Advanced Placement Statistics Teaching Knowledge Assessment

By Brenna Haines

B.S. in Resource Economics, June 1998, The University of Connecticut
M.S. in Applied Economics, August 2000, The University of Delaware

A Dissertation Submitted to

The Faculty of
The Graduate School of Education and Human Development
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Education

August 31, 2014

Dissertation directed by:

Curtis Pyke
Associate Professor of Curriculum and Pedagogy

The Graduate School of Education and Human Development of The George Washington University certifies that Brenna J. Haines has passed the Final Examination for the degree of Doctor of Education as of June 5th, 2014. This is the final and approved form of the dissertation.

Advanced Placement Statistics Teaching Knowledge Assessment

Brenna J. Haines

Dissertation Research Committee

Curtis Pyke, Associate Professor of Curriculum and Pedagogy, Dissertation Director

Karen Kortecamp, Adjunct Professor of Curriculum and Pedagogy, Committee Member

Brandi Weiss, Assistant Professor of Education, Committee Member

© Copyright 2014 by Brenna J. Haines
All rights reserved

Dedication

This dissertation is dedicated to Dr. Scott Ingles Morgan.

(September 10, 1951 - June 1, 2014)

Father - Mentor - Friend

Abstract of Dissertation

Advanced Placement Statistics Teaching Knowledge Assessment

Increasing student enrollment in high-school level Advanced Placement (AP) Statistics courses necessitates the need for teachers who are knowledgeable in the subject-area. However, no benchmark has been established that describes the amount or types of teaching knowledge that is required, or even desirable, of AP Statistics teachers. More specifically, there does not exist a criterion of reference to determine if an AP Statistics teacher does or does not possess the content-specific knowledge necessary to teach the subject. Therefore, a teacher may possess sufficient knowledge to teach mathematics but be deficient in the subject-specific knowledge necessary to teach AP Statistics.

This study had two main research goals. The first was to design an Advanced Placement Statistics Teaching Knowledge (APSTK) online assessment to investigate the content and pedagogical knowledge of secondary-level, in-service AP Statistics teachers. The second goal was to explore the relationships among individual teacher assessment scores and teacher characteristic variables including educational background, years of experience teaching AP Statistics, and a self-reported percentage of student success on the AP Statistics exam.

There were three primary methodological phases included in this study. Phase I consisted of item development and item-level analysis based on responses from a national sample of current AP Statistics teachers. Phase II consisted of completing a Confirmatory Factor Analysis (CFA) to evaluate the results of a measurement model and structural model using Structural Equation Modeling (SEM). Phase III consisted of a multiple

regression analysis to determine which teacher characteristic indicator variables predicted APSTK latent variable score (LVS).

Phase I resulted in a modified assessment with nine AP Statistics Content Knowledge (APSCK) and five AP Statistics Pedagogical Content Knowledge (APSPCK) multiple-choice items. Phase II produced a measurement model with acceptable fit, and proved that items designed to measure APSCK and APSPCK fit well within the model. In addition, a structural model produced good fit, and showed evidence that APSCK was a more reliable construct than APSPCK. However, APSPCK was found to be a stronger predictor of overall APSTK. Phase III concluded that a linear combination of teacher characteristic variables was a significant predictor of APSTK LVS. Specifically, the self-reported "Student Success on the AP Statistics Exam" variable was the only statistically significant variable in predicting APSTK LVS.

Table of Contents

Dedication iv
Abstract of Dissertation v
List of Figures x
List of Tables xi

CHAPTER ONE: Introduction

Overview of the Study 1
Statement of the Problem 2
Background on Secondary Statistics Teachers 4
Background on AP Statistics Teachers 5
Importance of Statistics as a Distinct Discipline 7
Purpose of the Study 8
Research Questions 9
Statement of Potential Significance 10
Definition of Key Terms 10
Conceptual Framework 11
Summary of the Methodology 12
Limitations 13

CHAPTER TWO: Review of Related Literature

Introduction 15
Background on AP Statistics Content 19
Background on Knowledge for Teaching 22
Knowledge for Teaching in General 22
Knowledge for Teaching Mathematics 25
Pedagogical Content Knowledge for Teaching Mathematics 27
Conceptual Framework 32
AP Statistics Content Knowledge 34
AP Statistics Pedagogical Content Knowledge 37
Empirical Evidence for the Conceptual Framework 46
Conclusions 55
Why not Pedagogic Knowledge? 55
Final Thoughts on Conceptual Framework 56
Limitations of the Literature 56
Discussion of Limitations 57
Final Thoughts on Limitations 58
Observations and Notes 59
Need for APSTK Assessment 62

CHAPTER THREE: Research Design and Methodology

Overview 63
Re-Statement of the Problem 64
APSTK Assessment Development 65
Purpose of the Assessment 66
Identify Behaviors 67
Constructing an Initial Pool of Assessment Items 67
Item Development Considerations 69
Item Review 71
Participant Selection 73
Data Collection and Analyses 74
Overview 74
Phase I: Item Development and Analysis 75
Phase II: Measurement and Structural Model Analyses 81
Phase III: Exploring Teacher Characteristic Variables 88
Institutional Review Board Considerations 88

CHAPTER FOUR: Results

Phase I: Item Development and Analysis 92
APSCK Item Analysis 93
APSPCK Item Analysis 95
Phase II: Measurement and Structural Model Analyses 97
Measurement Model Results 98
Validity and Reliability of Latent Variables 103
Structural Model Results 105
Calculating Latent Variable Scores 107
Phase III: Exploring Teacher Characteristic Variables 108
Results of Multiple Regression Analysis 109
Brief Summary of Results 114

CHAPTER FIVE: Implications and Recommendations

Re-Statement of the Problem 115
Overview of Findings 115
Phase I Results 116
Phase II Results 116
Phase III Results 121
Implications of Findings 121
Impact on Statistics Education 122
Impact on Statistics Teacher Preparation 123
Impact on Mathematics Licensure Requirements 125
Impact on Student Learning of Statistics 125
Recommendation for The College Board 126
Recommendations of Future Research 128

Phase I Future Research 128
Phase II Future Research 129
Phase III Future Research 130

References 132
Appendix A: APSTK Assessment version 1 141
Appendix B: APSTK Assessment version 2 150
Appendix C: Structural Model Path Diagram 157
Appendix D: Category Descriptions of Final Items 158
Appendix E: Frequency Distributions of Score Values 159
Appendix F: Assessment Score vs. Item-Level Score Data 163

List of Figures

Figures

1. Simple Conceptual Framework 12

2. Conceptual Framework 34

3. APSTK Measurement Model 83

4. APSTK Structural Model 87

5. APSTK Assessment Score Compared to Student Success on AP Exam 127

List of Tables

Tables

1. Search Terms and Databases 17

2. Database Search Results 18

3. AP Statistics Topic Alignment with National Guidelines 21

4. Measurable Characteristics of APSCK 36

5. GAISE Learning Objectives 43

6. GAISE Pedagogical Recommendations 44

7. Measurable Characteristics of APSPCK 45

8. Comparison of Foundational Ideas and Current Constructs 45

9. APSTK Assessment Construction Process 66

10. Specification Table for APSCK Items 68

11. Specification Table for APSPCK Items 69

12. General APSPCK Item Scoring Rubric 80

13. Acceptable Levels of Fit Indices 85

14. Difficulty and Discrimination Coefficients 94

15. Categories of Phase II APSCK Items 95

16. APSPCK Cohen's Kappa Coefficients 95

17. Categories of Phase II APSPCK Items 97

18. Measurement Model Descriptive Statistics 99

19. Measurement Model Fit Statistics 100

20. Measurement Model Parameter Estimates 101

21. Structural Model Fit Indices 105

22. Structural Model Parameter Estimates 106

23. Descriptive Statistics for Latent Variable Scores 108

24. Descriptive Statistics for Teacher Characteristic Variables 109

25. Correlation Matrix of Characteristic Variables 110

26. ANOVA Table 111

27. Model Summary 111

28. Regression Parameter Estimates 112

29. Regression Coefficients 113

30. Profile Characteristics of Respondent with Average Score on Assessment 126

Chapter 1

Introduction

Overview of the Study

Since the 1990's, statistics and probability concepts have increasingly become an integral part of mathematics standards across the K-12 curriculum (Common Core, 2010; NCTM, 2000). In addition to the increased interest in statistics and probability, the College Board's Advanced Placement (AP) Statistics course has also seen increasing student enrollment. Today, student enrollment in AP Statistics has jumped to nearly 17 times the number it was when it began in 1998 (Franklin, 2011). Consequently, the need for teachers who are qualified to teach the subject has also grown.

Recent research has shown that we do not know much about teachers who teach statistics (Mills, 2007); however, we do know that most mathematics teachers are not comfortable teaching statistics (Franklin, 2010). Furthermore, there often exists a lack of adequate preparation among secondary statistics teachers (including AP Statistics teachers), and minimal opportunities for in-service teachers to gain content and pedagogic knowledge. In fact, a majority of recent research on statistics teaching and learning claims that educational researchers need to look further into the pedagogic knowledge necessary for teaching secondary-level statistics (Casey, 2010; Groth & Bergner, 2005; Franklin, 2011; Hassad, 2006; Leavy 2010; Liu & Thompson, 2007; Phannkuch, 2006; Rossman, 2006).

In the United States, most secondary school teachers of statistics have backgrounds in mathematics but little or no training in statistics, and there is no specific teaching certification in statistics. This is problematic because although statistics is a

mathematical science, it differs from mathematics in several fundamental ways. Particularly, statistics has many teaching and learning characteristics that set it apart from Mathematics in general (Cobb & Moore, 1997). This, coupled with the increasing popularity of the AP Statistics course, reinforces the idea that there is a growing need for qualified teachers of statistics (Gould & Peck, 2004).

Finally, research has shown that statistics is not only a difficult subject for students to understand, but it can also be quite challenging to teach (Mills, 2007). In fact, the AP Statistics course requires teachers to not only cover basic, introductory statistical concepts, but also integrate statistical reasoning and literacy into the curriculum; as well as incorporate ever-changing technological advances in data analysis software. This presents many opportunities and challenges for those teachers willing to take on the task of acquiring and maintaining the content and pedagogic knowledge necessary to teach the subject.

Statement of the Problem

The problem of identifying a range or level of teaching knowledge for AP Statistics is based on the idea that a teacher may possess enough teaching knowledge to teach Mathematics but be deficient in the subject-specific teaching knowledge necessary to teach AP Statistics. In fact, recent research has uncovered the need to know what knowledge and pedagogical strategies are sufficient for teaching statistics to secondary students (Casey, 2008, 2010; Groth, 2008; Hassad, 2006; Leavy 2010; Liu & Thompson, 2007; Peters, 2009; Phannkuch, 2006). Even though a need exists, there is no established benchmark for the amount of teaching knowledge that is required, or even desirable, of AP Statistics teachers. More specifically, there does not exist a criterion of

reference to determine if an AP Statistics teacher does or does not possess the content-specific knowledge necessary to teach the subject.

In addition, recent attempts to measure statistics teaching knowledge have resulted in a lack of variety of psychometric techniques leading to research results that lack generalizability and produce inconclusive results (Casey, 2010; Groth, 2008; Hassad, 2006; Izsak, 2010; Krauss, et al. 2008; Leavy, 2006, 2010; Liu & Thompson, 2007; 2010; Pfannkuch, 2006; Sharma, 2007; Schmidt, et al., 2009; Tchoshanov, 2011; Watson, 2001; Watson et al., 2008, 2011) as well as lack of item level development and construct validation (Cummings, 2010; Watson, Callingham & Donne, 2008). Therefore, the field is lacking in measures that seek to define and categorize a range of teaching knowledge for secondary-level statistics.

Finally, statistics is most often situated in the Mathematics department at the secondary level. Therefore, it is possible to have knowledge of secondary mathematics, but not possess adequate knowledge, or the right types of knowledge, to teach statistics as a distinct subject area. In fact, this idea is not new. Dating back more than twenty years, those involved in statistics and statistics education were discussing the idea of statistics being a distinct, separate discipline from Mathematics and the concern of teacher knowledge and preparation in the subject.

In 1988, a primary topic at the International Statistical Institute (ISI) Round Table Conference was the lack of adequate preparation, and other deficiencies of teachers who teach statistics. Specifically, the conference discussed how, at the high school level there were not enough teachers with formal training in statistics and even fewer who had training in the pedagogy of teaching statistics (Hawkins, 1990). More recently, the

American Statistical Association (ASA) sponsored a conference in 2003 entitled the Statistics Teacher Education: Assessment, Methods and Strategies (TEAMS) Conference. Though fourteen years after the ISI conference, the ASA conference resulted in similar findings. Again, it was concluded that at the secondary level there is a need for more teachers to be knowledgeable and prepared to teach statistics. In addition, this conference linked the growing need for more knowledgeable statistics teachers with the increasing number of students taking AP Statistics (Franklin & Mewborn, 2006).

Background of Secondary Statistics Teachers

Since the mid-1990s, research has shown that most mathematics teachers are not comfortable teaching statistics. Today, much of the research that continues to problematize the potential lack of preparation for secondary level statistics teachers focuses on mathematics teacher preparation programs at the college or university level. In general, it has been found that statistics teachers most often gain their content knowledge from a single introductory college-level statistics course, but many teachers have not had the opportunity to take such a course (Franklin 2010; Franklin and Mewborn, 2006; Mills, 2007). As a consequence, those mathematics teachers who have never taken a formal statistics course do not feel comfortable teaching statistics concepts because they have not been adequately trained (Franklin, 2010). In fact, most mathematics teachers believe that their backgrounds are inadequate to properly teach statistics (Franklin, 2010).

As far as instructional preparation, many teachers do not have ample opportunities to develop their own statistical skills and understanding of statistical concepts before teaching them. In general, college-level introductory statistics courses are not geared

specifically for those intending to be teachers, and are not well suited for a teacher preparation program (Franklin, 2010; Franklin and Mewborn, 2006).

Furthermore, few programs in mathematics teacher preparation offer much in terms of statistics instruction, and what is provided often concentrates only on the mathematical aspects of statistics such as proficiency with calculations (Rossman, 2006). Because those preparing to teach secondary school mathematics may or may not have the opportunity to learn statistics and probability content, they are also not learning how the content links to grades 7-12 mathematics instruction. (Franklin & Mewborn, 2006). Even if one assumes that most graduate students preparing to teach introductory statistics have knowledge of the subject matter, research shows that they are oftentimes lacking the knowledge of curriculum, learners, and knowledge of teaching the subject (Garfield, 2009).

Finally, research has shown that most university faculty who provide statistical education for future teachers are either statistics or mathematics professors who have little or no formal experience with grades K-12 education, or they are mathematics educators whose preparation in statistics and probability is likely quite dated and not extensive (Franklin & Mewborn, 2006).

Background of AP Statistics Teachers

As with statistics teachers in general, research has shown that A.P. Statistics teachers are also in need of more subject-specific knowledge. As the number of schools offering AP Statistics and the number students taking the course continue to grow, the number of capable teachers has not kept pace (Franklin et al., 2011). There is a need for more efforts to enhance the knowledge of future and current AP Statistics teachers.

Recent feedback from College Board consultants and AP Statistics teachers calls for more training of AP Statistics teachers especially in the areas of statistics content and active learning (Franklin et al., 2011).

Even though the AP Statistics course is based on a one-semester college-level introductory statistics course, it is often taught over the span of an academic high school year. This extended time requires teachers to integrate active learning into their pedagogical practices and develop students' statistical reasoning skills (Franklin et al., 2011). In fact, the AP Statistics course content contains more content than the typical future teacher will see in an introductory college course (Franklin et al., 2011). In fact, several AP Statistics topics are typically covered in a second semester college statistics course. For example, topics such as sampling and experimental design are not covered in most undergraduate teacher-preparation programs.

For the AP Statistics course, the goal is for students to obtain statistical literacy and a working knowledge of statistics than can be applied to a particular context. The Advanced Placement Statistics course is not the traditional formula-oriented statistics often found in introductory statistics classes, and is intended to be taught at a conceptual level (Roberts, et al., 1999). This goal does not allow teachers simply to teach comfortable algorithms formulas that are easy to memorize. Today's AP Statistics students are encouraged to discover key ideas and concepts with hands-on activities or by using simulation and to be able to communicate their ideas. Furthermore, this discovery is often aided through computers or calculators—tools with which many in-service and prospective statistics teachers feel inadequate (Franklin, 2010).

