means from 50 samples. N=268. n=10. (Franklin & Kader, 2010).

(10 minutes) Question 1: Which distribution has more sample means around the population mean?

Answer: Random-Selection-sample-mean-distribution.

The sampling distribution (of self-selection-sample-means) shows that most of the sample means are bigger than population mean 4.3. That is, the means from the self-selection-samples tend to overestimate the population mean.

The sampling distribution (of random-selection-sample-means) shows that most of the sample means are around the population mean 4.3. That is, the means from the random-selection-samples tend to build a distribution that have the population mean 4.3 as a central tendency.

Question 2: Now, if teacher Ada chose 1 new sample (self-selection-sample), there are also 10 words in this sample; she calculated the length of each word and got a sample mean---Ada's sample mean. Meanwhile, teacher Bob chose 1 new sample (random-selection-sample), which contained 10 words also; he calculated the length of each word and got a sample mean---Bob's sample mean.

In your opinion, between Ada's sample mean and Bob's sample mean, which one has more confidence to claim that it is close to the population mean? Or, in other words, between

two samples, Ada's sample and Bob's sample, which sample has more confidence to claim that it can be used to generate inference about the population?

Answer: Bob's sample mean. A random sample is a proper representative of the population.

(5 minutes) Task 3: How to do random sampling? Introduction of Simple Random Sampling (SRS).

Simple random sampling refers to a sampling method that has the following properties.

- The population consists of N objects.
- The sample consists of n objects.
- All possible samples of n objects are equally likely to occur.

For example, we want to select 20 students from 1000 students. We can write down each student's name on a card, put all cards in a box, shake it, and then pick cards without looking. Or, we can give each student a number and use Random Number Generator to pick 20 numbers. See Random Number Generator at http://stattrek.com/statistics/random-number-generator.aspx.

(10 minutes) Task 4: A relationship among population, sample, random sampling, inference,

Question: Please fill in the blanks with proper choices. Instructor will do it together with the participants. **Choices:** Population, population mean μ, Sample 1, Sample 2, Sample 3, Sample mean of sample 1: $\bar{x}_1$, Sample mean of sample 2: $\bar{x}_2$, Sample mean of sample 3: $\bar{x}_3$, $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$.

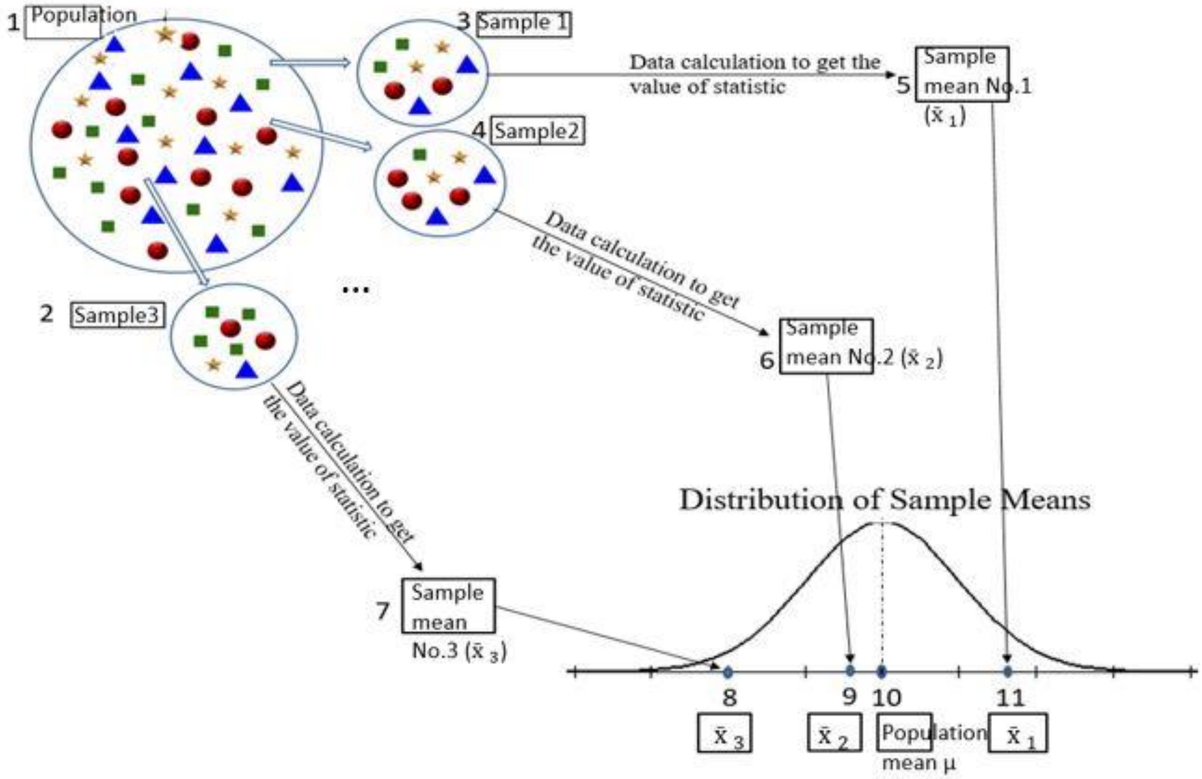


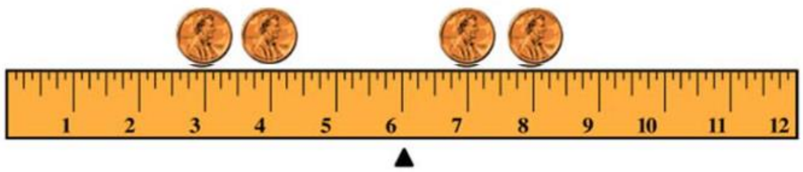
Figure 28. Distribution of Sample Means

Activity Two --- Pre-and Post-Test

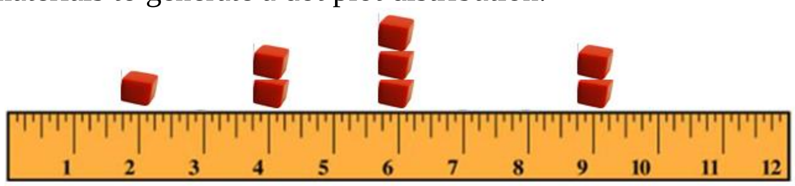
Name ______________________Grade _____________________Date _________________

1. Ada has 40 cookies. She wants to make a fair share of the cookies among her 4 kids. How many cookies each kid can get?

2. Bob has a ruler and some pennies. He wants to make a lever by using these materials. Do you think the following lever can balance? Why?



3. Cathy has 8 students. Cathy collected the amount of problems each student solved in math class today. Students solved 2, 4, 4, 6, 6, 6, 9, 9 problems. Cathy has a ruler and some cubes. She then used these materials to generate a dot plot distribution.



If she wants to put a fulcrum under the ruler and make a balance, where should she put? Why?

4. The following statement is from a local Newspaper.

ABOUT six in ten United States high school students say they could get a handgun if they wanted one; a third of them within an hour, a survey shows. The poll of 2,508 junior and senior high school students in Chicago found 15 percent had actually carried a handgun within the past 30 days, with 4 percent taking one to school. (a) Would you make any criticisms of the claims in this article? (b) If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona? Why or why not?

5. Please draw or write down the relationship among population, sample, sampling, and inference.

Appendix C: Activity Three and Four

Activity Three: Find Probability of Compound Events

Part One

Adam wants to get the probability of "getting at least one 6" when rolling two fair dices. Each dice has six surfaces. Let's help him design the experiment and find out the probability.

Question 1: What is the experiment?

Answer: the experiment is rolling two dices at the same time.

Question 2: How to calculate the probability of "getting at least one 6"?

Answer: We can repeat the experiment for a large number of times (use N to represent the total amount of trials), count the amount of the times that "getting sum 6" happened (use S to represent the total amount of successes). Then, we can calculate the **relative frequency** to represent the probability.

$$Relative\ frequency = P(getting\ at\ least\ one\ 6) = \frac{total\ amount\ of\ successes}{total\ amount\ of\ trials} = \frac{S}{N} \quad .$$

Part Two

Please roll two dice 10 times. After each roll, note whether any sixes were observed and record your results in the table below.

Roll	At least one 6 show up? Mark 1 if yes; leave blank if no.
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
Amount of yes/success =	
$\frac{amount\ of\ successes}{10} =$	
$\frac{amount\ of\ successes\ from\ everyone}{amount\ of\ trials\ from\ everyone} =$	

$Result: relative\ frequency = P(getting\ at\ least\ one\ 6) =$______.

Part Three

Without carrying out the experiment, let's calculate the theoretical probability of "getting at least one 6".

Step 1: The experiment is rolling two dices at the same time.

Step 2: Make a list of all the possible outcomes; then, count the total amount of these possible outcomes. Please make a table to list all possible outcomes. Please draw the table here:

Step 3: the theoretical probability of "getting at least one 6" =

$$P(getting\ at\ least\ one\ 6) = \frac{total\ amount\ of\ outcomes\ that\ has\ at\ least\ one\ 6}{total\ amount\ of\ possible\ outcomes} = \frac{}{}$$

Part Four

Let's upgrade activity 2 into a more professional level----using Terminologies.

Step 1: The experiment is rolling two dices at the same time.

Step 2: Please represent the <u>sample space.</u> S= {…} (No need to complete S here; please provide a few examples).

Step 3: Please represent the targeted <u>event</u>, which is "getting at least one 6". E = {…} (please complete E here).

Step 4: use the following formula

$$P(event) = \frac{number\ of\ outcomes\ in\ the\ event}{number\ of\ outcomes\ in\ the\ sample\ space}$$

Answer:

S= { }
(No need to complete S here; please provide a few examples);

E= { };

$$P(event) = \frac{number\ of\ outcomes\ in\ the\ event}{number\ of\ outcomes\ in\ the\ sample\ space} =$$

Activity Four: Design and Use a Simulation to Generate Frequencies

Example: Use random digits as a simulation tool to approximate the answer to the question: If 40% of donors have Type A blood, what is the probability that it will take at least 4 donors to find one with type A blood?

To simulate the question asked above, we could take the ten digits "0, 1, 2, 3, 4, 5, 6, 7, 8, 9" and let 40% of them represent those donors with Type A blood. So, let's say that we let the digits 0, 1, 2, 3 represent Type A Blood donors and the remainder of the digits 4, 5, 6, 7, 8, 9 represent the 60% of the donors not having Type A blood.

We want to find the probability that it will take at least 4 donors to find one with Type A blood. Thus, we want to find the probability that we <u>do not</u> get a Type A blood donor in the <u>first 3 random picks</u>. In other words, we want to find the probability that we do not generate a 0, 1, 2, or 3 within the first three picks. For example, let's say, we did a trial and picked up three numbers: (1, 5, 8), in this sample, there was "1", which represents Type A blood. It means, there was a Type A Blood donor in the first 3 random picks. We did a second trial and picked three numbers: (5, 8, 9), in this sample, there was no 0, 1, 2, or 3. It means, there is **NO** Type A blood donors.

Please use the Random Numbers Table to choose samples in the row where the row number is the same as your ID. Each sample should include 3 numbers. In a sample, if there is no 0, 1, 2, or 3, mark it under "NO"; otherwise, mark it under "YES". Please finish ten trials.

Trial	Sample (Numbers Picked)	NO Not Type A Within first three trials	YES Type A Within first three trials
Eg.	168		1
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Total			

After finishing 10 trials, please write down the total amount of "NO" on the Summary Sheet; instructor will use this Summary Sheet to calculate the relative frequencies.

This experimental answer should get closer to the theoretical answer of **0.216** as the number of trials increases.

Note: This activity was designed based on Ms. Linda Bridges' activity.

Bridges, L. (2012). 7th grade probability simulation activity. Retrieved on January 10, 2016, from https://lindabridgessite.wikispaces.com/space/content

Probability Pre-Test and Post-Test

Please write down your ID.
Please circle the correct answers or write down your answers.
There may be more than one correct answer.
Note: In this test, all coins are fair coins. Flip a coin, P (Head) =P (Tail) =0.5.

1. Flip two fair coins at the same time, what is the probability of getting one head and one tail? [7. SP.C.7.A&B]
   A. 1/2; B. 1/3; C.1/4; D. Other answer________________.

2. Flip a nickel and a dime at the same time, what is the probability of getting one nickel head and one dime tail? [7. SP.C.7.A&B]
   A .1/2; B. 1/3; C. 1/4; D. Other answer________________.

3. We know that the probability of giving birth to a boy is about 50%, the same as that of giving birth to a girl. In a certain town there are two hospitals, a small one in which there are an average of about 30 births a day and a big one in which there are an average of about 70 births a day. On a day, if there are more than 60 percent of the new born babies of the same gender, hospitals will mark the day as a UNEVEN day. The two hospitals kept record of UNEVEN days in the past two years. In which of the two hospitals were there more such UNEVEN days?
   Note: This problem is designed based on activities in Kahneman and Tversky (1972).

A. The big hospital.
B. The small hospital.
C. The numbers of UNEVEN days were equal in the two hospitals.
D. I don’t know.

5. "H" represents heads and "T" represents tails. Which of the following sequences is most likely to occur from flipping a fair coin five times?

a. HHHTT
b. THHTH
c. THTTT
d. HTHTH
e. (c) is less likely than (a), (b) and (d), but (a), (b) and (d) are equally likely.
f. All four sequences are equally likely or unlikely.
g. None of the above answers is correct.

6.

Clara and Luisa were each told to toss a coin 150 times. One did it properly. The other just made it up. They put 0 for Heads and 1 for Tails.

Clara: 01011001100101011011010001110001101101010110010001
01010011100110101100101100101100100101110110011011
01010010110010101100010011010110011101110101100011

Luisa: 10011101111010011100100111001000111011111101010101
11100000001000101001000000100011000101000000000011001
00000001111100001101010010010011111101001100011000

Question 1. Did Clara or Luisa make it up?