Importance of Statistics as a Distinct Discipline

Many in the Mathematics and statistics education community argue that while statistics is situated in the Mathematics K-12 curriculum, statistics is a separate discipline distinct from the field of Mathematics in general (Bessant & MacPherson, 2002; Cobb & Moore, 1997; Rossman, 2006). Specifically, Rossman (2006) states that although statistics certainly makes use of much Mathematics, it is a separate discipline and not a branch of mathematics. Also, it is true that statistics cannot function without mathematics, but the opposite argument does not hold up (Cobb & Moore, 1997). According to this argument, the primary difference between statistics and Mathematics is that in statistics context is crucial, because data are not just numbers they are numbers with a context. Conversely, Mathematics is an abstract field of study and can exist independently of context (Cobb and Moore, 1997).

Of particular importance to this study, research has found that teaching statistics requires a unique type of pedagogic knowledge apart from Mathematics (Bessant & MacPherson, 2002, Rossman, 2006). Statistics requires a different kind of thinking from teachers and students (Cobb and Moore, 1007) resulting in different definitions of student learning. Rossman (2006) explains how a distinct type of instructional preparation is needed for statistics teachers because students react differently to statistics than to Mathematics. He further states that this problem is exacerbated by the fact that few teacher preparation programs offer instruction in statistics, or only focus on the mathematical aspects of statistics.

In addition, unique to statistics students is the imperative need to relate their calculations to the context of the problem. Many students who are comfortable simply

making calculations do not initially expect such a strong focus on the need for interpretation and explanation regarding their conclusions. Statistics students are often very uncomfortable with uncertainty and the lack of definitive conclusions, so statistics teachers need to be prepared for this discomfort (Rossman, 2006).

Finally, there are those who do not want to see statistics as a separate discipline in a K-12 setting. For example, Scheaffer (2006) and Franklin and Mewborn (2006) advocate that statistics should not be entirely separated from Mathematics in the K-12 curriculum, because it might be spread out across the curriculum of other social sciences, and therefore become weakened and the disciplinary integrity compromised. Though Scheaffer (2006) and Franklin and Mewborn (2006) argue that statistics should not be treated as a separate discipline in the K-12 setting, they do agree with other researchers that schools of education should treat the content knowledge and pedagogy of statistics as a separate field distinct from mathematics.

Purpose of the Study

To date, there does not exist an established measure of AP Statistics teaching knowledge. Therefore, in order to further investigate the content and pedagogical knowledge of AP Statistics teachers the present study seeks to develop a normative instrument designed to capture a range and level of variability of AP Statistics teaching knowledge. To this end, a new construct entitled AP Statistics Teaching Knowledge (APSTK) is created for this study. APSTK is grounded in research relating to content and pedagogical content knowledge, and is comprised of two smaller constructs. First, AP Statistics content knowledge (APSCK) as presented by Shulman (1986) Schoenfeld (1985), Ball (1988) and others and also AP Statistics pedagogical content knowledge

(APSPCK) as it relates to subject specific student learning (Ball, 2008; Ball et al., 2008; Schoenfeld and Kilpatrick, 2008; Shulman, 1986).

The first purpose of this study was to create an assessment designed to identify individual AP Statistics teacher assessment Latent Variable Scores (LVS). The Advanced Placement Statistics Teaching Knowledge (APSTK) assessment worked to capture a range, average and level of variability in teaching knowledge among AP Statistics teacher respondents. The second purpose for developing the assessment was to explore the relationships among teacher scores on the APSTK LVS and teacher characteristic variables relating to educational background, years of experience teaching AP Statistics, and a self-reported rating of student success on the AP Statistics exam.

Research Questions

The research questions were as follows:

1. What items are suitable indicators APSCK and APSPCK for an assessment of APSTK?
2. Does a two-factor measurement model produce acceptable model fit?
3. What amount of shared covariance exists between APSPCK and APSCK in a two-factor measurement model?
4. Does a higher-order structural model produce acceptable model fit?
5. Does APSCK contribute to the overall measurement of APSTK?
    5a. What effect does APSCK have on APSTK?
6. Does APSPCK contribute to the overall measurement of APSTK?
    6a. What effect does APSPCK have on APSTK?
7. What is the latent variable APSTK score for each respondent?

8. What is an average level of APSTK among A.P. Statistics teachers?

9. Which independent variables are significantly related to APSTK score while controlling for other relevant variables?

Statement of Potential Significance

Development of an APSTK assessment will contribute to the field of statistics education in several key ways. First, research on secondary-level, in-service statistics teaching knowledge is incomplete, and a benchmark for teaching knowledge for A.P. Statistics does not exist. So, creating an assessment that works to identify a range or average amount of teaching knowledge among current AP Statistics teachers will help to fill this gap. Second, research shows the need for more confirmatory analyses that work to validate statistics teaching knowledge constructs including those that relate to student learning. Therefore, the APSTK assessment and resulting statistical analyses seek to validate the knowledge constructs of APSCK and APSPCK and establish an average level of teaching knowledge for AP Statistics.

Definition of Key Terms

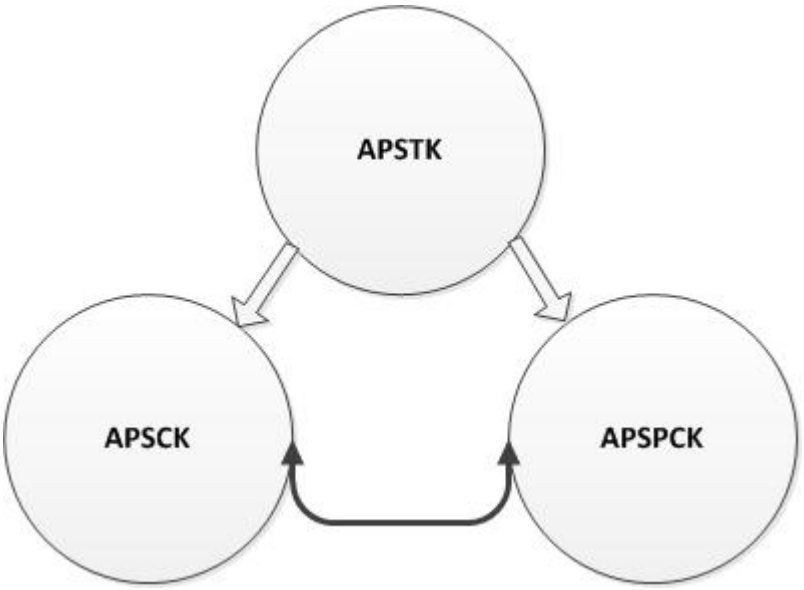
The APSTK construct is comprised of two smaller constructs, AP Statistics Content Knowledge (APSCK) and AP Statistics Pedagogical Content Knowledge (APSPCK). First, APSCK is defined as problem-solving content knowledge (Schoenfeld, 1985) combined with a specialized content knowledge specific to teaching (Ball, 1988, 2008) AP Statistics. This includes the ability to identify and recognize a variety of ways AP Statistics course content is organized, and the ability to unpack AP Statistics content as necessary to increase student understanding.

The second construct that constitutes APSTK is Pedagogical Content Knowledge for teaching AP Statistics or APSPCK. APSPCK is defined as a teacher's knowledge of how her students learn or do not learn specific topics (Shulman, 1986) within the AP Statistics curriculum. This includes being able to anticipate and interpret student thinking as it is happening (Ball, et al., 2008). Furthermore, a teacher's APSPCK allows her to recognize common errors and misconceptions (Shulman, 1986; Ball et al., 2008; Schoenfeld and Kilpatrick, 2008), and adjust instruction, curriculum choices and assessment strategies accordingly (Ball, 2008)

Conceptual Framework

The current study seeks to define and measure APSTK by quantifying APSCK and APSPCK. Each of these two constructs is grounded in the work of Schoenfeld (1985), Shulman (1986) and Ball (1988, 2003, 2005), and the more recent work of them and others. Because the work of Shulman (1986), Ball (2003, 2005) and others has been so influential in establishing frameworks for examining teaching knowledge in various contexts, this study considered the overlap or intersection that exists when defining the current constructs. Specifically, these intersections were used to define APSCK and APSPCK by transforming original notions of teaching knowledge into teaching knowledge that is specific to AP Statistics. Figure 1 shows a simple model of the conceptual framework for APSTK highlighting a proposed correlation between APSCK and APSPCK. A more detailed conceptual framework will be presented and discussed in Chapter 2.

Figure 1: Simple Conceptual Framework



Summary of the Methodology

The methodological steps for this study are centered around three main phases. The first phase included item development and analysis of APSCK and APSPCK items. Initially, an assessment was developed consisting of approximately 12 multiple-choice APSCK questions to include an equal number of items for each sub-component of APSCK (exploring data, sampling and experimentation, anticipating patterns, statistical inference) as well as 9 constructed response APSPCK questions. An initial group of current AP Statistics teacher respondents were invited to answer these items. This data helped to establish measured subscales of APSCK and item level scores of APSPCK.

Results from this initial assessment were used to conduct item level analysis, establish individual sub-scale scores for APSCK, and change the APSPCK items from constructed response to multiple-choice items. Next, a Confirmatory Factor Analysis

(CFA) was run using data collected from a second sampling of current AP Statistics teachers to confirm a measurement model. Then, an APSTK structural model was evaluated and confirmed. Finally, an APSTK LVS was calculated for each participant and compared to teacher characteristic descriptor variables using a multiple regression analysis.

Limitations

There are several limitations in the design of the study that may ultimately impact psychometric analyses and results. From an assessment design standpoint, there was one version of the APSTK assessment as well as one item developer and item rater. This may negatively affect reliability and validity measures. In addition, teacher participation in answering the assessment questions relied solely on intrinsic motivation. There are no incentives or rewards given to teachers who choose to respond except a chance to win one of four online gift cards. Also, because the assessment is entirely online, response rates may be lower than if it was administered in person (Cook, 2000). Finally, teacher characteristic data such as "Student Success on AP Statistics exam", and "Number of Years Teaching the Course" are self-reported and therefore potentially inaccurate or bias.

From a methodological standpoint, this study evaluated a two-factor structure (APSCK and APSPCK), which may not account for other important constructs that influence APSTK such as general pedagogic knowledge. In addition, there was a possibility that items and/or factors may not represent either construct or contribute to the CFA model.

From a conceptual standpoint, the framework for APSTK relied more heavily on mathematics teaching knowledge frameworks than statistical teaching knowledge

frameworks. In fact, the current framework was constructed based on Mathematics education literature because knowledge for teaching AP Statistics had not yet been studied. Therefore, this study assumed the more general model of content knowledge and pedagogical content knowledge for teaching Mathematics was valid across disciplines within the field of Mathematics; including statistics.

Chapter 2

Review of Related Literature

Introduction

Research regarding mathematics teaching knowledge has been active for the past thirty years (Ball, et al., 2001) and continues to grow and change. In particular, research in the area of statistics teaching knowledge has seen a dramatic increase over the past 20 years (Shaughnessy, 2005). Moreover, several key frameworks for teaching knowledge have been identified and used in empirical research (Shulman, 1986; Grossman, 1990; Fennema and Frank, 1992; Ball & Bass, 2000, 2003; Ball & Hill, 2005; Ball et al., 2008; Ma, 1999; Usiskin, 2000; Schoenfeld, 1985; Gal, 2004; Burgess, 2007; Wild and Pfannkuch, 1999).

In general, research on statistics teaching knowledge has been conducted on pre-service elementary and college-level teachers. Therefore, helping secondary-level, practicing teachers develop a better knowledge of statistics is still an area that needs to be further explored (Garfield and Ben-Zvi, 2007). Furthermore, studies that have examined statistics teaching knowledge tend to focus on the K-5, undergraduate, and graduate levels of education (Groth, 2003). Therefore, more research is needed at the secondary level (Garfield & Chance, 2000).

The purpose of this review of literature was to synthesize research pertaining to measuring in-service, secondary level statistics education teaching knowledge in order to substantiate a conceptual framework of AP Statistics Teaching Knowledge or APSTK. In addition, statistics learning frameworks were examined to the extent that they

influence secondary-level statistics teaching content knowledge and pedagogical content knowledge.

This literature review sought to answer the following questions:

- What conceptual frameworks have been applied to the study of teaching knowledge?
    - What frameworks of teacher knowledge have been used to study Advanced Placement or secondary level statistics?
    - How might these frameworks be adapted to examine the current construct of AP Statistics Teaching Knowledge (APSTK)?
- What frameworks have been used to study student learning and understanding in statistics?
- What research methodologies have been used to measure teaching knowledge for secondary level statistics?
    - What are the limitations of these methodologies?
    - How do these methodologies inform current methodological decisions?

In an effort to uncover recent research applicable to the focus of APSTK, database search efforts for this literature review were narrowed down in several ways. First, no date restriction was placed on key literature pieces that were seminal to the work of establishing or classifying teacher knowledge ideas and concepts. Second, studies concerning both primary and/or secondary statistics teacher knowledge were quite abundant, so those specific to PK-6 grade statistics teacher knowledge were not included in the review unless the methodology was of interest. Third, particular attention was paid

to more recent ideas (1990 or later) of teaching knowledge, because the A.P. Statistics exam was first offered to students in 1997. Finally, in order to examine methodologies focused on measuring secondary statistics teacher knowledge, the search was opened up to include any research that attempted to measure mathematics teacher knowledge quantitatively during the same time frame. The databases searched and search terms used are outlined in Table 1:

Table 1: Search Terms and Databases

Categories	Search Terms	Databases (sources)
Frameworks for Teaching Knowledge	Teacher Knowledge, Content Knowledge, Mathematics Teaching Knowledge	Academic Search Premier*, Academic Search Complete* Jstor*, ERIC*, Education Abstracts*, Dissertations and Theses Online*, Teacher Reference Center,
Statistics Teaching Knowledge	Statistics education, secondary mathematics, content knowledge, Pedagogical Content Knowledge	SERJ, NCTM, Conference Presentations, Causeweb.org
Measuring Secondary Math Teaching Knowledge	Teacher Knowledge, Secondary Mathematics, measurement, quantitative methods, Factor analysis, Item Response Theory, Confirmatory Factor Analysis, Exploratory Factor Analysis	SERJ, PsycInfo, PsycArticles, CauseWeb.org, Teacher Reference Center
Student Learning of Statistics	Statistical literacy, secondary-level, statistical reasoning, statistics learning frameworks, student learning,	SERJ, CauseWeb.org

*Database used in all searches

Table 2 outlines the database search results. The decision to include or not include articles was based on certain criteria. As previously mentioned, studies pertaining to elementary or college-level teacher knowledge were not included as well as studies identifying teacher attitudes or beliefs. Finally, studies conducted with a qualitative methodology were not included if they did not attempt to measure or quantify teacher knowledge. For example, search results for Student Learning of Statistics returned 56 results, but only 9 applied to secondary student learning of statistics over the past 20 years. Of the methodological studies included, the type of methodology used is also noted in Table 2. Limitations of the included studies are discussed at the end of this chapter.

Table 2: Database Search Results

Type of Study	Number of results	Number included
Frameworks for Teaching, Teaching Mathematics	57	8
Statistics Teaching Knowledge	29	9
Student Learning of Statistics	56	7
Measuring Secondary Math Teaching Knowledge	71	13

Type of Methodology	IRT/Rasch Model	Descriptive Stats/Correlation	Factor Analysis	SEM	Case Study
Number of Studies*	3	7	7	2	4

*Some studies used more than one method, so "number of studies" equals more than 13

Background on AP Statistics Content

First administered in 1998, the AP Statistics course has a strong emphasis on collecting data, including sampling and experimentation and de-emphasizes probability (outside of the context of inference) and computation. In addition, the course focuses on clear and accurate interpretation of statistical analyses, and communication is weighted heavily in the scoring of the AP Statistics exam. The exam also includes an investigative task that asks students to integrate statistical ideas and methods from the course to solve a problem of a type that they may not have previously encountered. Because of the clear emphasis on conceptual understanding and problem solving, teaching this course is a challenge for most secondary teachers. In fact, many, if not most, elementary teachers and even secondary mathematics teachers are not well prepared to teach statistical topics (Gould & Peck, 2004).

The AP Statistics course scope and sequence was developed in conjunction with the National Council of Teachers of Mathematics' (NCTM) Principles and Standards for School Mathematics (2000) and align with more recent recommendations of the American Statistical Association (ASA) (2007) and the Common Core Standards for Statistics and Probability (2010) as shown below in Table 3. Specifically, the AP Statistics course is centered around four broad conceptual themes: exploring data, planning a study, anticipating patterns, and statistical inference. AP Statistics teachers are free to teach course topics in any sequence, though most follow the College Board recommended timeline of first covering exploring data with graphs and numerical summaries, then association, sampling, probability distributions, inference, group comparisons, association between categorical and continuous variables, and finally an

introduction to multiple regression. These topics are most often taught over a two semesters during the high school academic year.

Table 3: AP Statistics Topic Alignment with National Guidelines

College Board (2010) Categories of AP Statistics Curriculum		**NCTM (2000) Data Analysis and Probability Standards**	**ASA (2007) Guidelines for Assessment and Instruction in Statistics Education (GAISE)**	**Common Core (2010) Statistics and Probability Standards**
Exploring Data: Describing patterns and departures from patterns (20-30%)		Select and use appropriate statistical methods to analyze data	Analyze Data	Interpreting Categorical and Quantitative Data
Sampling and Experimentation: Planning and Conducting a Study (10-15%)		Formulate questions that can be addressed with data and collect, organize and display relevant data to answer them	Formulate the question Collect the data	Making inferences and justifying conclusions
Anticipating Patterns: Exploring random phenomena using probability and simulation (20-30%)		Understand and apply basic concepts of probability	Formulate the Question Collect the data	Conditional Probability and the Rules of Probability Using Probability to Make Decisions
Statistical Inference: Estimating population parameters and testing hypotheses (30-40%)		Develop and evaluate inferences and predictions that are based on data	Interpret Results	Making inferences and justifying conclusions

The AP Statistics exam was developed by the College Board in 1998, and emphasizes four main areas: exploring data, sampling and experimentation, anticipating patterns and statistical inference. These objectives align with the NCTM Data Analysis and Probability Standards (2000) as well as the ASA's GAISE guidelines (2007) and the Common Core Statistics and Probability Standards (2010). All of these recommendations require students to analyze and describe data, formulate questions, plan experiments and apply inference techniques to justify conclusions. Rules of probability and simulation are included in all areas; however, the AP exam only focuses on probability as it pertains to inference.

Background on Knowledge for Teaching

Knowledge for Teaching in General

In order to consider the underpinnings of a conceptual framework surrounding the construct of APSTK, it was necessary to begin with a broad lens and examine frameworks put in place to explore teacher knowledge in general (Grossman, 1990; Lloyd and Wilson, 1998; Shulman, 1986). It seemed that there existed an uneven research base across domains of teaching knowledge, leading to a variety of working definitions for terms such as "content knowledge" and "pedagogical content knowledge" (Shulman, 1986). Such terms vary from study to study; therefore, it was necessary to mark-off boundaries for constructs leading to APSTK by reviewing original definitions as well as more contemporary notions of what it means to possess teaching knowledge.

Shulman (1986) was the first to separate the idea of pedagogic knowledge from the content knowledge of one's subject area. His seminal framework described three components of teachers' overall content knowledge: "subject matter knowledge",

"pedagogical content knowledge", and "curricular knowledge". In general, his efforts focused on examining how teachers' overall content knowledge included deciding what content to teach, how to represent what they teach, how to question students about it, and how to deal with misunderstandings.

Shulman (1986) used a hierarchical system to categorize teacher knowledge by examining how successful college graduates transformed knowledge of a subject into a form used for teaching. In the broadest of terms, Shulman (1986) believed that "content knowledge" encompassed the total amount and organization of knowledge in the mind of the teacher.

He also argued that teachers must know a variety of ways to organize the content of a discipline as well as how competing representations may or may not be warranted within a given subject area. Therefore, the idea of "subject matter knowledge dictates" that, "a teacher need not only understand that something is so; the teacher must further understand why it is so (p. 9). In this way, it is expected that the subject matter knowledge of a teacher must be greater than or equal to that of a colleague who simply has a degree in the subject area. In addition, Shulman (1986) states a teacher's subject matter knowledge allows her to make decisions about which topics are central to the subject and which are peripheral. In addition, a teacher's ability to judge central verses peripheral topics within a subject area also has many pedagogical consequences.

Shulman's (1986) idea of "pedagogical content knowledge" (p. 9) has also been significant to the work of understanding teacher knowledge (Casey, 2008). For Shulman (1986) the pedagogical relationship to subject matter knowledge lies within a teacher's pedagogical content knowledge or ability to understand representations, analogies,

illustrations and explanations of topics specific to her discipline. Shulman (1986) states that pedagogical content knowledge depicts the type of knowledge teachers must have to identify errors and misconceptions, change explanations, and formulate the subject to accommodate shifting student interpretation. Important to the current study is Shulman's (1986) distinction that pedagogical content knowledge is specific to one's subject area.

Finally, Shulman (1986) determined that the third and final component of his framework included ideas relating to general teaching strategies such as determining instructional materials, determining which materials to use at varying instructional levels, and how these materials aid in presenting content and evaluating student work. For him, this form of teacher knowledge is entitled "curricular knowledge" (p. 10). In one way, curricular knowledge serves as a means to enrich the subject matter by providing cross-curricular examples and making connections to other subject areas. However, curricular knowledge also serves to vertically integrate curriculum within a single subject. Specifically, this knowledge type allows the teacher to determine what topics precede or follow the current topic within the same content area.

Shulman's (1986) framework for teacher knowledge laid the groundwork and paved the way for future teacher knowledge research. For example, Peterson (1988) kept Shulman's three main components of teacher knowledge and added a fourth: teachers' "meta-cognitional knowledge". This framework argued that it is not only important for teachers to know how students think and how to facilitate growth in student thinking but also possess a self-awareness of her or his own thinking in mathematics. (Casey, 2008)

Grossman (1990) also transformed Shulman's (1986) framework for teaching knowledge. She noted four categories: "subject matter knowledge", "pedagogical content

knowledge", "general pedagogic knowledge" and "context knowledge". In addition, Lloyd and Wilson (1998) also extend Shulman's types of knowledge by defining content "conceptions" that, "encompass the range of conceptions that a teacher might hold about a particular topic or concept that is taught" (p. 250). The word conceptions for Lloyd and Wilson (1998) refer not only to knowledge but also to beliefs, understandings, preferences, and views of a particular topic.

As the conceptualization of teacher knowledge becomes more complex, the challenge for researchers becomes blending all the components or domains to describe teaching knowledge that is subject specific. Using frameworks established by Shulman (1986), Grossman (1990) and others, mathematics education researchers attempt to detail and categorize the components of teaching knowledge for mathematics.

Knowledge for Teaching Mathematics

Frameworks for teaching knowledge in general have led to the development of more content specific frameworks that describe mathematical teaching knowledge (Ball & Bass, 2000, 2003; Ball & Hill, 2005; Fennema and Frank, 1992; Lappan, 2000; Ma, 1999; Schoenfeld, 1985; Usiskin, 2000;). In order to consider teaching knowledge frameworks specific to the field of mathematics teaching and learning, it is necessary to discuss the work of Schoenfeld (1985), Ball and colleagues (1988, 2005, 2008), and others.

To begin, Schoenfeld (1985) was one of first to consider the components of content knowledge (Shulman, 1986) for teaching mathematics. His ideas centered around two main components of content knowledge: "declarative knowledge" and "problem-solving knowledge". Schoenfeld (1985) defined declarative knowledge as

simply that which can be learned through memorization consisting of algorithmic and routine procedures, facts, and rules. Specifically, he thought of declarative knowledge as information teachers must retrieve from memory without hesitation, or information they know at a glance. He describes how this type of knowledge is the foundation for problem solving. In addition, Schoenfeld (1985) cautions that incorrect or flawed declarative knowledge can interfere with mathematical understanding and performance. In this way, a strong foundation in declarative knowledge is essential to the development of higher-order mathematical knowledge pertaining to problem solving and mathematical reasoning.

This higher-order mathematics knowledge is essential to the development of problem solving and mathematical reasoning (Schoenfeld, 1985). The second type of content knowledge Schoenfeld (1985) describes is entitled problem-solving knowledge and requires the mathematics teacher to develop, apply and reflect upon mathematical strategies. Schoenfeld (1985) describes how knowledge of mathematical problem solving is essential to further development of "higher-order thinking skills" (p. 26). Problem solving knowledge is important because the mathematics teacher must possess the expertise and flexibility to use problem solving knowledge efficiently when called upon. Here, the teacher must know how to select and implement the appropriate content knowledge as needed to help students realize the consequences for their solution choices. Therefore, mathematical problem solving and reasoning requires using existing knowledge to reason, interpret, and make connections outside of the original context in which it was learned.

In 1988, Deborah Ball began the work of unpacking her own definitions and categories of subject matter knowledge for teaching mathematics by hypothesizing that there exists a "pedagogical equation" (p. 53) made up of several key categories of teacher knowledge specific to mathematics, much like Shulman (1986) categorized for teaching in general. For Ball (1988), the teacher knowledge equation included "subject matter knowledge", "knowledge of teaching and learning", and "knowledge of learners and context".

Ball (1988) originally described subject matter knowledge as a teacher's understandings of and about mathematics. In the simplest of terms, subject matter knowledge is the idea that a teacher cannot teach what she herself does not understand. As a result, a teacher who does not have a high level of subject matter knowledge may provide inaccurate representations that are misleading to students.

Pedagogical Content Knowledge for Teaching Mathematics

Ball's (1988) original framework uncovered a separate, but heavily related, construct entitled knowledge of teaching and learning. This type of pedagogic knowledge was grounded in a teacher's ability to decide which instructional tasks are most appropriate for an individual student or specific class. For Ball (1988), this second type of teacher knowledge was heavily influenced by the individual teacher's assumptions, thoughts and beliefs about how well her students are learning, or how well her students' needs are being met. In essence, a teacher's knowledge of how to make on the spot pedagogical decisions in addition to content-driven adjustments.

The final type of pedagogic knowledge introduced in Ball's (1988) original notions of mathematics teaching knowledge was knowledge of learners and context. As

with the other two types of teacher knowledge, knowledge of learners and context was heavily connected to knowledge of teaching and learning in such a way that one moves from considering teacher thinking to teacher action. In this way, knowledge of learners and context was characterized by a teacher's ability to develop a variety of tasks or in-class activities to encourage student learning, decide whether or not to consider student suggestions, and to facilitate class discussions focused on specific topics.

More recently, Ball and her colleagues (Ball & Bass, 2003; Ball & Hill, 2005; Ball, Thames & Phelps, 2008; Hill & Ball, 2004; Hill & Ball, 2009; Thames & Ball, 2010) have proposed newer frameworks for teaching knowledge of mathematics. For example, Ball, Thames and Phelps (2008) defined a new construct entitled, "Mathematical Knowledge for Teaching" and discussed two main parts: "subject matter knowledge" and "pedagogical content knowledge". Within the construct of subject matter knowledge there are three smaller components: "common content knowledge", "specialized content knowledge for teaching" and "horizon content knowledge". Common content knowledge is mathematical content that the average person knows while specialized content knowledge is the mathematical knowledge and skill unique to teaching or knowledge not typically needed for purposes other than teaching. Horizon content knowledge is a teacher's knowledge of how mathematical topics are related across different curricula.

Also within this framework is a larger construct entitled pedagogical content knowledge (Ball, Thames & Phelps, 2008) and includes three components: "knowledge of content and students", "knowledge of content and teaching", and "knowledge of content and curriculum". Ball et al., (2008) point out that ideas regarding pedagogical

content knowledge draw heavily from literature on mathematical student thinking and learning. For Ball et al., (2008), knowledge of content and students is characterized by a teacher's level of familiarity with common errors and deciding which of those errors students are most likely to make. Knowledge of content and teaching is a mathematical knowledge of the design of instruction and the ability to evaluate representations and consider how these decisions impact student learning. Finally, for Ball et al., (2008) knowledge of content and curriculum is a blended category of the subject matter knowledge teachers use to make curricular decisions.

Several other recent frameworks have been established to examine mathematical teaching knowledge as a unique, content-focused area. For example, Usiskin (2000) argued that there exists a construct that includes knowledge of mathematics as well as a pedagogical content knowledge component, and includes topics such as making mathematical connections and analyses across the K-12 curriculum. In addition, Ma (1999) defined a type of mathematical content knowledge entitled, "fundamental mathematics" defined as a teacher's knowledge of foundational concepts that are necessary to pursue formal learning of the subject. This knowledge type is illustrated by a teacher's ability to make connections, promote multiple perspectives and emphasize coherent basic concepts.

Van Driel and Berry (2012) describe how Shulman's (1986) pedagogical content knowledge, "includes teachers' understanding of how students learn, or fail to learn, specific subject matter" (p. 26). They argue how research on pedagogical content knowledge should not result in one ideal or optimal way to teach certain subject matter. Instead, Van Driel and Berry (2012) describe the complex nature of pedagogical content

knowledge as a form of teachers' professional knowledge that is highly topic, person, and situation specific.

In another example, Fennema and Frank (1992) developed a model for mathematics teachers' knowledge that centered on teachers' knowledge as it happens in the context of the classroom. This framework included the components of "knowledge of mathematics", "pedagogical knowledge" and "knowledge of learners' cognitions in mathematics". The last component noted that student learning was part of an interactive, dynamic relationship with teachers' context-specific knowledge.

Lappan (2000) argues that teacher knowledge types do not act in isolation and views the components (or domains) of teacher knowledge in a nonlinear fashion. He proposes that all the domains interact together to produce knowledge for teaching. The domains include: "content for teaching", "pedagogy", "knowledge of student development", "vertical vision of curriculum", and "evaluation".

In this model, the domain content for teaching includes not just the content of mathematics but also knowledge of the mathematics taught in preK-12 programs and how the ideas at different levels relate. Lappan's (2000) pedagogy domain includes "compelling examples of tasks that can be used to elicit particular kinds of thinking and investigation on the part of the learner" (p.322). Knowledge of student development as it is related to learning and teaching mathematics is included in the learning domain. It also includes knowledge of and experience in listening to what students can make sense of and where they have problems. Lappan (2000) also mentions as an important domain, a vertical vision of curriculum that includes the lesson, unit, school year and pre- K-12 level. Finally, Lappan (2000) also consider an evaluation domain, which includes

assessment of students' knowledge and evaluation of programs.

In an attempt to build a cognitive model of teacher knowledge, Leinhardt, Greeno, Putman, Stein and Baxter (cited in Fennema & Franke, 1992) propose that the skill of teaching is determined by at least two fundamental and related systems of knowledge: "subject matter" (content knowledge) and "lesson structure" (practical knowledge). For Leinhardt and her colleagues, knowledge of lesson structure consists of organizing the lesson with an overall goal, some procedures or activities to achieve the goal, and some routines that enable the class to function smoothly. Knowledge of the subject matter includes not only mathematical knowledge but knowledge of curricular activities, methods of presentation, and assessment procedures.

These frameworks describe teacher knowledge in a way that generates a conceptualization of what it means to have teaching knowledge. The components of pedagogical content knowledge have been refined over the past 20 years to emphasize a teacher's knowledge of subject-specific, student learning. The focus now is on how teachers combine content knowledge and knowledge of their students as learners in the context of a particular class. Today, the make-up of pedagogical content knowledge is complex, and encapsulates more than just recognizing student errors and misconceptions. The idea that using one's mathematical knowledge to make informed pedagogical decisions affecting student learning hints at a correlation between content knowledge and pedagogical content knowledge. Thus, a new framework of teaching knowledge specific to AP Statistics is conceived.

Conceptual Framework

The exploration of research on secondary-level statistics teacher knowledge using the terms and databases listed in Table 2 turned up many studies that are not included here because they pertain to elementary, middle school or college-level teacher knowledge, and/or focused on pre-service (as opposed to in-service) teacher knowledge (Berger, 2005; Bright, Friel & Berenson, 1993; Groth, 2008; Groth & Bergner, 2005; Lancaster, 2007; Leavy, 2006, 2010; Metz, 2010; Mickelson & Heaton, 2004; Pfannkuch, 2006; Saderholm et al., 2010; Sorto, 2004; Thomas, 2002). In fact, the majority of studies found did not attempt to measure secondary-level statistics teaching knowledge of in-service teachers. Furthermore, many of these studies were also excluded from this discussion because they were qualitative in nature, and therefore examined topics such as teacher attitudes, perceptions or understandings of statistical concepts as opposed to knowledge (Arnold, 2008; Bergner, 2005; Groth & Bergner, 2005; Lancaster, 2007; Levey, 2006; Mills, 2007; Peters, 2004; Pfannkuch, 2006).

In general, a review of research on statistics teaching found a lack of content knowledge and a lack of information regarding pedagogical content knowledge needed for teaching secondary level statistics. In fact, the majority of studies found that educational researchers need to look further into the pedagogic knowledge necessary for teaching statistics at all levels (Casey, 2010; Groth & Bergner, 2005; Hassad, 2006; Leavy 2010; Liu & Thompson, 2007; Phannkuch, 2006). Furthermore, while most studies began by seeking to quantify, perceive, or organize content knowledge, many concluded that the next research step is to consider how this lack of content knowledge is

affecting a teacher's pedagogical content knowledge (Groth, 2008; Groth & Bergner, 2005; Hassad, 2006; Leavy, 2010).