Question 2. How can you tell?

Cited from *A Survey of probabilistic concepts in 3000 pupils aged 11-16 years* (Green, 1983).

Answer to question 1:
Answer to question 2:

7. Adam flipped a fair coin 9 times. Heads came up all 9 times. Adam intends to flip the coin again. What is the chance of getting a head in the 10^{th} time?

a. Smaller than the chance of getting a tail

b. Equal to the chance of getting a tail

c. Greater than the chance of getting a tail

d. None of the above answers is correct. The correct answer is:

8.Adam wants to know the probability of having a girl and a boy when people have two children. He lists the possible pairs as: boy and girl, boy and boy, girl and girl. So, his answer is 1/3. Please circle the right options below.

A. He is correct.

B. He is wrong. The correct answer is ________.

C. I don't know.

Appendix D: LOCUS

Introduction of Levels of Conceptual Understanding in Statistics Test (LOCUS)

Pre-test and post-test of teachers' content knowledge of statistics and probability are part of the professional development (PSPD) and study design. The test is provided online and developed by researchers from the University of Florida. Study designers got permissions to use this test in the PSPD to detect participants' content knowledge in statistics and probability. Please see the following paragraph for detailed information of LOCUS test. The following description of the test is copied directly from LOCUS official website: https://locus.statisticseducation.org/

> LOCUS is an NSF Funded DRK12 (DRL-1118168) project focused on developing assessments of statistical understanding. These assessments measure students' understanding across levels of development as identified in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework. LOCUS assessments measure statistical understanding at two levels: Beginning/Intermediate and Intermediate/Advanced. The intent of these assessments is to provide teachers, educational leaders, assessment specialists, and researchers with a valid and reliable assessment of conceptual understanding in statistics consistent with the Common Core State Standards (CCSS). We look forward to hearing from you about what you think of these tests and the information they provide you. If you have any questions regarding the assessments or Institutional Review Board approval for the project, please contact Tim Jacobbe at the University of Florida.

Appendix E: SETS

Introduction of SETS: Self-Efficacy to Teach Statistics.

Pre-and post-self-efficacy survey will be used to detect teachers' self-efficacy in teaching statistics. This survey was one part of the professional development (PD). The survey was developed by Leigh M. Harrell-Williams (University of Memphis), M. Alejandra Sorto (Texas State University), Rebecca L. Pierce (Ball State University), Lawrence M. Lesser (The University of Texas at El Paso), and Teri J. Murphy (Northern Kentucky University). The survey is called Self-Efficacy to Teach Statistics (SETS). Researchers got permission to use this survey in this PSPD to detect participants' self-efficacy to teach statistics. Please check the following link for details about Self-Efficacy to Teach Statistics (SETS):

http://www.amstat.org/publications/jse/v23n1/harrell-williams.pdf

Appendix F: Interview

General questions.

1. Where did you learn pedagogies to teach statistics and probability content?
2. What factors influence your self-efficacy in teaching statistics and probability the most?

Questions about PSPD.

1. Did this PD help you in getting/understanding stochastic content knowledge?
2. Did this PD help you in improving pedagogies in teaching stochastic content?
3. Did you use PD materials for teaching in your classroom?
4. Did you utilize the GAISE investigation cycle in your classroom?
5. What is your favorite part of this PD?
6. What do you expect to see in a statistics and probability PD?

Questions about class.

For the content that is taught today,

1. What does prerequisite knowledge include?
2. What are you going to teach next?
3. What are students’ common misconceptions?
4. What are the most difficult parts to teach and what are the easiest parts to teach?
5. Where did you get the teaching materials from?

Thank you so much for answering these questions.

References

ABC Science. (2008). Flipping coins. Retrieved from http://www.abc.net.au/science/surfingscientist/pdf/lesson_plan19.pdf

Arshavsky, N., Edmunds, J., Charles, K., Rice, O. (2012). *STEM Classroom Observation Protocol Training Manual*. Greensboro, NC: The SERVE Center, University of North Carolina at Greensboro. Available at http://www.serve.org/STEM.aspx\

Ball, D. L., Hill, H.C, & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29*(1), p. 14-17, 20-22, 43-46.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching what makes it special? *Journal of teacher education*, *59*(5), 389-407.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman

Bar-Hillei. M. (1982). Studies of representativeness. In Kahneman, D., Slovic. P., Tversky, A. (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 69-83). Cambridge: Cambridge University Press

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational behavior and human performance*, *9*(3), 396-406. doi:10.1016/0030-5073(73)90061-5

Batanero, C., & Sánchez, E. (2005). What is the nature of high school student's conceptions and misconceptions about probability? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 260–289). Dordrecht, Netherlands: Kluwer

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, 558-567. Retrieved from http://www.ugr.es/~batanero/

Batanero, C., Arteaga, P., Ruiz, B. y Roa, R. (2010). Assessing pre-service teachers' conceptions of randomness through project work. En C, Reading (Ed.), *Proceedings of the Eight International Conference on Teaching Statistics.* [CD-ROM]. Lubjana: International Association for Statistical Education.

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment, in: J.B. Garfield, G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics.* Proceedings of the 1996 IASE Round Table Conference. International Statistical Institute, Voorburg, The Netherlands, 1997, pp. 191–205

Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of statistics Education*, *12*(1), 1-19.

Batanero, C., Godino. J. D. & Cañizares, M. J. (2005). Simulation as a tool to train pre-service school teachers. *Proceedings of First ICMI African Regional Conference.* Johannesburg: ICMI. Retrieved from http://www.ugr.es/~batanero/

Batanero, C., Gómez, E., Gea, M. M. y Contreras, J. M. (2014). Assessing and developing prospective teachers' understanding of random sequences. En U. Sproesser, S. Wessolowski, y C. Wörn (Eds.) *Daten, Zufall und der Rest der Welt - Didaktische Perspektiven zur anwendungsbezogenen Mathematik*, (pp. 1-11). Heidelberg: Springer Fachmedien Wiesbaden. Retrieved from http://www.ugr.es/~batanero/english%20papers%20and%20books.htm

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology*, *29*(1), 113-123.

Baumer, B., Cetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating A Reproducible Analysis Tool into Introductory Statistics. *Technology Innovations in Statistics Education, 8*(1). Retrieved from: http://escholarship.org/uc/item/90b2f5xh

Beauchamp, L., Klassen, R., Parsons, J., Durksen, T., & Taylor, L. (2014). Report: Teacher Efficacy Through Professional Learning. Retrieved from https://goo.gl/B24VRN

Begg, A. and Edwards, R. (1999), "Teachers' Ideas About Teaching Statistics," in *Proceedings of the 1999 Combined Conference of the Australian Association for Research in Education and the New Zealand Association for Research in Education*, Melbourne: AARE and NZARE. Retrieved from http://files.eric.ed.gov/fulltext/ED455195.pdf

Biehler, R. (2001). *Girls (tend to) watch less television than boys – Students' hypotheses and data exploration strategies in group comparison tasks*. Talk presented at LOGOS #10, Mathematics Education Unit, Department of Mathematics, the University of Auckland, New Zealand.

Borovcnik, M., and Peard, R. (1996), Probability. In Bishop, et al (eds.), *International handbook of mathematics education*, Dordrecht: Kluwer, pp. 239-288.

Boyd, W., Foster, A., Smith, J., & Boyd, W. E. (2014). Feeling good about teaching mathematics: Addressing anxiety amongst pre-service teachers. Creative Education, 5, 207-217.

Bremigan, E.G. (2003). Developing a meaningful understanding of the mean. *Mathematics Teaching in the Middle School, 9* (1), 22-26.

Bridges, L. (2012). 7th grade probability simulation activity. Retrieved on January 10, 2016, from https://lindabridgessite.wikispaces.com/space/content

Burgess, T. A. (2006). A framework for examining teacher knowledge as used in action while teaching statistics. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics ICOTS8* (2010) Invited Paper Refereed Burgess International Association of Statistical Education (IASE) www.stat.auckland.ac.nz/~iase/ education: Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS 7), Salvador, Brazil. Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications.

Carbone, R. E. (1998). *A sequel to SEQuaL: Quantitative literacy workshops and their effects on teaching practices*. Dissertation: University of Pittsburgh.

Carr, D. (2013). Giving Viewers What They Want. New York Times. Retrieved from http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?_r=0

Carter, T. A. (2005). *Knowledge and Understanding of Probability and Statistics Topics by Preservice PK-8 Teachers.* Doctoral dissertation, Texas A&M University. Texas A&M University. Available electronically from http://hdl.handle.net/1969.1/2567

Cave, A., & Mulloy, M. (2010). How do cognitive and motivational factors influence teachers' degree of program implementation? A qualitative examination of teacher perspectives. *National Forum of Educational Administration and Supervision Journal*, *27*(4).

Chance, B. L., & Rossman, A. J. (July, 2016). Using simulation to teach and learn statistics. In *Proceedings of the Seventh International Conference on Teaching Statistics*. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/17/7E1_CHAN.pdf

Chance, B. L., & Rossman, A. J. (2006). *Investigation Statistical Concepts, Applications, and Methods*, (pp. 172-189). Belmont, CA: Thomson Publishing.

Chance, Beth; Ben-Zvi, Dani; Garfield, Joan; & Medina, Elsa. (2007). The Role of Technology in Improving Student Learning of Statistics. *Technology Innovations in Statistics Education, 1*(1). Retrieved from: https://escholarship.org/uc/item/8sd2t4rr

Cherry, K. (2014). *What Is an Illusory Correlation?* Retrieved from http://psychology.about.com/od/iindex/g/illusory-correlation.htm

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, New Jersey, USA.

Cohen, J. (2012). *The Common Core State Standards: Where Do Probability and Statistics Fit In?* Retrieved from http://www.amstat.org/education/stn/pdfs/stn79.pdf

Cohen, L., Manion, L. & Morrison, K. (2000). *Research methods in education 5th edition.* London: Routledge Falmer.

Conference Board of the Mathematical Sciences (2001). The Mathematical Education of Teachers. Providence RI and Washington DC: American Mathematical Society and Mathematical Association of America. Retrieved from http://www.cbmsweb.org/MET_Document/ and http://cbmsweb.org/MET2/met2.pdf

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. (2011). *Best practices for mixed methods research in the health sciences*. Bethesda, MD: National Institutes of Health, 10.

Czernaik, C. M. (1990). A study of self-efficacy, anxiety, and science knowledge in preservice elementary teachers. *Paper presented at the National Association for Research in Science Teaching,* Atlanta, GA.

Day, K. L. (2016). *Teacher self-efficacy, instructional practices, and student achievement in mathematics: A correlational study* (Order No. 10248996). Available from ProQuest Dissertations & Theses Global. (1855473481). Retrieved from http://ezproxy.library.unlv.edu/login?url=http://search.proquest.com/docview/1855473481?accountid=3611

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), pp. 181 - 199

Desimone, L.M., Porter, A.C., Garet, M.S., Yoon, K.S., & Birman, B.F. (2002). *Educational Evaluation and Policy Analysis, 24* (2), pp. 81-112

Driskell, S., Bush, S.B., Ronau, B., Pugalee, D., & Rakes, C.R. (2016). Research in Mathematics Educational Technology: Trends In Professional Development over 40 Years of Research. Presented in 2016 AMTE.

Ekeland, I. (1988). *Mathematics and the Unexpected.* Chicago: Chicago University Press.

Filliben, J. J. (2013). Exploratory Data Analysis. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Fischbein, E. (1975). *The Intuitive Sources of Probabilistic Thinking in Children.* Dordrecht, The Netherlands: Reidel.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational studies in mathematics, 15*(1), 1-24.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, 96-105.

Fischbein, E., Barbat, I., & Minzat, I. (1971). Primary and secondary intuitions in the introduction of probability. *Educational Studies in Mathematics, 4*, 264-280.

Flores, A. (2008). The mean as the balance point: thought experiments with measuring sticks. *International Journal of Mathematical Education in Science and Technology, 39*(6), 741-748.

Florida State University. (2013). *Collaborate-Plan-Align-Learn-Motivation.* Retrieved from http://www.cpalms.org/Public/

Foley, G. D., Strayer, J. F., & Regan, B. (2010). High school mathematics teacher professional development in data analysis, probability, and statistics. *Journal of The Research Center For Educational Technology, 6*(2), 4-17. Retrieved from http://www.rcetj.org/index.php/rcetj/article/viewArticle/132

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive psychology*, *18*(3), 253-292.

Franklin, C., & Kader, G. (2010). Models of teacher preparation designed around the GAISE framework. In *8th International Conference on Teaching Statistics, Ljubljana, Slovenia.* Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_3E3_FRANKLIN.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* Alexandría: American Statistical Association.

Franklin, C.A., Bargagliotti, A.E., Case, C.A., Kader, G.D., Scheaffer, R.L., & Spangler, D.A. (2015). *The Statistical Education of Teachers (SET)*. Retrieved from http://www.amstat.org/education/SET/SET.pdf

Frederick & Roberts, D. (2016). Activity: Estimating Ages of Famous People. Retrieved from MathBits.com

Friel, S. N. & Joyner, J. (Eds.) (1997). Teach-Stat for teachers: Professional development manual. Palo Alto, CA: Dale Seymour Publications.

Gravetter, F. J. & Wallnau, L. B. (2013). Essentials of Statistics for the Behavioral Sciences. 8th edition. Wadsworth Cengage Publishing: Belmont, CA.

Garet, M. S., Poter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915–945.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for research in Mathematics Education*, 44-63.

Garfield, J., & Ahlgren, A. (1994). Student reactions to learning about probability and statistics: evaluating the quantitative literacy project. *School science and mathematics, 94*(2), 89-95.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning*, *2*(1-2), 99-125.