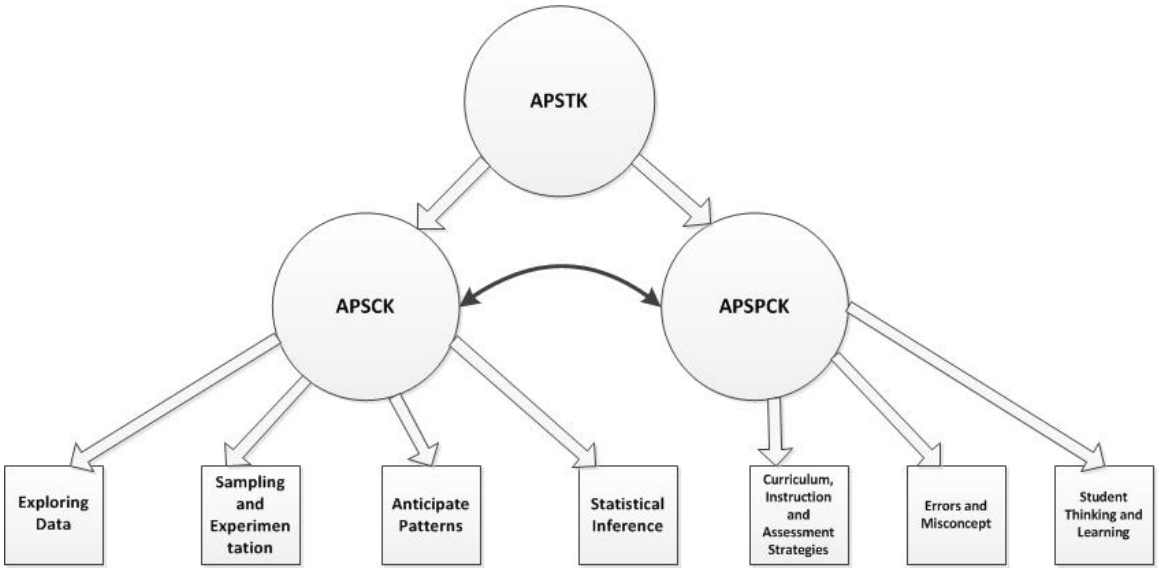
Frameworks regarding knowledge for teaching AP Statistics, or secondary statistics in general have not been established (Casey, 2008). The limited amount of research in the area of statistics teaching has focused on teachers' level of statistical literacy, thinking or content knowledge of a particular statistical concept (Burgess, 2007; Casey, 2005; Gal, 2004; Makar & Confrey, 2004; Sorto, 2004; Wild and Pfannkuch, 1999). In fact, most research on statistical content knowledge has used previously established teaching knowledge frameworks (ie. Ball & Hill, 2005; Wild & Pfannkuch, 1999; Franklin et al., 2005) to explore statistical teaching knowledge.

For example, Burgess (2007) worked to blend the teaching knowledge frameworks established by Ball and Hill (2005) and Wild and Pfannkuch (1999). This study found that all previously established knowledge types also applied to teaching statistics, and therefore took an initial step in describing a framework for teaching knowledge of statistics. In another example, Sorto (2004) used the framework of Ball and Hill (2005) to examine statistical knowledge of middle school teachers for teaching data analysis. This study validated the existence of a common content knowledge construct and specialized content knowledge construct for teaching data analysis. Finally, Casey (2008) also used the knowledge framework established by Ball and Hill (2005) and found 29 categories on content knowledge for teaching statistical association at the secondary level.

The current study seeks to define and measure APSTK by examining AP Statistics teachers' content knowledge (APSCK) and AP Statistics pedagogical content

knowledge (APSPCK). Each of these two constructs was grounded in the work of Schoenfeld (1985), Shulman (1986) and Ball (1988, 2003, 2005), as well as more recent work of others. Because the work of Shulman (1986), and Ball (2003, 2005) had been so influential in establishing frameworks for examining teaching knowledge in various contexts, it was necessary to consider the overlap or intersection that existed when defining the current constructs. For this study, these intersections were used to define APSCK and APSPCK by transforming original definitions of teaching knowledge into a definition of teaching knowledge that is specific to AP Statistics. Figure 2 shows the conceptual framework for APSTK.

Figure 2: Conceptual Framework



AP Statistics Content Knowledge

AP Statistics Content Knowledge (APSCK) was defined as problem-solving content knowledge (Schoenfeld, 1985) combined with a specialized content knowledge specific to teaching (Ball, 1988, 2008) AP Statistics. This included the ability to identify and recognize a variety of ways AP Statistics course content is organized, and the ability

to unpack AP Statistics content as necessary to increase student understanding. Specifically, the current idea of APSCK is an intersection of Shulman's (1986) concept of "subject matter content knowledge", Schoenfeld's (1985) "declarative" and "conceptual" levels of "content knowledge" as well as Ball et al.'s, (1988, 2008) construct of "subject matter knowledge".

Shulman (1986) stated that "subject matter knowledge" is a teacher's ability to understand the variety of ways that content of a discipline is organized as well as competing representations of the same concept. Schoenfeld (1985) recognized that mathematics teachers must use both "declarative knowledge" and "problem-solving knowledge" to, not only memorize formulas, but also disentangle mathematical problems along the way. Finally, Ball et al. (1988, 2008) showed that "subject matter knowledge" for teaching mathematics is more than general mathematical knowledge or ability.

Combining Shulman's (1986) idea that "subject matter content knowledge" must be greater than or equal to that of someone who simply has a degree in the subject and Ball et al.'s (2008) separation of "common content knowledge" and "specialized content knowledge" for teaching along with Schoenfeld's (1985) levels of "content knowledge" leads to the definition of APSCK. This definition was important during item development to differentiate APSCK items from original AP Statistics content knowledge items. APSCK items were aligned with AP Statistics content, but they also included additional components specific to the definition of APSCK such as specialized content knowledge specific to teaching. APSCK items were meant to identify more than simply AP Statistics students' content knowledge and were different than already existing items.

Table 4 outlines the characteristics for APSCK based on this definition.

Table 4: Measurable Characteristics of APSCK

A.P. Statistics Themes	Definitional Components of APSCK Problem Solving Knowledge	Organize Content	Student Learning
Exploring Data	Interpret information from graphical and numerical displays Analyze beyond the data to make generalizations beyond the given context Evaluate the benefits and shortfalls of choosing one technique over another	Decide when graphical and numerical techniques should be taught	Explain how graphical and numerical techniques are used to visualize data Create explanations for why departures from patterns might exist
Sampling and Experimentation	Understand sampling designs with random selection Apply experimental designs using randomization Differentiate methods of data collection and analysis Develop appropriate conjectures based on a well-developed plan Design experiments to test conjectured relationships among variables	Determine when sampling and experimentation topics should be taught	Model situations that use various types of sampling procedures

Anticipating Patterns	Measure uncertainty using probability models and theory Experiment with a variety of probability models Compare and contrast continuous and discrete probability distributions	Decide when probability topics should be taught	Model the use of probability theory in real-world scenarios
Statistical Inference	Recognize unique characteristics of each type of hypothesis test Use statistical inference to draw conclusions and measure uncertainty Evaluate appropriateness of conclusions made from hypothesis tests	Determine when statistical inference topics should be taught	Create an opportunity to use hypothesis testing in real-life situation that is meaningful to students Model the use of probability rules in connection with inference procedures

AP Statistics Pedagogical Content Knowledge

The second construct that constitutes APSTK is Pedagogical Content Knowledge for teaching AP Statistics or APSPCK. APSPCK is defined as a teacher's knowledge of how her students learn or do not learn specific topics (Shulman, 1986) within A.P. Statistics. This includes being able to anticipate and interpret student thinking as it is happening (Ball, et al., 2008). Furthermore, a teacher's APSPCK allows her to recognize common errors and misconceptions (Ball et al., 2008; Schoenfeld and Kilpatrick, 2008; Shulman, 1986;) and adjust instruction, curriculum choices and assessment strategies

accordingly (Ball, 2008). Emerging from research over the past 25 years, APSPCK reflects the knowledge a teacher has of subject-specific student learning. Ball et al. (2008) states that, "many demands of teaching require knowledge at the intersection of content and students" (p. 401). Central to the idea of APSPCK, is a teacher's ability to not only recognize a wrong answer, but also decide which wrong answers students are most likely to make.

Ball (1988), Ball et al. (2008), Grossman (1990), Hill and Ball (2008), Schoenfeld and Kilpatrick (2008), Shulman (1986), and Van Driel and Barry (2010) state that pedagogical content knowledge is inherently subject-specific. Shulman (1986) described pedagogical content knowledge as the ways in which a teacher interprets a subject to make it more comprehensible to others. Furthermore, according to Shulman (1986), pedagogical content knowledge is the teacher's ability to identify errors and misconceptions, change explanations, and formulate the subject to accommodate shifting student interpretation, thus making the learning of specific topics easy or difficult. A key point here is that pedagogical content knowledge is specific to one's subject area.

Since Shulman's (1986) initial approach, the idea of pedagogical content knowledge has undergone many definitional revisions and re-conceptualizations. Similar to Shulman (1986), Ball et al. (2008) classifies knowledge of student misconceptions under the pedagogical content knowledge heading. However, Ball (Ball et al., 2008) goes on to say that pedagogical content knowledge is an "interaction between specific mathematical understanding and an understanding of pedagogical issues that affect student thinking" (p. 394). Therefore, pedagogical content knowledge requires the teacher to be able to anticipate and interpret students' thinking about a concept as it is

emerging or taking shape.

Furthermore, Ball et al. (2008) describes how pedagogical content knowledge is a teacher's ability to arrange and deliver instruction in a way that facilitates student learning. As previously mentioned, Ball's (1988) concept of "knowledge of learners and context" is the ability of a teacher to develop activities that facilitate student learning. This would include making appropriate curricular choices, sequencing topics in an order that makes sense, being familiar with how students think about a particular concept, and deepening student thinking to go beyond simple memorization. Finally, Schoenfeld and Kilpatrick (2008) concur that pedagogical content knowledge includes knowledge of typical student understandings and misunderstandings and of ways to deal with them. They say that a teacher needs to recognize which errors are frequent occurrences among students and prepare for it by developing a set of responses to help them see clearly what is wrong.

Watson and Callingham (2003) state that very little is known about how secondary school students learn statistics, how the misconceptions they bring to the subject impede their learning, and what should be taught or assessed. Furthermore, it has been found that most students have misconceptions about statistics that are difficult to change. Because of this, statisticians and educators have begun to model the intellectual activity required of students within the statistics classroom (Gardner, 2007). For this study, it was necessary to uncover research-based examples of secondary-level statistics student learning, because many components of APSPCK rely heavily on a teacher's understanding of student learning.

To begin, there are two primary frameworks established in the field of statistics education that outline statistical thinking and learning. First, the Pfannkuch and Wild (2002) model provides a comprehensive description of the processes involved in statistical thinking, from problem formulation to conclusion. The model has four separate dimensions (the investigative cycle, types of thinking, the interrogative cycle, dispositions) and claims that people think within each dimension simultaneously.

Next, the Ben-Zvi and Friedlander model (1997) seeks to define and characterize the levels of thinking that were observed for 13 to 15 year olds in a statistics teaching experiment. The model consists of four stages (Mode 0: Uncritical thinking, Mode 1: Meaningful use of representation, Mode 2: Meaningful handling of multiple representations/developing metacognitive abilities, Mode 3: Creative thinking) in which the categories emerged from students' work on data analysis and statistical investigations. While the Ben-Zvi and Friedlander model identified stages of development it was not determined if the students were going through the stages in any particular order. These two models begin to provide a sense of what statistical thinking is understood to mean among educators and statisticians (Garfield, 2007).

Several empirical studies have used the Pfannkuch and Wild (2002) and Ben-Zvi and Friedlander (1997) models as a basis for organizing student learning and thinking in statistics (Garfield, 2007; Garfield & delMas, 2010; Hubbard, 1997). For example, Garfield and delMas (2010) define students' statistical literacy as "understanding words and symbols and being able to read and interpret graphs and terms", statistical reasoning as "reasoning with statistical information" and statistical thinking as, "asking questions and making decisions involving statistical information" (p. 2).

Hubbard (1997) recommends techniques for constructing assessment questions that increase student learning by focusing on the understanding of statistical concepts over memorizing calculations, procedures and shortcuts. Therefore, open-ended questions are encouraged, because they highlight a student's understanding of concept. Hubbard's (1997) analysis suggests that if teachers assess students in this way they will, in turn, study at a deeper, more intrinsic level because they anticipate having to provide more than a pre-determined answer. Overall, teachers should use the following types of questions to assess students' statistical understanding: questions that reinforce overall goals of the course, questions that differentiate between deep and surface learning, questions that require more than a pre-determined answer, and questions that test understanding of computer printouts in a context or non-trivial way.

In a final example, Garfield (2007) asked students in an introductory statistics course questions regarding what types of statistics they learned and how they learned during the course. She found that students classified what they learned into three categories: statistics as basic mathematics skills, statistics as procedures for data and statistics as standards for research. Garfield also uncovered four categories of how students learned introductory statistics: being able to recall facts and procedures, being able to understand when, how and why concepts are used, being able to relate statistics concepts to life experiences, and being able to uncover attributes of other things using statistics. Her findings show that a hierarchical structure exists in students' conceptions of statistics and statistical thinking skills.

A more recent framework for statistical student learning was developed in response to the American Statistical Association's call for a conceptual framework

outlining K-12 statistical learning objectives. In 2005, Franklin, et al. created the American Statistical Association Guidelines for Assessment and Instruction in Statistics Education or GAISE. This guideline established recommendations for curriculum, instruction and student learning for K-12 statistics education based on the NCTM's *Principles and Standards for School Mathematics* (2000). GAISE's learning objectives are broken down into three levels, A, B, and C based on student development rather than age. Casey (2008) states that objectives for levels B and C are relevant to secondary students. These objectives are outlined in Table 5.

Table 5: GAISE Learning Objectives

Level B:	Level C:
Students become more aware of the statistical question distinction (a question with an answer based on data that vary versus a question with a deterministic answer).	Students should be able to formulate questions that can be answered with data.
Students make decisions about what variables to measure and how to measure them in order to address the question posed.	Students should be able to devise a reasonable plan for collecting appropriate data through observation, sampling or experimentation.
Students use and expand the graphical, tabular and numerical summaries introduced at Level A to investigate more sophisticated problems.	Students should be able to draw conclusions and use data to support these conclusions.
Students develop a basic understanding of the role that probability plays in random selection when selecting a sample and in random assignment when conducting an experiment.	Students should be able to understand the role that variability plays in the decision making process.
Students investigate problems with more emphasis placed on possible associations among two or more variables and understand how a more sophisticated collection of graphical, tabular and numerical summaries is used to address these questions.	
Students recognize ways that statistics is used or misused in their world.	

GAISE also provides six recommendations for statistics curriculum choices: emphasize statistical literacy and develop statistical thinking, use real data, stress conceptual understanding rather than mere knowledge of procedures, foster active

learning, use technology to develop conceptual understanding and to analyze data, and use assessments to improve and evaluate learning. Finally, Groth (2008) outlined five GAISE pedagogical recommendations that built upon current discourse themes within the statistics education community. They are outlined in Table 6.

Table 6: GAISE Pedagogical Recommendations

Statistical literacy should be a prominent curricular goal because of the central role it plays in democratic citizenship, personal choices, careers, and evaluating scientific findings.
Statistics and mathematics differ as disciplines. Statistics utilizes mathematics but should not be mistaken for a branch of mathematics.
The study of variability should have a central role in school statistics. Students should understand a variety of types of variability, including measurement, natural, induced, and sampling variability.
Statistical problem solving is heavily reliant upon context. It is not possible to give plausible interpretations of data without some knowledge of the context that generated them.
Pre-college experiences with statistics require an intuitive grasp of probability. Probability is an important tool in statistical analysis, but doing mathematical probability problems should not be mistaken for doing data analysis.

Based on the definition of APSPCK, relying on frameworks of student learning in statistics, GAISE recommendations and the ideas of Shulman (1986), Grossman (1990), Ball (1988, 2008) and others, this study attempted to measure APSPCK. Table 7 describes the measurable characteristics of APSPCK.