Garvin-Doxas, K., & Klymkowsky, M.W. (2008). Understanding randomness and its impact on student learning: Lessons learned from building the biology concept inventory. *CBE Life Sciences Education, 2008 summer, 7*(2), 227-233. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2424310/

Gencturk, C.Y. (2012). *Teachers' mathematical knowledge for teaching, instructional practices, and student outcomes*. Doctoral dissertation, University of Illinois at Urbana-Champaign. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/31173/CopurGencturk_Yasemin.pdf?sequence=1

Gilovich, T. & Savitsky, K. (2002). Like goes with like: The role of representativeness in erroneous and pseudoscientific beliefs. In T. Gilovich, D.W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 617-624). Cambridge, England: Cambridge University Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. London, UK: Weidenfeld & Nicholson.

Glesne, C. (2011). *Becoming Qualitative Researchers: An Introduction (*4th Edition). Pearson.

Green, D. R. (1981). *Probability concepts in school pupils aged 11-16 years*. Unpublished doctoral dissertation. Loughborough University of Technology. Loughborough, UK.

Green, D. R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In *Proceedings of the first international conference on teaching statistics* (Vol. 2, pp. 766-783). Sheffield, England: Teaching Statistics Trust.

Green, D. R. (1991). *A longitudinal study of pupils' probability concepts* (Tech. Rep. ME90/91). Loughborough, England: Loughborough University of Technology

Gresham, G. (2008). Mathematics anxiety and mathematics teacher efficacy in elementary pre-service teachers. *Teaching Education*, 19(3), 171-184.

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

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. Reynolds (Ed.), *The Knowledge Base for Beginning Teachers* (pp. 23-36). New York: Pergamon.

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

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8* (3/4), 381-391

Halgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor of Quantitative Methods in Psychology*, *8* (1), 23-34.

Haller, S. (1997). *Adopting Probability Curricula: The Content and Pedagogical Content Knowledge of Middle Grades Teachers*. Unpublished doctoral dissertation. University of Minnesota, Twin Cities, USA.

Halvorsen, K. T. (2010). In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics* (ICOTS8, July 2010), Ljubljana, Slovenia. Voorburg, The

Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41.

Hand, D. J. (1994). Deconstructing statistical questions. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 317-356.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer Netherlands.

Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2014), Validation of Scores from a Measure of Teachers' Self-efficacy to Teach Middle Grades Statistics. *Journal of Psychoeducational Assessment, 32* (1), 40-50.

Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics in middle grades. *Journal of Statistics Education, 23* (1), 1-20. Retrieved from https://ww2.amstat.org/publications/jse/v23n1/harrell-williams.pdf

Heck, D. J., Weiss, I. R., & Pasley, J. D. (2011). A Priority Research Agenda for Understanding the Influence of the Common Core State Standards for Mathematics. Technical Report. *Center for Mental Health in Schools at UCLA*. Retrieved from http://files.eric.ed.gov/fulltext/ED544302.pdf

Heffernan, C. (n.d.). *Intervention studies*. Retrieved from http://www.drcath.net/toolkit/intervention.html

Hellman, G. (1978, January). Randomness and reality. *In PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 79-97). Philosophy of Science Association. Chicago: University of Chicago Press

Henriques, A., & Ponte, J. P. (2014). Preparing teachers to teach statistics: Developing professional knowledge and practice. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.

Higazi, S. M. (2002). Teaching statistics using technology. *ICOTS6 Proceedings*. Retrieved from http://iase-web.org/documents/papers/icots6/5f4_soha.pdf

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794-831.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Education Research Journal, 42*(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *Elementary School Journal, 105*, 11–30

Hogarth, R.M. (2001). *Educating Intuition.* Chicago: University of Chicago.

Hong, S. B., & Trepanier-Street, M. (2004). Technology: A Tool for Knowledge Construction in a Reggio Emilia Inspired Teacher Education Program. *Early Childhood Education Journal 32*(2): 87-94.

Hopfensperger, P., Jacobbe, T., Lurie, D., & Moreno, J. (2012). *Bridging the Gap between Common Core State Standards and Teaching Statistics*. Alexandria, VA: American Statistical Association

Hoy, A. W., Hoy, W. K., & Davis, H. A. (2009). Teachers' self-efficacy beliefs. In K. R. Wentzel and A. Wigfield (Eds.) *Handbook of Motivation* (pp. 627-653). New York: Routledge Press. Retrieved from http://iase-web.org/documents/papers/icots8/ICOTS8_3E3_FRANKLIN.pdf

Huinker, D. and Madison, S. (1997), Preparing Efficacious Elementary Teachers in Science and Mathematics: The Influence of Methods Courses. *Journal of Science Teacher Education, 8*(2), 107-126.

Huinker, D. and Madison, S. (1997). Preparing efficacious elementary teachers in science and mathematics: The influence of methods courses. *Journal of Science Teacher Education*, 8(2), 107-126.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives, 13*(10), 1-26.

Inhelder, B., & Piaget, J. (1958). The Growth of Logical Thinking from Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures. Psychology Press. Routledge.

Ives, S. E. (2009). *Learning to teach probability: Relationships among pre-service teachers' beliefs and orientations, content knowledge, and pedagogical content knowledge of probability*. Unpublished doctoral degree dissertation. North Carolina State University, Raleigh, USA. Retrieved from http://repository.lib.ncsu.edu/ir/handle/1840.16/4058

Ives, S. E. (2011). The Relationship between pre-service teachers' conceptions of randomness and their pedagogical content knowledge of probability. *Mathematics Education into the 21 Century Project*. Retrieved from http://math.unipa.it/~grim/21_project/21_Charlotte_IVESPaperEdit.pdf

Iyer, N. N., & Wang, J. T. (2013). Perceptions of Elementary Pre-service Teachers' Confidence to Teach Mathematics. *Proceeding of Northeastern Educational Research Association (NERA) Annual Conference in 2013*. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1022&context=nera_2013

Jacobbe, T. (2008). Elementary school teachers' understanding of the mean and median. *Proceedings of the ICMI Study, 18.* Chicago. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Jacobbe.pdf

Jacobbe, T. (2012). Elementary school teachers' understanding of the mean and median. *International Journal of Science and Mathematics Education*, 10(5):1143–1161.

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the validity of the LOCUS assessments through an evidenced-centered design approach. In Sustainability in Statistics Education. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July 2014), Flagstaff, Arizona, USA.

Jones, G. A., & Thornton, C. A. (2006). An overview of research into the teaching and learning of probability. In Jones (Ed.), *Exploring Probability in School* (pp. 65-92). New York, NY: Springer US.

Kader, G., Jacobbe, T., Wilson, P., & Zbiek, R. M. (2013). *Developing essential understanding of statistics for teaching mathematics in grades 6 - 8*. (1st edition). National Council of Teachers of Mathematics.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kahneman, D., & Tversky, A. (1974). Subjective probability: A judgment of representativeness. In C.-A. Staël von Holstein (Ed), *The Concept of Probability in Psychological Experiments* (pp. 25-48). Dordrecht, Netherlands: Springer Netherlands.

Kalil T. (2012, March 29). *Big Data is a Big Deal*. Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal

Kenney, M. M. (2016). How does professional development improve teaching? *Review of Educational Research. 86*(4), 945–980

Kenyon, J. J. (2013). *Evaluation of Clark County School District' s Alternative Route to Licensure Program from the Program Participants' Perspectives*. Dissertation for Ph.D. degree at Florida State University. Retrieved from http://diginole.lib.fsu.edu/cgi/viewcontent.cgi?article=7809&context=etd

Kirkebøen G (2009). Decision behavior – improving expert judgment. In Williams T et al (eds). *Making Essential Choices with Scant Information*. Palgrave MacMillan: Basingstoke, UK 2009.

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

Larrick R (2004). Debiasing. In: D Koehler and N Harvey (eds). *Blackwell handbook of judgment and decision making.* (pp. 316-337). Blackwell: Malden, MA.

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.

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

Lee, H. J. (2005). Developing a Professional Development Program Model Based on Teachers' Needs. Professional educator, 27, 39-49.

Lee, H. S., & Hollebrands, K. (2008). Preparing to teach mathematics with technology: An integrated approach to developing technological pedagogical content knowledge. *Contemporary Issues in Technology and Teacher Education [Online serial], 8*(4). Retrieved from: http://www.citejournal.org/vol8/iss4/mathematics/article1.cfm

Lee, H.S., & Mojica, G. F. (2007). Teachers' use of probability experiments and simulations. *Proceedings of the Twenty-Ninth Annual Meeting of the Psychology of Mathematics Education-North American Chapter*. Lake Tahoe, UT, October.

Leek, J. (2012). The statisticians at Fox News use classic and novel graphical techniques to lead with data. Retrieved from http://simplystatistics.org/2012/11/26/the-statisticians-at-fox-news-use-classic-and-novel-graphical-techniques-to-lead-with-data/

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, *26*(6), 952.

Li, J. (2013). 5 Ways to avoid being fooled by statistics. Retrieved from http://www.iacquire.com/blog/5-ways-to-avoid-being-fooled-by-statistics

LOCUS Committee. (2015). Levels of Conceptual Understanding in Statistics Evaluation (LOCUS) test. Retrieved from https://locus.statisticseducation.org/

Lortie, D. C. (1975). School Teacher: A Sociological Study, Chicago, IL, University of Chicago Press.

Loucks-Horsely, S., Stiles, K. E., Mundry, S. E., Love, N. B., & Hewson, P. W. (2010). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School science and mathematics*, *99*(5), 258-271.

Madden, S. R. (2008). *High School Mathematics Teachers' Evolving Understanding of Comparing Distributions*. Unpublished PhD Dissertation. Western Michigan University: Kalamazoo, Michigan.

Makar, K., & Confrey, J. (2002). Comparing two distributions: Investigating secondary teachers' statistical thinking. *In Sixth International Conference on Teaching Statistics (ICOTS-6),* Cape Town, South Africa.

Malloure, M., Richardson, M, & Rogness, N. (2012). *An A-MAZE-ING Comparison.* Retrieved from http://rpdp.net/admin/images/uploads/resource_11751.pdf

Marasinghe, M. G., Meeker, W. Q., Cook, D., & Shin, T. S. (1996). Using Graphics and Simulation to Teach Statistical Concepts. *The American Statistician, 50*(4), 342–351. http://doi.org/10.2307/2684932

Marrongelle, K., Sztajn, P., & Smith, M. (2013). Scaling up professional development in an era of Common Core Standards. *Journal of Teacher Education, 64*(3), 202-211.

Mayring, P. (2000). Qualitative content analysis. Qualitative Social Research, 1 (2). Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman.

McClain, K. (2002). Supporting teachers' understanding of statistical data analysis: Learning trajectories as tools for change. In B. Phillips (Ed.), Proceedings of the Sixth International Conference on the Teaching of Statistics (On CD). Hawthorn, VIC: International Statistical Institute.

Meletiou-Mavrotheris, M., Paparistodemou, E., & Stylianou, D. (2009). Enhancing statistics instruction in elementary schools: Integrating technology in professional development. *The Montana Mathematics Enthusiast*, *6*(1), 57-78.

Meletiou-Mavrotheris, M., Paparistodemou, E., Mavrotheris, E., Azcárate, P., Serradó, A., & Cardeñoso, J. M. (2008). Teachers' professional development in statistics: The Early Statistics European project. In Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study and 2008 IASE Round Table Conference, Monterrey, México. Retrieved from http://iase-web.org/documents/papers/rt2008/T6P6_Meletiou.pdf

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics,*14, 415-429

Moore, D. S. and McCabe, G. P. (2006). *Introduction to the Practice of Statistics*, 5th edition, New York, NY: Freeman.

Morton, N. (2015). CCSD starting school short on teachers. Retrieved from http://www.reviewjournal.com/education/ccsd-starting-school-short-teachers

Mullins, D., Lepicki, T., & Glandon, A. (December 2010). A professional development evaluation framework for the Ohio ABLE System. Prepared by Center on Education and Training for Employment, the Ohio State University. Retrieved from

http://uso.edu/network/workforce/able/reference/development/PD_Eval_Framework_Report.pdf
National Center for Teacher Effectiveness. (2016a). Logistics and Training. Retrieved from http://isites.harvard.edu/icb/icb.do?keyword=mqi_training&tabgroupid=icb.tabgroup120175
National Center for Teacher Effectiveness. (2016b). Measuring MQI. Retrieved from http://isites.harvard.edu/icb/icb.do?keyword=mqi_training&pageid=icb.page394700
National Council of Teachers of Mathematics (NCTM). (2000). *Principles and Standards for School Mathematics*. Reston, VA: Authors.
National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards (CCSS). Washington, DC: Authors.
Nevada Mathematics Project Committee. (2014). Nevada Mathematics Project. Retrieved from http://nevadamathproject.com/
New York State Education Department. (2016). Engage NY. Retrieved from https://www.engageny.org/mathematics
Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review,90* (4), 339.
Nishimura, T. (2014). Effective professional development of teachers: A guide to actualizing inclusive schooling. *International Journal of Whole Schooling. 10* (1). 18-53
Pfannkuch, M., Budgett, S., Parsonage, R., & Horring, J. (2004). Comparison of data plots: Building a pedagogical framework. Paper presented at the Tenth International Congress on Mathematics Education (ICME-10), Copenhagen, Denmark, 4-11 July.
Piaget, J. (1952). *The origins of intelligence in the child* (M. Cook, Trans.). New York, NY: International Universities Press.
Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in students* (L. Leake, Jr., P. Burrell, & H. D. Fischbein, Trans). New York, NY: Norton (Original work published 1951).
Piatetsky, G. (2012). Taking misleading statistics to a new level. Retrieved from http://www.kdnuggets.com/2012/12/taking-misleading-statistics-to-a-new-level.html
Pitman, J. (1993). *Probability*. New York: Springer
Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics,12* (2), 191-204. Retrieved from http://link.springer.com/article/10.1007/BF00305621
Putney, L.G. (1997). *Collective-individual development in a fifth-grade bilingual classroom: An interactional ethnographic analysis of historicity and consequentiality.* Unpublished Dissertation, University of California, Santa Barbara.
Reaburn, R. (2008). The hospital problem revisited. Tertiary student's perceptions of a problem involving the binomial distribution, In M. Goos, R. Brown, and K. Makar (eds.) *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia*. Brisbane: MERGA. Retrieved from http://www.academia.edu/4343061/The_Hospital_Problem_Revisited._Tertiary_Students_Perceptions_of_a_Problem_Involving_the_Binomial_Distribution
Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51-65.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7* (2), 5-19. Retrieved from www.stat.auckland.ac.nz/serj/.