Table 7: Measurable Characteristics of APSPCK

Characteristics of APSPCK	**Adjust curriculum, instruction and assessment strategies**	**Recognize and interpret errors and misconceptions**	**Aid in student thinking and learning**
	Change instructional strategies based on students learning or not learning	Identify errors and misconceptions on student work, quizzes, projects and tests	Analyze student learning in a way that encourages conceptual thinking

Table 8 shows a comparison between Shulman's (1986) teacher knowledge constructs, Ball and Ball et al's. (1988, 2008) mathematics teacher knowledge constructs and the current definitions of APSCK and APSPCK.

Table 8: Comparison of Foundational Ideas and Current Constructs

	APSTK	Shulman (1986)	Ball (1988, 2008)
APSCK	Organize content Accurate representations* Conceptual understanding Student understanding Curricular decisions	Organize content Conceptual understanding Curricular decisions	Organize content Accurate representations Conceptual understanding Perceptions and Beliefs Student learning
APSPCK	Identify misconceptions/errors Connected to student learning Accommodate student interpretation Subject specific Assessment choices	Identify misconceptions/errors Connected to student learning Accurate representations Accommodate student interpretation Subject Specific	Identify misconceptions/errors Connected to Student learning Subject Specific Curricular decisions

*Terms shown in red highlight unique components of APSTK definitions

Empirical Evidence for the Conceptual Framework

Recent research has shown a difference between defining a construct and identifying it or measuring it in practice (Ball, 2002; Sorto, 2004). Ball (2002) found that many studies work to clarify the content knowledge necessary to teach particular concepts. However, according to Ball (2002), simply describing teachers' mathematical knowledge does not bring the field any closer to understanding what content knowledge teachers need. In fact, these studies do not help describe the type of knowledge teachers' need or how they use it in their work. Therefore, further research is needed to uncover what knowledge is sufficient for teaching statistics to secondary students, what pedagogical strategies and instruction enable teachers to construct that knowledge, and how teachers can draw on that knowledge to help their students develop similar knowledge. According to Peters (2009) these large goals require a collection of powerful strategies and tools to investigate the nature and development of statistics teacher knowledge.

Therefore, developing a methodology to measure and define APSTK required an initial uncovering of the current state of methodologies used to measure teacher knowledge to inform the conceptual framework and current effort to develop an APSTK instrument. In this review of research, only a few quantitative studies were found that measure teaching knowledge for statistics specifically, so the search was expanded to include studies that attempted to measure mathematics teaching knowledge in general. Table 2, shown above, includes the types of methodologies used in the studies examined for this part of the literature review.

To begin the analysis, Watson (2001) designed an instrument to serve as a profile of teacher achievement and teacher content knowledge needs with respect to the probability and statistics strands in the mathematics curriculum in Australia. The instrument was administered to 43 Australian mathematics teachers including 15 primary teachers and 28 secondary teachers with a follow-up interview of 21 teachers. The study attempted to address multiple domains of teacher knowledge including self-efficacy, beliefs and confidence in teaching statistical concepts, but also pedagogical content knowledge using items in which teachers responded to questions based on student survey items. A unique feature of this study is the instrument's attempt to go beyond a documentation of number of courses as evidence of subject matter knowledge.

In more recent work, Watson teamed up with Callingham and Donne (2008) to continue her examination of statistics teaching knowledge using an updated version of her previous instrument. This particular study considered the pedagogical content knowledge of 42 teachers (20 of which were high school teachers) in Australia selected to be part of a professional learning program in statistics entitled the StatSmart project. A profile for each teacher was established measuring many aspects of teacher confidence, beliefs, teaching practice, assessment practice, and background. In addition, pedagogical content knowledge was captured through responses to a teacher survey, and other survey items answered by students. An example of a pedagogical content knowledge item from this survey is:

> Briefly outline how you might teach "sample" to your students.
>
> • How would you introduce the topic?
>
> • What resources or materials would you use?

- How much class time would you spend?
- What teaching methods and groupings of students would you use?
- What lessons might precede or follow the topic?

These questions help in the conceptualization of items for the current study by providing examples of items that seek to capture and measure secondary-level pedagogical content knowledge for teaching statistics.

For this study, a Rasch model Item Response Theory (IRT) analysis was used to obtain a measure of teacher ability in relation to pedagogical content knowledge. This produced a model where all individual items showed good fit to the model, indicating that the instrument provided a defensible measure of a single construct, labeled pedagogical content knowledge. The IRT methodology had several limitations that will be discussed later in the limitations section of this chapter.

In an effort to further explore important studies pertaining to quantifying or measuring knowledge for teaching statistics, the review of research was widened to include studies that measured other mathematical teaching knowledge. Examining these studies helped identify a wider variety of methodologies used to study teaching knowledge. To begin, Krauss, et al. (2008) set out to investigate whether the distinction between "pedagogical content knowledge" (Krauss et al., 2008) and "content knowledge" varied according to the level of mathematical expertise among 192 current 10^{th} Grade Mathematics teachers in Germany.

Krauss, et al (2008) hypothesized that the association between the two categories of knowledge is substantially closer in teachers high in mathematical expertise (academic track GY teachers). The study utilized confirmatory factor analysis to examine multi-

group comparison (df = 16, $\chi^2 = 15.79$, $p = 0.47$) to examine latent means and correlations between the two groups. This analysis concluded that the degree of cognitive connectedness between pedagogical content knowledge and content knowledge in secondary mathematics teachers is a function of the degree of mathematical expertise, or more academic training in Mathematics.

In another example, authors Manizade and Mason (2011) researched how to measure math teacher pedagogical content knowledge for a specific topic in Geometry (measurement) by developing an instrument. The goal of this study was strictly to explore the feasibility of designing an instrument to assess teachers' pedagogical content knowledge and ensure validity. The researchers utilized a Delphi methodology where the development and administration of a survey are interconnected. A diverse panel of 20 experts, including mathematicians, mathematics educators, teachers, and teacher leaders, were asked to develop an instrument through multiple revisions of writing items until a consensus was reached on which items to include. Researchers were able to establish the instrument's content validity based on the experts' judgments that the content of the items was consistent with what was being measured. Limitations of the Delphi methodology will be discussed below in the limitations section of this chapter.

Another instrument designed to measure mathematics teacher knowledge is outlined in Tchoshanov's (2011) study entitled *Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics*. A sample of 102 middle school mathematics teachers (grades 6–8) was selected for the study from a number of school districts in an urban area in the

Southwestern US. Teachers were tested using a Teacher Content Knowledge Survey developed at the University of Texas.

The larger research study consisted of three small studies: first, a quantitative study (an initial sample of N=102 middle grades mathematics teachers) on the association between cognitive type of teacher content knowledge and student achievement; second, a mixed methods study (N=10, a subset of the initial sample) on the correlation between cognitive type of teacher content knowledge and lesson quality; and third, a case study (N=2, a subset of two teachers from the sub-study 2) on examining middle grades mathematics teachers' knowledge and understanding of fraction division

Correlation statistics were analyzed and a two-way chi-squared test of Independence ($df = 2$, $\chi^2 = 6.99$, $p = 0.03$) found a positive correlation between teacher's level of content knowledge and student achievement. In addition, it was found that teachers whose mathematical knowledge appeared to be connected and conceptual (Schoenfeld, 1985) were also more conceptual in their teaching. This study suggests that teacher knowledge of concepts and connections has the potential to be a good predictor of successful teachers who might positively impact middle grades students' mathematics achievement.

Margarita Cummings (2010) at the University of Utah completed a dissertation by developing her own teacher survey that was administered to 96 new (less than 5 years) secondary mathematics teachers to identify four proposed domains of teacher knowledge: "mathematical knowledge", pedagogical knowledge", "pedagogical content knowledge" and "curricular knowledge". Cummings (2010) used an exploratory principal components factor analysis ($\alpha = 0.05$) with Varimax rotation and Kaiser normalization to establish

four domains of knowledge: mathematical knowledge, pedagogic knowledge, pedagogical content knowledge and curricular knowledge and Cronbach's Alpha ($\alpha >$ 0.8) was examined to establish internal consistency. Items with rotated factor loads greater than 0.4 were retained. Consequently, 12 items were retained for mathematical knowledge, 4 items for pedagogical content knowledge, 4 items were retained for pedagogic knowledge and 3 items for curricular knowledge.

An example of a mathematical knowledge item topic is:

> Explain the difference between polynomial and exponential functions.

An example of a pedagogical content knowledge item topic is:

> Explain mathematics symbols in a way that helps students understand their mathematical meaning (ie. Helping students understand the difference between 2X, X^2 and 2^X).

An example of a pedagogic knowledge item topic is:

> Use effective verbal and nonverbal communication strategies to guide student learning and behavior.

An example of a curricular knowledge item topic is:

> Evaluate the usefulness and appropriateness of mathematics curriculum materials for your students.

Cummings' (2010) work to establish domains of secondary-level mathematics teaching knowledge aids in the conceptualization of teaching knowledge domains for more specific content areas; such as A.P. Statistics. In addition, the development of items that attempt to measure secondary-level mathematics teaching knowledge will help inform the development of items designed to measure A.P. Statistics teaching knowledge.

In 2006, Ferrini-Mundy and Senk developed test items and teaching knowledge constructs for Algebra, and piloted the test items for the Knowing Mathematics for Teaching Algebra (KAT) project at Michigan State University. They used the works of Shulman (1986), Ma (1999), Ball & Bass (2000) and others to develop a conceptual framework for teaching secondary-level Algebra. Ferrini-Mundy and Senk (2006) piloted the test items among 810 pre-service and in-service secondary mathematics teachers. They analyzed the results using item level difficulty parameters and descriptive statistics, to find three over-arching domains of KAT: Knowledge of School Algebra, Advanced Knowledge, and Teaching Knowledge. Their work in item development and knowledge construct validation for Algebra will also help to inform current efforts in item development and creation of the APSTK assessment.

In another attempt to clarify and/or create mathematics teacher knowledge constructs, Izsak et al. (2010) created a multiple-choice instrument to measure the mathematical knowledge needed for teaching arithmetic with fractions, decimals, and proportions. The instrument was developed for an ongoing NSF-funded research project called Does it Work? Building Methods for Understanding Effects of Professional Development (DiW). The final instrument consisted of 64 items, 45 taken from the middle grades Learning Mathematics for Teaching (LMT) project (Ball & Hill, 2007) and 19 from Izsak et al. (2010). The instrument was administered to a convenience sample of 201 middle grades mathematics teachers with follow-up interviews conducted with 16 of the same teachers. An Item Response Theory (IRT) mixed Rasch model was used to help identify two subgroups of middle grades teachers that differed in their understanding of arithmetic with rational numbers.

In another important work, Hill, Schilling and Ball (2004) outlined efforts to design and empirically test measures of teachers' content knowledge for teaching elementary mathematics. This analysis sought to answer two questions: how teachers' mathematical knowledge for teaching is organized, and whether developed items can reliably measure teachers' mathematical knowledge for teaching? Hill, Schilling and Ball's (2004) assessment was pilot tested in California's Mathematics Professional Development Institutes (MPDIs). These institutes were designed to increase California teachers' knowledge of subject matter in mathematics and served about 23,000 K–12 teachers in the first 3 years, and altogether 1,552 responses from all three forms were used to conduct statistical analyses.

An Exploratory factor analysis (EFA) (df. = 25, $\chi^2 = 44.82$, $p = 0.05$) of all items on the first form of the assessment suggested that there were three underlying constructs: (a) knowledge of content in number concepts and operations; (b) knowledge of content in patterns, functions, and algebra; and (c) knowledge of students and content in number concepts and operations.

From a measurement standpoint, these results suggested that common and specialized mathematical knowledge are related yet are not completely equivalent. So, the possibility exists that individuals might have well-developed common knowledge yet lack the specific kinds of knowledge needed to teach. Results also indicated that individuals might develop the specialized knowledge for teaching mathematics from teacher preparation, professional development, or working with students or curriculum materials, without having otherwise expert knowledge of mathematical content. Finally, Hill, Schilling and Ball (2004) suggest these results reveal that rather than focusing on

how much mathematics an individual knows, researchers must also ask how a teacher can use mathematical knowledge to generate representations, interpret student work, or analyze student mistakes.

Ball and Hill (2005) worked to confirm this framework in a later study with elementary mathematics teachers. A factor analysis showed evidence of the proposed specialized content knowledge for teaching (Ball & Bass, 2003) and in fact showed an even more specific mathematical knowledge pertaining to even smaller domains of number sense and operations.

In a final example of recent research designed to measure teacher knowledge for mathematics, a 2010 study entitled, "Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS)" by Saderholm et al. (2010) tested an assessment to categorize and measure four different knowledge types needed to teach middle school mathematics: algebra, geometry, number computation and probability and statistics. The DTAMS assessment, composed of 20 items, 10 multiple-choice and 10 open-response items, was sent to 1,700 middle school mathematics teachers in 17 states.

The DTAMS assessments were found to accurately measure middle-school level teacher knowledge for algebra and geometry. Confirmatory factor analysis was used to assess the fit of items within each of the separate structures. In addition, separate structural equation models (SEM) were constructed for each of the four knowledge type (for statistics and probability, df = 389, $\chi^2 = 1008.77$, $p = .01$) Through these analyses, Saderholm et al. (2010) proved the statistics and probability assessment did not accurately measure middle-school level teacher knowledge possibly due to the high level of statistical knowledge required to successfully complete the assessment.

Conclusion

The review of research on statistics teaching knowledge frameworks, statistics student learning, and quantitative methodologies used to measure mathematics teaching knowledge, is designed to aid in the conceptualization of the proposed framework for APSTK. Research on secondary-level statistics teaching knowledge proved difficult to find. In fact, many studies suggest the need to further explore how secondary-level students learn statistics, as well as the need for further research on teaching knowledge of A.P. Statistics teachers. In addition, studies using various quantitative methodologies for measuring mathematics teaching knowledge were explored to support the methodological choices for this study. These findings, along with the conceptual framework, will inform APSCK and APSPCK item development and subsequent analyses.

Why not Pedagogic Knowledge?

Ball (2000) suggested the need for a bridge between content knowledge and pedagogy and sought to answer three questions to fill the gap: how to determine which content matters for teaching, how subject matter must be understood to be usable in teaching, and how to create opportunities to learn to use subject matter knowledge in the context of practice. Ball et al. (2008) later stated that pedagogical content knowledge is the bridge between content and pedagogy, or the intersection of content and students. In addition, Lappan (2000) describes how teacher knowledge types are varying and changing, and do not act in isolation. As seen in many frameworks, teacher knowledge definitions often over-lap and interact, and separating constructs has become increasingly difficult.

For this study, it was thought that APSPCK combines knowledge of content and

knowledge of general pedagogy in a specific context of teaching A.P. Statistics. Once the context of the study was constrained to AP Statistics teaching and learning, general pedagogic knowledge is inextricable from APSPCK. Furthermore, knowledge of teaching strategies and instructional and curriculum choices cannot be separated from the student-learning component of APSPCK.

Final Thoughts on Conceptual Framework

Empirical research pertaining to the measurement of in-service, secondary-level statistics teaching knowledge is lacking. However, several different frameworks have been developed to serve as a foundation for teaching knowledge constructs and for insight into how students learn statistics. Student learning of statistics at the secondary level plays an integral part in defining and measuring APSPCK, therefore student-learning frameworks and recommendations impact the way in which APSPCK is conceived and measured in the current study. Furthermore, because APSCK, for this study, includes a teacher's ability to use her content knowledge in a way that increases student understanding, APSCK and APSPCK are thought to correlate in the conceptual framework for APSTK. In fact, a unique interaction construct is included in the conceptual framework because of the hypothesized nonlinear structure of APSTK.

Limitations of the Literature

Overall, there were several limitations found among the studies detailed in this review. From a topic perspective, there exists a lack of literature describing statistics teaching knowledge at the secondary level. This review found that most research has focused on elementary and college-level statistics teaching, and furthermore focuses primarily on content knowledge leaving out pedagogical content knowledge

considerations. From a methodology standpoint, limitations include qualitative work that has resulted in small sample size, and non-generalizability as well as quantitative work that has resulted in non-random samples, lack of item level development and analysis, or other psychometric concerns specific to a particular study. In general, a lack of variety among quantitative methodologies was discovered throughout.

Discussion of Limitations

To begin, many studies in this review made somewhat limited conclusions due to small sample size and lack of generalizability (Casey, 2010; Groth, 2008; Hassad, 2006; Izsak, 2010); Krauss, et al. 2008; Leavy, 2006, 2010; Liu & Thompson, 2007; 2010; Pfannkuch, 2006; Sharma, 2007; Schmidt, et al., 2009; Tchoshanov, 2011; Watson, 2001; Watson et al., 2008, 2011). For example, several studies used a case study methodology, focus groups, classroom observation and/or interview protocol to draw conclusions that were often compartmentalized and restricted (Casey, 2010; Groth, 2008; Leavy, 2010; Liu & Thompson, 2007, 2010; Pfannkuch, 2006; Sharma, 2007; Tchoshanov, 2011). Therefore, generalizability and, for some studies, external reliability may be compromised making it difficult to establish that a study's results will be consistent in other scenarios. Finally, many studies used a non-random sample, which may affect validity of the research findings (Hill, Schilling & Ball, 2004; Leavy, 2006; Liu & Thompson, 2007; Mills, 2007; Watson, 2001).

The next limitation noted was a lack of item development, alignment or analysis (Cummings, 2010; Watson, Callingham & Donne, 2008). Cummings (2010) noted that no item-level analysis was conducted and her study needed more items to help improve validation of the four constructs. In addition, Watson, Callingham and Donne (2008)

detailed the limitations of the types of PCK questions asked. It was felt the PCK questions were too basic or simplistic in nature, so they did not accurately reflect teachers' knowledge in this area.

There were several other limitations relating to psychometric or measurement concerns specific to the methodologies of individual studies (Cummings, 2010; Hill, Schilling & Ball, 2004; Izsak, 2010; Koi, Chai & Tsai, 2010; Krauss, et al., 2008; Manizade & Mason, 2011; Tchoshanov, 2011). For Cummings (2010), a main concern was the self-reported perceptions of preparedness from teachers, in addition to a low response rate among participants. Izsak, et al. (2010) emphasized a lack of information provided by the IRT Rasch model while Koi, Chai and Tsai (2010) called for a construct validation study to be done with in-service teachers rather than pre-service. In addition, Manizade and Mason, 2011 point to limitations of using a Delphi methodology. They noted the difficulty of finding and working with panels of experts who do not have enough time or interest to commit to the project.

Furthermore, Krauss et al. (2008) note that because the knowledge constructs were assessed in a standardized testing situation there is a lack of implication for authentic learning situations. Tchoshanov (2011) discovered that because the focus was solely on teacher content knowledge the findings were similar to those of other studies while Hill, Schilling and Ball (2004) felt that results were limited because they measured expert teachers rather than typical teachers.

Final Thoughts on Limitations

Overall, this review of literature found a distinct lack of research on the use of instruments for assessing secondary statistics teacher knowledge. Furthermore, it was

discovered that recent efforts to measure teacher knowledge in statistics education, as well as the larger field of mathematics have resulted in a lack of variety in measurement techniques.

These results confirm previous literature review attempts by Izsak and Orrill (2010) and Orrill (2010) to outline recent work in the area of using quantitative research methods to measure mathematics teacher knowledge. Specifically, Orrill's (2010) literature review efforts found that of 17 studies designed to measure primary and/or secondary mathematics teacher knowledge by using a test or assessment, the majority used some variety of correlation and/or regression techniques while only 1 study used Hierarchical Linear Modeling (H.L.M.) and none used Structural Equation Modeling (S.E.M.). These limitations suggest the need for further study into secondary-level statistics teaching knowledge and along with this, a need for more creative ways to measure content knowledge and pedagogical content knowledge for teaching statistics.

Observations and Notes

Most of the articles reviewed conclude with implications and recommendations for future research efforts. Recommendations ranged from advice for teacher preparation programs to ideas for professional development improvements as well as an increased focus on how teacher knowledge affects student learning. In general, the majority of research shows that teacher educators, teachers and educational researchers need to look further into the pedagogic knowledge necessary for teaching statistics at various levels (Casey, 2010; Groth, 2008; Hassad, 2006; Leavy 2010; Liu & Thompson, 2007; Phannkuch, 2006).

Many studies designed to quantify or outline components of content knowledge, also emphasize the next research step is to consider how a lack of content knowledge affects a teacher's pedagogic knowledge for teaching statistics (Groth, 2008; Hassad, 2006; Leavy, 2010; Saderholm et al., 2010). In addition, several academic dissertations on statistics teacher knowledge have revealed the fact that further research is needed to uncover what knowledge is sufficient for teaching statistics to secondary students and what pedagogical strategies and instruction enable teachers to construct that knowledge (Casey, 2008. Peters, 2009).

Other studies came to similar conclusions but also recommended future research by giving advice for teacher educators (Cummings, 2010; Hassad, 2006; Leavy, 2006; Mills, 2007; Schmidt, et al., 2009). In one example, Jamie Mills (2007) calls for the need to design a college-level course specifically for statistics teachers. Leavy (2006) suggests that college-level mathematics methods courses need to be more conceptual in nature.

Another group of studies recommended that future research focus on the need for increased professional development opportunities to provide support for developing teacher knowledge for mathematics and statistics (Hill, Schilling & Ball, 2004; Manizade & Mason, 2011; Watson, 2001; Watson, Callingham & Donne, 2008). To begin, Watson (2001) suggests that various sections of her instrument could be used to design individual professional development programs for teachers. In a later study, Watson, Callingham and Donne (2008) describe how there is a need for professional development to increase not only teachers' knowledge of the content but also their appreciation of how their students learn that content.

In addition, Manizade and Mason (2011) suggest that future studies should include a confirmatory factor analysis to ensure the reliability of each item of the instrument is essential. Moreover, an instrument designed in this way would allow subsequent professional development efforts to be evaluated.

Many researchers in this literature review call for future research to include how teacher knowledge constructs affect student learning and achievement (Hill, Schilling & Ball, 2004; Krauss, et al., 2008; Tchoshanov, 2011; Watson, Callingham & Donne, 2008). For example, Hill, Schilling and Ball (2004) recommend assessing teachers on pedagogic knowledge they will use in the classroom and use this information to gain more knowledge about student achievement. Furthermore, Krauss et al. (2008) claim that additional research is needed to examine precisely how pedagogical content knowledge and content knowledge regulate teaching behavior, and impact student learning. Similarly, Tchoshanov (2011) calls for a more integrated, complete picture of teacher knowledge to help describe how teacher knowledge impacts student achievement.

Finally, two studies had unique recommendations that did not fit into one of the previous categories. Manizade and Mason (2011) recommended creating individual rubrics for each item to measure teaching knowledge. Also, Izsak et al (2010) calls for the development of new instruments and types of psychometric analyses because of the potential to provide more detailed information not only about mathematical knowledge that teachers have already constructed, but also about the growth and change of that knowledge.

Need for APSTK Assessment

This literature review sought to uncover studies that examine statistics teaching knowledge and learning frameworks; as well as empirical research pertaining to measuring and validating secondary-level statistics teaching knowledge constructs. There were many gaps found in the literature with regard to secondary-level, in-service statistics teaching knowledge, and no literature was found that measures, or even describes knowledge for teaching AP Statistics. Therefore, there is a need to define a range, average and level of variability in APSTK that exists among current AP Statistics teachers.

Furthermore, research shows the need for more confirmatory analyses that work to validate statistics teaching knowledge constructs including those that relate to student learning. So, the APSTK survey, and resulting statistical analyses, seeks to validate the constructs of APSCK and APSPCK, and produce APSTK latent variable total scores in order to compare those scores with certain teaching characteristics of current AP Statistics teachers.

Chapter 3

Research Design and Methodology

Overview

The purpose of this chapter was to outline the process by which the current APSTK construct was developed and assessed to measure APSCK and APSPCK among current AP Statistics teachers. This chapter described item development and statistical analyses by using construct definitions and the conceptual framework, both developed in Chapter 2, to answer key research questions.

Initially, an APSTK assessment was developed consisting of twelve multiple-choice content knowledge questions to include an equal number of items for each sub-component of APSCK (exploring data, sampling and experimentation, anticipating patterns, statistical inference) as well as nine constructed-response APSPCK questions (see Appendix A for assessment). A sample of 500 current A.P. Statistics teachers was invited to answer these questions. These data helped to establish indicators of APSCK and the final set of items to measure APSPCK.

The updated APSTK assessment instrument was used to collect data from a second sampling of 1000 current A.P. Statistics teachers. Then, a measurement model (see Figure 3) and a structural model (see Figure 4) were evaluated using SEM. Finally, a latent variable APSTK total score (LVS) was calculated for each participant and used in a multiple regression analysis. The multiple regression analysis determined which teacher characteristic descriptor variables are related to APSTK LVS.

Re-Statement of the Problem

Increasing popularity of the A.P. Statistics course and growing numbers of students taking A.P. Statistics has produced a need for qualified teachers of A.P. Statistics. However, recent research efforts have shown that most secondary-level mathematics teachers are not comfortable teaching statistics, and are in fact not adequately prepared to do so (Franklin, 2010; Mills, 2007). Also, a distinct type of teaching knowledge and instruction is needed for statistics as opposed to Mathematics in general (Cobb and Moore, 2007; Rossman, 2006). Therefore, it is possible to teach Mathematics, but not possess enough knowledge, or the right types of knowledge, to teach A.P. Statistics.

Specifically, recent research has shown that a need exists to learn and understand more about teaching knowledge for secondary-level statistics. Yet, there was no established benchmark for the amount of teaching knowledge that is desirable or required of A.P. Statistics teachers. Therefore, there was a need to uncover a range, average and level of variability in APSTK among current A.P. Statistics teachers.

In addition, efforts to validate and measure A.P. Statistics teaching knowledge constructs have been lacking in the research community. In fact, existing research oftentimes uses qualitative measures that often produce inconclusive and un-generalizable results (Casey, 2010; Groth, 2008; Leavy, 2010; Liu & Thompson, 2007, 2010; Pfannkuch, 2006; Sharma, 2007; Tchoshanov, 2011). Also, there existed a distinct lack of research on the development and use of instruments to measure secondary-level statistics teaching knowledge, and lack of variety in psychometric techniques used to validate teaching knowledge constructs.

Therefore, the current study attempted to measure APSTK and validate a two sub-factor structure comprised of APSCK and APSPCK constructs. The study addressed the following research questions:

1. What items are suitable indicators APSCK and APSPCK for an assessment of APSTK?

2. Does a two-factor measurement model produce acceptable model fit?

3. What amount of shared covariance exists between APSPCK and APSCK in a two-factor measurement model?

4. Does a higher-order structural model produce acceptable model fit?

5. Does APSCK contribute to the overall measurement of APSTK?

5a. What effect does APSCK have on APSTK?

6. Does APSPCK contribute to the overall measurement of APSTK?

6a. What effect does APSPCK have on APSTK?

7. What is the latent variable APSTK score for each respondent?

8. What is an average level of APSTK among A.P. Statistics teachers?

9. Which independent variables are significantly related to APSTK score while controlling for other relevant variables?

APSTK Assessment Development

Steps for assessment construction were adapted from the processes outlined by Crocker and Algina (2008). As shown in Table 9, initial steps included identifying a purpose for the assessment and deciding teaching knowledge components and behaviors that match APSCK and APSPCK definitions. Next, items were developed, reviewed by a panel of experts and revised according to their feedback. The initial assessment was

administered to a sample of 500 teachers. Items were revised, edited or deleted from the survey based on the results of the item-level analysis. Then, a revised assessment was administered to another sample of 1000 teachers.

Table 9: APSTK Assessment Construction Process

Step 1: Identify the purpose for which the assessment scores will be used
Step 2: Identify behaviors that represent the construct
Step 3: Prepare a set of assessment specifications, delineating the proportion of items that should focus on each type of behavior identified in step 2
Step 4: Construct an initial pool of assessment items
Step 5: Have items reviewed (and revise as necessary)
Step 6: Hold preliminary item tryouts (and revise as necessary)

Purpose of the Assessment

The first purpose was to create an assessment designed to identify individual A.P. Statistics teacher latent variable scores (LVS). The Advanced Placement Statistics Teaching Knowledge (APSTK) assessment worked to capture a range, average and level of variability in teaching knowledge among AP Statistics teacher respondents. The second purpose for developing the assessment was to explore the relationships among teacher scores on the APSTK survey and teacher characteristic variables relating to educational background, years of experience teaching AP Statistics, and a self-reported

rating of student success on the A.P. Statistics exam.

Identify Behaviors

Initially, a set of construct behaviors and behavior characteristics were developed to begin to conceptualize the individual components of each construct.

Therefore, a list of behaviors that represent the constructs of Content Knowledge, Pedagogic Knowledge and Pedagogical Content Knowledge were compiled. Crocker and Algina (2008) recommend using techniques such as "expert judgment" (p. 68) to verify that specific assessment items do in fact make-up the construct being measured. In terms of "expert judgment", input was obtained from one or more individuals who have experience with the construct. For this study, the researcher has spent six years teaching A.P. Statistics and used this knowledge to help outline teacher behaviors that represent the definitions of APSCK and APSPCK.

Constructing an Initial Pool of Assessment Items

According to Crocker and Algina (2008), there should be a balance of items so that the different constructs, or components of a single construct, are represented in proportion to their importance. Therefore, the number of initial 12 multiple-choice APSCK items was balanced with an equal number per sub-domain (exploring data, sampling and experimentation, anticipating patterns and statistical inference). In addition, APSCK measurable characteristics outlined in Chapter 2 guided how many items are created per sub-domain of A.P. Statistics content (problem-solving knowledge, organizing content and student learning). As previously mentioned, APSCK items were developed to include content knowledge for teaching, and meant to identify more than College Board identified content knowledge for the subject.

Overall, an attempt was made to create at least three items per sub-domain of APSCK. Table 11 provides a specification of number of actual items (listed by item number) per measurable characteristics of APSCK.

Table 10: Specification Table for APSCK Items (Listed by Item Number*)

	APSCK Measurable Characteristics		
A.P. Statistics Content Sub-Domains	Problem Solving Knowledge	Organizing Content	Student Learning
Exploring Data	4	7	5
Sampling and Experimentation	1	6	10
Anticipating Patterns	2	11	9
Statistical Inference	3	12	8

*See Appendix A

In addition, twelve constructed-response APSPCK items were designed based on the definition of APSPCK established in Chapter 2. APSPCK items were written to reflect A.P Statistics curriculum while keeping in mind the work of theorists (Shulman,1986; Ball et al., 2008; Schoenfeld and Kilpatrick, 2008) as well as established statistics student learning frameworks, and GAISE (2007) recommendations. Table 12 provides a specification of number of proposed items per measurable characteristics of APSPCK.

Table 11: Specification table for APSPCK Items (Listed by Item Number*)

A.P. Statistics Content Sub-Domains	APSPCK Measurable Characteristics		
	Adjust curriculum, instruction and assessment strategies	Recognize and interpret errors and misconceptions	Student thinking and learning
Exploring Data		17	21
Sampling and Experimentation	19	16	
Anticipating Patterns	13	14	20
Statistical Inference	15		18

* See Appendix A

For both APSCK and APSPCK, assessment items were developed to reflect a measurable characteristic that matches a particular AP Statistics content area. In addition, an attempt was be made to create an equal number of items per sub-domain and measurable characteristic so that each category and definitional component was adequately represented. Finally, it was estimated that the expected time required of participants to complete all the items on the original survey would be approximately 45 minutes.

Item development considerations

According to the American Association for the Advancement of Science (AAAS) Project 2061 Symposium on Assessment National Association for Research on Science Teaching (NARST) Annual Conference (2007), item comprehensibility, test-wiseness, and appropriateness of task context issues are some of the most common threats to construct validity. Specifically, test items may contain construct irrelevant features that

make it difficult to rely on the answers that examinees choose. In general, teacher respondents should choose the correct answer when they know the idea, and they should choose an incorrect answer when they do not know the idea. Specific guidelines for writing the multiple-choice APSCK items can be found under the Phase I APSCK Items heading below. A more detailed description of the AAAS NARST conference recommendations for construct validity is as follows:

Comprehensibility.

The NARST annual conference (2007) concluded that in order to make valid interpretations of what a teacher knows, a test question must be understandable. Therefore, to ensure that valid interpretations of teacher knowledge it must first be clear what question is being asked. Furthermore, the question must use familiar general vocabulary that is clearly defined and cannot use unnecessarily complex sentence structure or ambiguous punctuation that makes the question difficult to comprehend. Finally, the question cannot include words and phrases that have unclear, confusing, or ambiguous meanings. APSCK and APSPCK items were developed with these principles in mind.

In addition, consideration was given to ensure that items include accurate information (including information on diagrams and data tables), so are not confusing to teachers who have a correct understanding of A.P. Statistics. For example, respondents should not have to spend time wondering if there is a mistake in the question. Finally, diagrams, graphs, and data tables were designed to be clear and comprehensible.

Appropriateness of task context.