Rossman, A. J., & Chance, B. L. (1999). Teaching the reasoning of statistical inference: A "top ten" list. *The College Mathematics Journal*, *30*(4), 297-305. Retrieved from http://www.rossmanchance.com/papers/topten.html

Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In Vere-Jones (Ed.). *Proceedings of the third international conference on teaching statistics. 1*, 314-319. AZ Voorburg, The Netherlands: The International Statistical Institute.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10* (3). Retrieved from www.amstat.org/publications/jse/v10n3/rumsey2.html

Saldanha, L., & Thompson, P. W. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257–270. doi:10.1023/A:1023692604014

Scheaffer, R. (1988). Statistics in the schools: The past, present and future of the Quantitative Literacy Project. *In American Statistical Association 1988 Proceedings of the Section on Statistical Education* (pp. 71-78). Alexandria, VA: American Statistical Association.

Scheaffer, R. (1990). The ASA-NCTM Quantitative Literacy Project: An Overview. Proceeding in The International Conference on Teaching Statistics (ICOTS). Retrieved from http://iase-web.org/Conference_Proceedings.php?p=ICOTS_3_1990

Scheaffer, R., Tabor, J., & Hirsch, C. (2008). Statistics in the high school mathematics curriculum: Building sound reasoning under uncertain conditions. *The Mathematics Teacher*, 56-61.

Schenker, N., Davidian, M., & Rodriguez, R. (June 2013). The ASA and big data. Amstat News. Retrieved from http://magazine.amstat.org/blog/2013/06/01/the-asa-and-big-data/

Schieb, L. J., & Karabenick, S. A. (2011). Teacher Motivation and Professional Development: A Guide to Resources. Math and Science Partnership – Motivation Assessment Program, University of Michigan, Ann Arbor, MI. Retrieved on June 16, 2017, from http://mspmap.org/wp-content/uploads/2011/10/TeachMotivPD_Guide.pdf

Schreiber, J. M. (2014). Cognitive Processes Associated with the Perception of Randomness. *Journal of Educational and Developmental Psychology, 4* (1), p.84.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10* (1), 33-51. doi: 10.1002/(SICI)1099-0771(199703)10:1<33::AID-BDM244>3.0.CO;2-6

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465–494). New York: Macmillan.

Shih, J.C., Ing, M. & Tarr, J.E. (2015). Addressing measurement issues in two large-scale mathematics classroom observation protocols. In Cai, Hwang, and Middleton (ed.) *Large-Scale Studies in Mathematics Education* (pp. 363-371). Springer International Publishing.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Skorkowsky, P. (2015 Jan. 27). CCSD looks to ARL program to hire more teachers. Retrieved from http://ccsd.net/district/pat-personally/2015/jan/27/ccsd-looks-to-arl-program-to-hire-more-teachers

Smith, M. K. (2014). Why is random sampling important? Retrieved from https://www.ma.utexas.edu/users/mks/statmistakes/RandomSampleImportance.html

Sparrow, M. (2010). Probability, statistics & logical reasoning through games of strategy. Retrieved from https://faculty.unlv.edu/jbailey/2010spring/Sparrow_Probability-Final.pdf

Steele, M. & McLeod, K. (2015). *Common Core High School Math Leadership (CCHSML) Project*. Retrieved from http://uwm.edu/education/research/centers/cmser/cchsml/#a0

Steinbring, H. (1991). The concepts of chance in everyday teaching: aspects of a social epistemology of mathematical knowledge. *Educational Studies in Mathematics (22*), 503-522.

Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.). *Exploring probability in school: Challenges for teaching and learning* (pp. 345-366). New York: Springer.

Sukta, J. (2014). *Business Statistics.* Retrieved from http://gezzie2.com/SXU132/lesson12a.htm

Swars, S. L. (2005). Examining Perceptions of Mathematics Teaching Effectiveness Among Elementary Preservice Teachers with Differing Levels of Mathematics Teacher Efficacy. *Journal of Instructional Psychology, 32*(2), 139-147.

Swenson, K. A. (1998). *Middle school mathematics teachers" subject matter knowledge and pedagogical content knowledge of probability: its relations to probability instruction* (Disertación Doctoral, Universidad del Estado de Oregon, 1998). Dissertation Abstracts International, 59(02), 440A. Retrieved from http://ir.library.oregonstate.edu/xmlui/handle/1957/33752

University of California, Los Angles. (2007). *Introduction to SAS.* University of California, Los Angles: Statistical Consulting Group. Retrieved on June 5, 2016, from http://www.ats.ucla.edu/stat/sas/notes2/

Trochim, W. M.K. (2006). Types of Reliability. Retrieved from http://www.socialresearchmethods.net/kb/reltypes.php (accessed September 5, 2016)

Tschannen‐Moran, M., & McMaster, P. (2009). Sources of Self‐Efficacy: Four Professional Development Formats and Their Relationship to Self‐Efficacy and Implementation of a New Teaching Strategy. *The Elementary School Journal, 110*(2), 228-245. doi:10.1086/605771

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110. Retrieved from http://pirate.shu.edu/~hovancjo/exp_read/tversky.htm

Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Uccellini, J. C. (1996). Teaching the Mean Meaningfully. *Mathematics Teaching in the Middle School, 2*(2), 112-115.

Ünlü, M. & Ertekin, E. (2013). The relationship between mathematics teaching self-efficacy and mathematics self-efficacy. *Social and Behavioral Sciences, 106,* p. 3041-3045. Retrieved from https://doi.org/10.1016/j.sbspro.2013.12.350

Verhoeff, T. (1993). The laws of large numbers compared. Retrieved from http://www.dklevine.com/archive/strong-law.pdf

Virginia Department of Education. (2004). *Probability and Statistics for Elementary and Middle School Teachers: A Staff Development Training Program to Implement the 2001 Virginia Standards of Learning.* Retrieved from http://www.doe.virginia.gov/testing/sol/standards_docs/mathematics/2001/resources/elementary/probability_module/mprobstatentire.pdf

Watson, G. (2006). Technology professional development: Long-term effects on teacher self-Efficacy. *Technology and Teacher Education, 14*(1), 151-165.

Watson, J. M. (1998), Professional Development for Teachers of probability and Statistics: Into an Era of Technology. *International Statistical Review, 66:* 271–289. doi: 10.1111/j.1751-5823.1998.tb00373.x

Watson, J. M. (2000). Pre-service mathematics teachers' understanding of sampling: Intuition or mathematics. *Mathematics Teacher Education and Development, 2*, 121-135.