As items were developed, consideration was given to the appropriateness of the

context. According to the American Association for the Advancement of Sciences (AAAS) Project 2061 (NARST, 2007), in order to make valid interpretations of what teachers know any item that involves a situation with people, objects or events in the real world or otherwise is checked to be understandable, interesting, and sensible to the A.P. Statistics teacher. Finally, one group of respondents should not be advantaged or disadvantaged relative to another.

Analyzing test-wiseness.

To make valid interpretations of APSCK and APSPCK, items were written so that they are not answerable by using test-taking strategies that do not depend on knowing the ideas being tested. According to the AAAS Project 2061 NARST conference (2007), test-wiseness addresses characteristics of the task that might allow respondents to make a satisfactory response using only general test-taking skills without understanding the idea being tested or mislead respondents into choosing an incorrect answer. For example, in multiple-choice questions, one answer (especially the correct answer) should not be distinctly different from the other answers. Finally, items were checked so that qualifying terms such as "always," "never," or "everyone" was avoided as well as logical opposites, or other obvious distractors that may make it easy to eliminate answers.

Item Review

According to Crocker and Algina (2008), a content validation study seeks to assess whether the items adequately represent a construct of particular interest. In order to prepare for a content validation study, this study used a panel of independent experts to judge whether proposed assessment items represent knowledge constructs pertaining to

knowledge for teaching A.P. Statistics. Crocker and Algina (2008) suggest the following four steps to be undertaken during the construction of a content validation study:

1. Define the behavior or performance domain of interest
2. Select a panel of qualified experts in the content area
3. Provide a structured framework for the process of item alignment
4. Collect and summarize the content alignment feedback from the experts

To this end, two current A.P. Statistics teachers as well as two professors in mathematics education at two different local universities vetted the initial items designed to assess APSCK and APSPCK. Items were vetted for content alignment criteria set forth by the American Association for the Advancement of Sciences (AAAS) Project 2061 to ensure construct validity in terms of comprehensibility, appropriateness of task context and test-wiseness. The following list outlines the alignment criteria questions provided to reviewers for each item:

- Is it clear what the question is asking?
- Are mathematical and statistical representations accurately stated?
- Does the question use familiar general vocabulary (not mathematical terminology)?
- Does the question use unnecessarily complex sentence structure or ambiguous punctuation that makes the question confusing?
- Does the question use words and phrases that have unclear, confusing, or ambiguous meanings?
- Does the question include inaccurate information that may be confusing to teachers who have a correct understanding of the subject?

- If present, are diagrams, graphs, and data tables clear and comprehensible?
- Is the context of the question familiar to A.P. Statistics teachers?
- Is the context of the question easy to understand so that respondents do not have to spend a lot of time trying to figure out what the context means?
- Is the information and quantities that are used reasonable and believable?

Each reviewer analyzed 30 potential APSCK items and 18 potential APSPCK items. Each item was deemed acceptable if at least three out of four reviewers identified it as passing 75% of the alignment criteria questions. Items that were deemed acceptable by 3 out of 4 reviewers on less than 75% of the alignment criteria questions were revised. Items that were not deemed acceptable by 3 out of 4 reviewers were not included in the initial item review.

Participant Selection

In order to contextualize the population of interest for this study, it was necessary to differentiate between the general population of current high school statistics teachers and the study population of current Advanced Placement Statistics teachers. For this study, a sample of potential participants was selected from the population of current A.P. Statistics teachers who have passed College Board audit procedures. In order to pass the College Board's audit process a teacher must submit an AP Statistics course syllabus that explicitly outlines how the course meets certain requirements by describing what content is covered and by detailing assignments and assessments given.

A public use database containing school information for each audit-approved A.P. Statistics course can be found within the Advanced Placement course ledger at https://apcourseaudit.epiconline.org/ledger/. This database includes a search engine that

will allow the researcher to select all schools with a College Board audit approved AP Statistics program for current or previous years. A systematic sample was drawn to determine an initial group of 500 current AP Statistics teachers across the United States. For Phase II, a new sampling of 1000 teachers was asked to respond to the revised questions. Attention was given to drawing a representative sample based on state-by-state geographic area only. Information concerning years of experience teaching, gender, or educational background are not given on the College Board database, so any particular similar or dissimilar characteristics among those who were invited, or agreed to participate, were purely coincidental.

Cook (2000) estimated that the response rate for a web-based survey or assessment is usually around 40-50%. Therefore, the desired sample size for each phase was between 200-250 (Phase I) or 450-500 (Phase II) teacher participants. To encourage participation, potential respondents were reminded of the needs and benefits completing the survey might bring to the general A.P. Statistics teacher population. Also, potential participants were informed that the expected time to complete all of the items was approximately 45 minutes for the revised assessment. They also had the option of stopping and starting the survey at their convenience. These accommodations are discussed in more detail in the Institutional Review Board (IRB) section at the end of the current chapter.

Data Collection and Analyses

Overview

The research design and analysis consisted of three significant phases. Phase I began by collecting data from 500 teacher participants on APSCK and APSPCK items.

An item level analysis was conducted on the APSCK items to establish support for four unique subscales, and a score was calculated for each. APSPCK items were analyzed and actual teacher responses were used to change constructed-response items into multiple-choice items to improve response rates.

For Phase II, the edited and revised items were administered to a new sample of 1000 teacher participants. In Step 1 of Phase II, a confirmatory factor analysis (CFA) was conducted on the new items to confirm a measurement model. In Step 2 of Phase II, latent variable path analyses were examined to confirm a structural model. Finally, Phase III established a range, average and level of variability in overall APSTK among current AP Statistics teachers by calculating a latent variable score for each respondent. Then, individual APSTK scores were compared to other participants' scores to establish a range of overall APSTK. In addition, individual APSTK assessment scores were analyzed with respect to several teacher characteristics variables collected from each respondent pertaining to educational background and teaching experience.

Phase I: Item Development and Analysis

The Phase I design began by inviting 500 potential respondents to participate in the online APSTK survey. It was expected that at least 200 examinees would agree to answer an initial set of 12 (three for each subscale) dichotomous content knowledge items and 9 constructed-response pedagogical content knowledge items. According to Crocker and Algina (2008) item level analysis can be conducted with relative stability for samples of at least 200 respondents (p. 322). Therefore, it was hoped that at least 200 current A.P. Statistics teachers would agree to participate.

APSCK items.

For the design portion of Phase I, APSCK items were presented in multiple-choice format. Item writing guidelines for APSCK multiple-choice questions followed rules set forth by Halaydna and Downing (1989) and Frey (2005) regarding general item construction, content and answer options. For example rules such as, "using a best answer format", "minimize examinee reading time" and "avoid trick items" (Halaydna & Downing, 1989) were used during item construction. For item content, rules such as "base each item on an educational or instructional objective", "use multiple-choice to measure higher level thinking", "and keep items independent of one another" were also used (Halaydna & Downing, 1989). For answer options, rules such as "answer options should only include one correct answer", "answer options should be homogenous", "there should be 3-5 answer options", and "negative wording should not be used" were considered (Frey, 2005).

For the analysis portion of Phase 1, APSCK items were scored dichotomously and examined using item-level analysis. Respondents were given a correct score if they chose the best possible answer or an "incorrect" score if they chose an answer other than the best answer. A numeric score of "1" per item was given to correct answers, and a "0" was given for incorrect answers. Total scores for APSCK were added together and combined with an APSPCK score to determine a total latent variable score for APSTK.

Also, descriptive statistics were gathered at the APSCK item level to determine item mean and variance. Also, APSCK assessment items were examined for both item level difficulty and discrimination to determine if an item should be kept or eliminated. An item's difficulty level was defined as being the proportion of examinees that

answered the item correctly, and produced a value in a range from 0.00 to 1.00. So, the easier an item was to respondents the higher item difficulty the item had. For example, an item answered correctly by 85.00% of respondents had a p-value of 0.85, while an item answered correctly by 50.00% of respondents had a lower p-value of 0.50. An item's difficulty level essentially controls the item's variance (σ^2) because variance $\sigma^2 = p_i q_i$ where p_j = (number of persons with a score of 1 on item j)/ N and $q_j = (1-p_j)$.

In addition, it was assumed that there exists a constant degree of correlation among items, therefore total assessment score variance was considered to be maximized when item difficulty is $p_i=0.50$. According to Crocker and Algina (2008) item difficulties (or *p*-values) for score interpretation usually fall into a range of 0.60 to 0.80 (p. 312). For this study, an acceptable difficulty level for an item was set at 0.40 to 0.70. Therefore, items falling below or above this range did not pass the item difficulty analysis and were dropped from further analysis.

For those items that passed the item difficulty analysis, an item discrimination value was calculated using a criterion-based index of discrimination (D). Because the purpose of the assessment was to provide information about individual differences on how the content knowledge construct was measured by the content knowledge questions, a discrimination parameter was used to examine how effectively an item discriminates between respondents who have a relatively high level of content knowledge from those with a relatively low level (Crocker and Algina, 2008).

For this purpose, an initial APSCK assessment score was established for each individual respondent. For Phase I only, the purpose of calculating an individual APSCK total assessment score was solely meant to establish a criterion score distribution (or cut

scores) to determine the item level discrimination parameter (D). Therefore, in this phase examinees were split into two groups: those who scored below and above the cut scores. Again, these cut scores were used for item-analysis purposes only, and not meant to imply an average level or range of teaching knowledge among respondents.

Determining the upper and lower bound or cut score has been done several ways in past research. For example, Kelly (1939) found a stable index discrimination index to include the upper 27% and the lower 27% of the examinees total assessment scores (in Crocker & Algina). In another analysis, (Beuchart and Mendoza (1979) and Englehart (1965) found that when the sample size is relatively large the same results can be obtained by having an upper and lower bound of 30% and 50%. For the APSTK assessment, the cut score was set at 30% for the upper group and 50% for the lower group. So, for the purpose of item-analysis only, those respondents who scored in the upper 30% possess an adequate level of APSCK and those that scored in the lower 50% do not.

Once the upper and lower groups were determined the index of discrimination is calculated as $D = p_u - p_i$ where p_u was the proportion in the upper group who answered the item correctly and p_i was the proportion in the lower group who answered the item correctly. The value of D ranged from -1.00 to 1.00 where positive values indicated that the item discriminates in favor of the upper group and negative values indicated that the item favors the lower scoring group. Specifically, the following guidelines (Ebel, 1965) as quoted in Crocker & Algina (2008) will be used:

If $D \geq 0.40$ the item if functioning satisfactorily

If $0.30 \leq D \geq 0.39$, little or no revision required

If $0.20 \leq D \geq 0.29$ the item is marginal and needs revision

If $D \leq 0.19$ the item should be eliminated or completely revised

According to Crocker & Algina (2008), using the D-value was acceptable because of ease of computation and straightforward interpretation. Examining correlations ensured items measure the same thing before they were combined to make up individual subscale scores. Once support was shown for each subscale, items scores were added together to get four total-subscale APSCK scores.

APSPCK Items.

In addition, a Likert-scale was used to score APSPCK constructed response items. Table 12 outlines a general scoring rubric; however, each APSPCK item had a unique corresponding scoring rubric assigned to it. The APSPCK item scoring rubric format was adapted from the scoring rubric used to grade free-response questions on the A.P. Statistics exam from 2012. This kept the assessment of APSPCK items closely aligned with student expectations on the AP Statistics exam. Finally, the College Board (2012) Scoring Guidelines suggest that "essentially correct", "partially correct" or "incorrect" response ratings depend on the particular question being asked. The same held true for APSPCK responses in the current study. For example, if an item had three sections and all three sections were answered correctly, the grade given was a "4" for a complete response.

Table 12: General APSPCK Item Scoring Rubric

Score/Description	Ratings
4 / Complete Response	All three parts* essentially correct
3 / Substantial Response	Two parts essentially correct and one part partially correct
2 / Developing Response	Two parts essentially correct and one part incorrect OR One part essentially correct and two parts partially correct OR One part essentially correct and one part partially correct OR All three parts partially correct
1 / Minimal Response	One part essentially correct and two parts incorrect OR Two parts partially correct and one part incorrect

* Assuming a three-part question

Inter-rater reliability.

To increase reliability, there were two raters of the APSTK survey items to determine individual scores. Specifically, an attempt was made to have 10-20% of all responses rated by both raters. In order to measure the level of agreement among raters, a Cohen's Kappa (κ) coefficient (Cohen, 1960) was calculated and analyzed. Cohen's Kappa was defined as:

$$\kappa = \frac{Pr\ (a) - Pr\ (e)}{1 - Pr\ (e)}$$

where Pr (*a*) was defined as the observed percentage of agreement and Pr (*e*) was defined as the expected percentage of agreement. In general, a Cohen's Kappa value of 1.00 implies perfect agreement and -1.00 implies imperfect agreement. For the current study, a value of 0.70 was considered a satisfactory percentage of agreement among raters.

Phase II: Measurement and Structural Model Analyses

A Confirmatory Factor Analysis (CFA) was conducted using Structural Equation Modeling (SEM) to evaluate the APSTK measurement and structural models. Joseskog and Sorbom (2001) claim the objective of SEM is to test the hypothesis that a population covariance matrix, £, is equivalent to an observed covariance matrix, S, that is produced from collected data on the constructs. Furthermore, Schumacker and Lomax (2010) state, "the goal of structural equation modeling is to achieve a parsimonious model with few substantive meaningful paths and a non-significant chi-square value close to the saturated model value of zero." For the confirmatory phase, the null hypothesis is that the model does fit the data well while the alternative hypothesis is the model does not fit the data well.

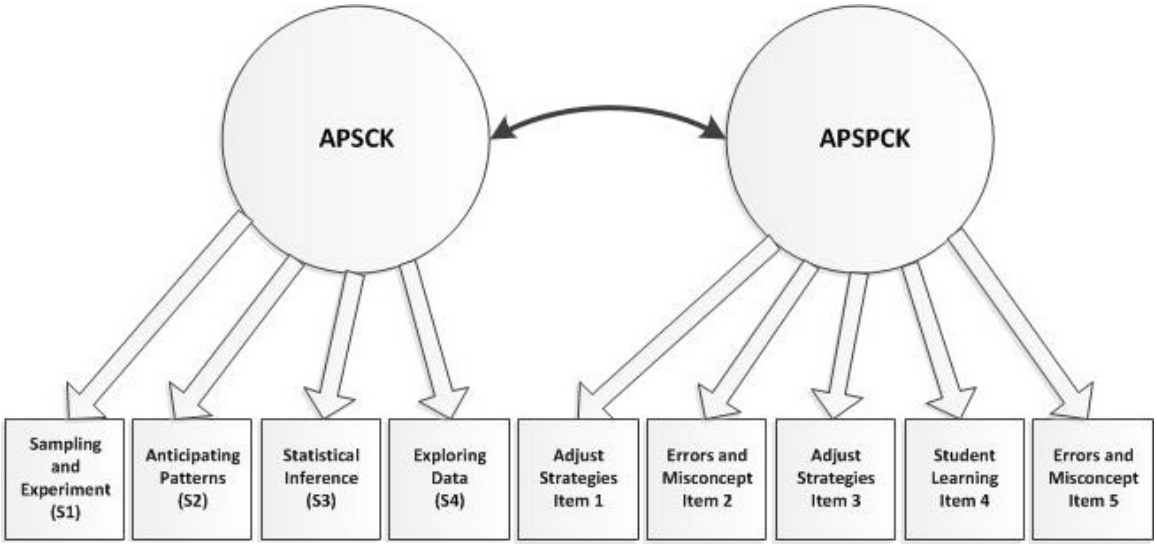
Schumacker and Lomax (2010) suggest 5 steps for conducting SEM: Model Specification, Model Identification, Model Estimation, Model Testing and Model Modification. Model Specification involves determining the best possible theoretical model that generates the sample variance-covariance matrix. Model Identification is necessary to resolve any problems pertaining to the number of observations compared to the number of free parameters in the model. Model Estimation requires the researcher to decide which estimation technique will be used to estimate the population parameters from sample data. Model Testing is designed to establish how well the theoretical model is supported by the sample data. Model Modification is recommended when the theoretical model proves to be not acceptable.

For Phase II a second group of 1000 potential teacher respondents were invited to respond to the revised items from Phase I. Again, it was thought that approximately 200+ complete surveys would be returned. The first step of Phase II was to conduct a Confirmatory Factor Analysis (CFA) to confirm a measurement model. The second step of Phase II was to conduct a latent variable path analyses in order to confirm a structural model using LISREL v8.80 student edition software. The third step of the Phase II analysis was to calculate individual latent variable APSTK scores for each respondent.

Step 1: measurement model evaluation.

To begin, a CFA was conducted to evaluate a measurement model as shown below in Figure 3. The measurement model was assumed to represent the best possible fit given the data. Therefore, it was used to examine and confirm theoretical relationships and further establish validity and reliability.