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, *4*(4), 305-337.

Watson, J. M. (2004). Developing reasoning about samples. In Dani Ben-Zvi & Joan Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer.

Watson, J. M. (2005). The probabilistic reasoning of middle school students. In G. Jones (Ed.). *Exploring probability in school: Challenges for teaching and learning* (pp. 145-169). New York: Springer.

Watson, J. M., & Moritz, J. B. (1998). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*(2), 145-168.

Watson, J. M., & Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. *The Journal of Mathematical Behavior, 19*(1), 109-136

Watson, J., and Baxter, J. (1997). Learning the Unlikely at Distance as an Information Technology Enterprise: Development and Research. In J. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 285--313). Voorburg, The Netherlands: International Statistical Institute. Available at http://www.pucrs.br/famat/viali/tic_literatura/livros/IASE.book.pdf#page=288

Wessels, H. (2014). Developing statistical knowledge for teaching of variability through professional development. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July 2014), Flagstaff, Arizona, USA. Voorburg, the Netherlands: International Statistical Institute. www.iase-web.org

Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review / Revue Internationale De Statistique*, *67* (3), 223–248. http://doi.org/10.2307/1403699

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological bulletin*, *116* (1), 117.

Wolf, K. J., Foster, D. D., & Birkenholz, R. J. (2010). The relationship between teacher self-efficacy and the professional development experience of agricultural education teachers' candidates. *Journal of Agricultural Education, 51*(4), 38.

Yin, R. K. (2003). *Applications of Case Study Research*. Newbury Park, CA: Sage Publications.
Yin, R. K. (2014). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage.

Curriculum Vitae

Date: July 2017

Lina DeVaul, Ph.D.

Department of Teaching and Learning
College of Education
University of Nevada, Las Vegas
LINA.DEVAUL@UNLV.EDU

EDUCATION

2017 August. (Anticipated Completion Date)	Ph.D. Curriculum and Instruction Emphasis: Mathematics Education Minor: Teacher Education Department of Teaching and Learning University of Nevada, Las Vegas
2010	Master of Science Emphasis: Curriculum and Teaching Methodology in Mathematics Education Department of Mathematics East China Normal University, Shanghai, China
2007	Bachelor of Science Emphasis: Mathematics and Applied Mathematics Department of Mathematics Liaocheng University, Shandong, China

PUBLICATIONS

Stohlmann, M., **DeVaul, L.**, Allen, C., Adkins, A., Ito, T., Lockett, D., & Wong, N. (2016). What is known about secondary grades mathematical modeling-a review. *Journal of Mathematics Research*, *8*(5), 12-28

Schactera, J., Shih, J., Allen, C. M., **DeVaul, L.,** Adkins, A., Ito, T., & Jo, B. (2015). Math Shelf: A Randomized Trial of a Pre-Kindergarten Tablet Number Sense Curriculum. *Early Education and Development, 26(*9*),* 1-15.

PROCEEDINGS

Stohlmann, M., **DeVaul, L.,** Allen, C., Adkins, A., Ito, T., Lockett, D., & Wong, N. (2015). *Effectiveness of Mathematical Modeling in Secondary School and Future Research Opportunities.* Paper presented at the International Community of Teachers of Mathematical Modelling and Applications (ICTMA) conference. Nottingham, UK: ICTMA-17.

Shih, J. C., Adkins, A. **DeVaul, L**, Ito, T., & Allen, C. M. (2015). Early childhood gender differences in number sense when learning with iPads. In *Proceeding of 2015 Research Council on Mathematics Learning* (pp.174-179). Las Vegas, NV: Research Council on Mathematics Learning (RCML)

Olson, T., Olson, J., Olson, M., Capen, S., Shih, J., Adkins, A., **DeVaul, L.** & Thomas, A. (2015). Exploring 1:1 Tablet Technology Settings: A Case Study of the First Year of Implementation in Middle School Mathematics Classrooms. In *Proceedings of Society for Information Technology & Teacher Education International Conference 2015* (pp. 2802-2807). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

CONFERENCES PRESENTATIONS AND POSTERS

DeVaul, L., Olson, T., & Adkins, A. (2017, April). *Utilizing Random Number Sheets to Carry Out Simulations Aligning with CCSS.* National Council of Teachers of Mathematics Annual Meeting & Exposition (NCTM). San Antonio, Texas.

DeVaul, L., & Olson, T. (2017). *Strengthening Middle School Mathematics Teachers' Knowledge of Statistics and Probability via Professional Development*. Presented at the Twenty-first Annual Conference of Association of Mathematics Teacher Educators (21st AMTE). Orlando, FL.

DeVaul, L., Stohlmann, M., & Adkins, A. (2016). *Using a Computerized Algebra Program ALEKS for Intermediate Courses.* Presented at the National Council of Teachers of Mathematics Annual Meeting & Exposition (NCTM). San Francisco, CA.

DeVaul, L., Adkins, A. Wong, N., Shih, J., & Warren, C. B. (2016). *Improve Elementary Teachers' Understanding of Base-ten Place Value System.* Presented at the Annual Conference of Research Council on Mathematics Learning (43rd RCML). Orlando, FL.

DeVaul, L. (2015). *Assessing Pre-service Mathematics Teachers' Misunderstanding of Probability.* Presented at the 42th Annual Conference of Research Council on Mathematics Learning (42nd RCML). Las Vegas, NV.

DeVaul, L. & Xu, Z.T. (2014). IT application in Chinese Elementary Mathematics Classroom. Presented at the Summer Conference of the Association of Teacher Educators (ATE), Niagara, NY.

DeVaul, L. (2013). An Analysis of Chinese Han and Mongolian Students' Mathematics Understanding. Presented at the 57th Annual Conference of the Comparative and International Education Society (CIES). New Orleans, LA, U.S.

Adkins, A., Shih, J., & **DeVaul, L.** (2015). *Lessons Learned about Preschool Children's Use of iPads.* Presented at the **N**ineteenth Annual Conference of Association of Mathematics Teacher Educators (19th AMTE). Orlando, FL.

Olson, T., Shih, J., Thomas, A., **DeVaul, L.,** and Adkins, A. (2014). iPad Apps Utilized in Middle School Mathematics Classrooms with a 1:1 Setting. Presented at the Eighteenth Annual Conference of Association of Mathematics Teacher Educators (18th AMTE). Irvine, CA, U.S.

TEACHING EXPERIENCE AND TEACHER EDUCATION PROGRAM EXPERIENCE

- ❖ Instructor: Math 95, Math 96, Teaching High School Mathematics, Teaching Middle School Mathematics, Teaching Elementary School Mathematics, Introduction to Statistical Methods, Introduction to Elementary Education, secondary school math.
- ❖ Teacher Education Program School Site Facilitator: Elementary Teacher Education Program, Rebel Academy Alternative Route to Licensure Program (Secondary).

For a detailed Curriculum Vitae, please contact Lina DeVaul via lina.devaul@unlv.edu