Figure 3: APSTK Measurement Model



Validity considerations.

To evaluate validity in a confirmatory setting , Hancock and Mueller (2006) recommend examining the variance extracted or the average squared standardized loading of a factor's indicators. Therefore, the variance-extracted coefficient = $\frac{\sum_{i=1}^{p} l^2}{p}$ will be calculated for each latent variable in order to explain the percent of variance in the indicators due to the latent variable.

Reliability Considerations.

With regard to reliability in a CFA, Hancock and Mueller (2006) recommend reporting the Coefficient H construct reliability value (or maximal reliability index) H = $\frac{1}{1+\frac{1}{\frac{l^2}{1-l^2}+\cdots\frac{l^2}{1-l^2}}}$ to estimate the "correlation the factor is expected to have with itself over repeated administrations" (p. 381). Therefore, Coefficient H was calculated for both APSCK and APSPCK.

Measurement model fit.

The measurement model was over-identified because there are 45 observations and 19 parameters. The measurement model was tested and model-fit indices were evaluated to verify that the hypothesized model was supported by the observed data.

Next, appropriate model fit indices such as the Standardized Root Mean Square Residual (SRMR), the Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI) (Hancock and Mueller, 2006) were examined. Many researchers suggest that judgments about the fit of the model should be based on a variety of different goodness-of-fit indices (Schumacker and Lomax, 2010). For the present study, model fit will be based on several indices: Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Goodness of Fit Index (GFI), Standardized Root Mean Square Residual (SRMR) and Nonnormed Fit Index (NNFI).

The Root Mean Square of Approximation (RMSEA) procedure developed by Steiger (1990) is an index that assesses how closely the model fit approximates the population model. Schumacker and Lomax (2010) defines a model with close fit as one in which the RMSEA value is less than or equal to 0.05. However, Brown and Cudeck (1993) and Schermelleh-Engel et al. (2003) indicate that an RMSEA value between 0.05 and 0.08 can be considered as acceptable fit.

The Comparative Fit Index (CFI) developed by Bentler (1990), is a fit index that works well with small samples and it is less affected by sample size. The CFI values can range from zero to one, with values closer to one indicating better fit. According to Schumacker and Lomax (2010) values of 0.95 can be considered as acceptable, where as values > 0.97 indicate a well-fitted model.

One of the first fit indices proposed was the Goodness of Fit Index (GFI) index to highlight the percent of variance explained by the model, with values closer to one indicating better fit (Joreskog, 2001). The Standardized Root Mean Square Residual (SRMR) describes the standardized mean absolute value of the covariance residuals. Higher SRMR values indicate worse fit, so a value less than 0.05 was considered acceptable.

The Nonnormed Fit Index (NNFI) was developed by Bentler and Bonnet (1980) in an effort to fix the sample size sensitivity. Values of the NNFI generally will be between zero and one, however, because this index's values are normed, the values may exceed one (Schermelleh-Engel et al., 2003). Recent simulation studies suggest that the NNFI is more robust than the NFI. Table 13 details acceptable levels for fit indices used in this study.

Table 13: Acceptable Levels of Fit Indices

	RMSEA	NFI	NNFI	GFI	SRMR
Acceptable Level	0.05-0.08	1	≥1	1	<0.05

For the measurement model, an adjusted chi-squared statistic was also calculated. The APSPCK items were Likert-type responses, and without necessary data transformations of Likert-type responses, SEM will produce biased results (Joreskog, 2002). According to Schermelleh-Engel et al. (2003) and Satorra and Bentler (1999), the chi-square test statistic is problematic when used with data that is not multivariate normal, it is sensitive to sample size, and the chi-square test statistic value decreases as model complexity increases; therefore, it is affected by the number of parameters in the model.

As an alternative, the Satorra-Bentler scaled chi-square test adjusts the maximum likelihood estimators downward, by a constant value, which reflects the degree of the observed kurtosis, in an effort to minimize the effects of non-normality. So, the Satorra-Bentler scaled chi-square test statistic was reported here to highlight necessary adjustments made for non-normal data produced by the likert-type APSPCK items.

Finally, the squared multiple correlation coefficient, R^2 was also evaluated. R^2 gives the variance ratio (true score to observed score) for a specific item, and indicates the degree to which each item adequately measures its respective underlying construct (Bollen, 1989). In other words, R^2 is the percentage that a dependent variable is explained by other variables or the proportion of variance in the measured variable that is "explained" by variance in the latent variable(s). For the measurement model, it represents item reliability. According to Fomell and Larker (1981) items with squared multiple correlations of .50 or greater are demonstrate acceptable levels of reliability.

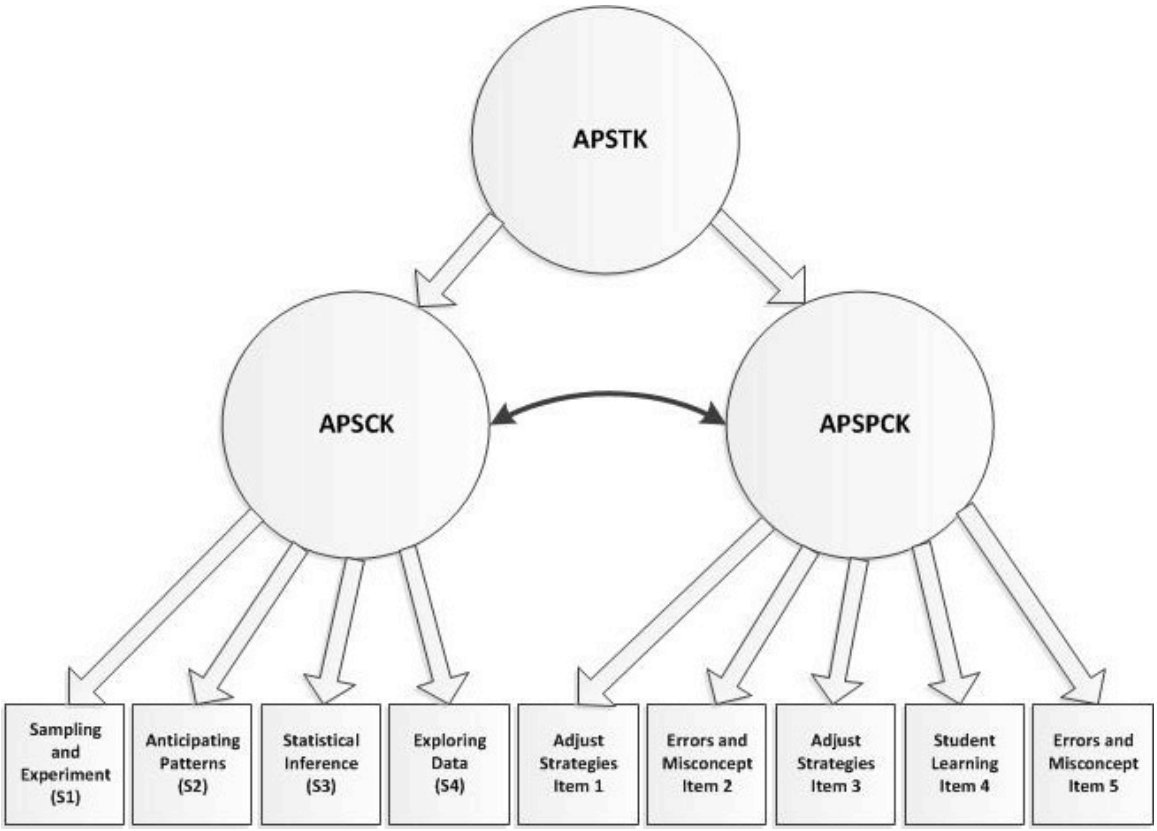
Step 2: Structural Model Evaluation.

After estimating a measurement model in CFA, the next step was to assess how well the model matches the observed data in a structural model (see Figure 5 below). Byrne (1998) suggests that due to the complex nature of the analytic process in SEM, changes to a hypothesized model should be made due to model misfit only when a sound theoretical basis exists for the change. For the APSTK model, it was thought that there were no theoretical changes in the 2 sub-components of APSTK.

The current structural model was over-identified with forty-five observations and twenty-two parameters. Structural model fit indices were analyzed to determine model fit. In addition, parameter estimates, *t*-values and R^2 coefficients were analyzed to

determine which latent variables (APSCK, APSPCK) had the most impact on overall APSTK.

Figure 4: APSTK Structural Model



Step 3: APSTK Latent Variable Scores.

When adequate APSTK structural model fit was achieved, a total APSTK LVS was calculated for each respondent. According to Anderson and Rubin (1956), calculating a latent variable score for each respondent will remove error as opposed to calculating a traditional total test score. The four total subscale APSCK scores from the second sampling of teacher respondents will be added together to obtain an individual latent variable APSCK score. In addition, APSPCK item scores from the second sample will also be added together to produce an individual latent variable APSPCK score.

Finally, these two scores will be added together to produce a latent variable APSTK score for each respondent.

Phase III: Exploring Teacher Characteristic Variables

Individual teacher APSTK assessment LVS were compared to one another in order to establish a range, average and level of variability in APSTK among current AP Statistics teachers who responded to the survey. Then, a multiple regression analysis was conducted to evaluate how well predictor variables such as "Number of Years Teaching A.P. Statistics", "Number of College-Level Statistics Courses Completed", and "Having a Degree in Statistics" predicted the criterion variable of APSTK LVS.

Multiple regression coefficients were calculated using the SPSS v22 linear regression option. The sample multiple correlation coefficient (R^2) was analyzed to determine the relative strength of individual predictors. This coefficient also indicated how much variance in the dependent variable (APSTK survey score) each independent variable explains or predicts. In addition, the F critical value and F test statistic were examined to determine if individual predictor variables had a statistically significant and/or practically significant impact on the APSTK score (Cohen, et al. 2003).

Institutional Review Board Considerations

Potential participants were informed that the benefit from the study comes from the importance of research knowledge gained about the types and levels of statistical knowledge used by current A.P. Statistics teachers. In addition, the benefit of doing this study was to help determine if current A.P. Statistics teachers have content mastery of teaching knowledge in A.P. Statistics in addition to a license to teach mathematics. In general, it was hoped that the knowledge gained from this study would help support

future preparation and professional development needs of A.P. Statistics teachers. Specifically, the field of statistics would continue to grow in importance and as a separate discipline from Mathematics by knowing more about knowledge teachers' use when teaching A.P. Statistics.

For this study, teacher participants were assured they would experience minimal risk while participating. The probability and magnitude of harm to the participants was minimal, and their privacy and confidentiality was ensured through anonymous online log in identification and password protection. Potential participants understood that willingness to participate in the study was completely confidential, and completion of the survey was not timed in any way. However, participants were notified that there were 25 initial items, and it was estimated that completion of the initial survey would take approximately 45 minutes. For the revised survey, participants were notified that there were 12 items, and it was estimated that completion of the survey would take approximately 20 minutes.

Also, each participant could have chosen at any point to stop participating in the survey. Furthermore, a participant's refusal to answer one question did not preclude him from answering a subsequent question on the survey. The researcher was able to see number of surveys completed, but not determine which particular survey an individual participant completed.

Each participant was given an individual, anonymous log in code to complete the survey and teacher characteristics questions. This log in code ensured that each individual could not complete more than one survey, but it also let the researcher know when a survey was complete. In addition, the researcher was not able to tell which survey

belonged to which participant. Individual responses were automatically separated from log in information once the respondent submitted his/her answers.

Chapter 4

Results

This chapter presents the results of the three methodological phases outlined in Chapter 3. Phase I included item development, initial data collection and item level analysis of APSCK and APSPCK items. Data from this phase served to establish measurable subscales of APSCK and item level scores of APSPCK items.

Next, a Confirmatory Factor Analysis (CFA) was completed using a new set of data collected from a second sample of current AP Statistics teachers to confirm a measurement model. Then, a full structural model was evaluated to determine the relationship among the latent variables APSPCK, APSCK and overall APSTK. Finally, an APSTK latent variable score (LVS) was calculated for each participant and compared to teacher characteristic descriptor variables using a multiple regression analysis for Phase III.

The results outlined in this chapter sought to answer the following research questions:

1. What items are suitable indicators of APSCK and APSPCK for an assessment of APSTK?
2. Does a two-factor measurement model produce acceptable model fit?
3. What amount of shared covariance exists between APSPCK and APSCK in a two-factor measurement model?
4. Does a higher-order structural model produce acceptable model fit?
5. Does APSCK contribute to the overall measurement of APSTK?
    5a. What effect does APSCK have on APSTK?

6. Does APSPCK contribute to the overall measurement of APSTK?

6a. What effect does APSPCK have on APSTK?

7. What is the latent variable APSTK score for each respondent?

8. What is an average level of APSTK among A.P. Statistics teachers?

9. Which independent variables are significantly related to APSTK score while controlling for other relevant variables?

Phase I: Item Development and Analysis

The initial APSTK assessment consisted of 12 multiple-choice content knowledge items that included an equal number of items for each sub-component of APSCK (exploring data, sampling and experimentation, anticipating patterns, statistical inference) as well as 9 constructed-response APSPCK questions (see Appendix A). A sample of 500 current AP Statistics teachers across the United States was invited to answer these questions through an online assessment. However, only 465 assessments reached the intended respondents due to incorrect or invalid email addresses listed in the College Board database.

Initial assessment data aimed to establish subscales of APSCK and the final set of items to measure APSPCK. The initial data collection resulted in 202 partial and 108 complete assessment responses. Response rates were approximately 30.00% for the multiple-choice APSCK items and 10.00% for the constructed-response APSPCK items. Many attempts were made to increase response rates over a period of ten weeks. Two reminder emails were sent to potential respondents as well as a reminder email to those who started the assessment but did not finish. Respondent characteristics and descriptive statistics were detailed in Table 26, found later in this chapter.

APSCK Item Analysis

Next, the difficulty coefficient and discrimination parameter for each item were calculated. During this analysis, emphasis was placed on maintaining at least two items per sub-group, so that one subscale score was not based only on one item. As stated in Chapter 3, an item's difficulty level is defined as being the proportion of examinees that answer the item correctly, and will produce a value in a range from 0.00 to 1.00. So, the easier an item is to respondents the higher item difficulty the item has. For this analysis, an acceptable difficulty level for an item will be set at 0.40 to 0.70. Therefore, items falling below or above this range will not pass the item difficulty analysis and will be dropped from further review.

For those items that passed the item difficulty analysis, an item discrimination value was also calculated using a criterion-based index of discrimination (D). Respondents were split into two groups: those who scored below and above the cut scores for the APSTK survey. The cut score was set at 30.00% for the upper group and 50.00% for the lower group. So, those respondents who scored above 70.00% possess an adequate level of APSCK and those that scored below 50.00% do not. Once the upper and lower groups were determined the index of discrimination was calculated as $D = p_u - p_i$ where p_u is the proportion in the upper group who answered the item correctly and pi is the proportion in the lower group who answered the item correctly (Schumacker and Lomax, 2010).

If $D \geq 0.40$ the item if functioning satisfactorily

If $0.30 \leq D \geq 0.39$, little or no revision is required

If $0.20 \leq D \geq 0.29$ the item is marginal and needs revision

If $D \leq 0.19$ the item should be eliminated or completely revised

For comparison purposes, JMetrik software (Meyer, 2009) was also used to calculate item-level difficulty and discrimination parameters. JMetrik is a free, open-source software application designed for classical and modern psychometric analyses. Difficulty and Discrimination coefficients for the initial APSCK items are listed in the following table. In some cases, there was a noticeable discrepancy between the hand-calculated and JMetrik coefficients. This was due to the handling of missing or incomplete data. Specifically, the hand-calculation method ignored missing data while the JMetrik software treated missing data as zero.

Table 14: Difficulty and Discrimination Coefficients (Items found in Appendix A)

APSCK Item	Diff*	Discri*	JMetrik Diff	JMetrik Discr
10	88	56	84	-0.8
11	85	45	4.4	3
12	80	64	57	26
13	42	39	41	41
14	95	-6	0	29
15	67	50	38	0
16	95	50	0	8
17	41	29	21	49
18	34	14	32	50
19	94	-4	0	12
20	95	13	41	60
21	40	36	14	14

*Difficulty and Discrimination Coefficients of Initial Content Knowledge Items

At this point, items were examined for a final time and those with high difficulty parameters, and/or zero or close to zero discrimination parameters were considered for deletion or revision. Ultimately, items were deleted if they had a negative discrimination parameter from either source. The following table details a summary of APSCK items that were kept, revised and/or deleted based on the item-analysis. Please note that item characteristics for final assessment items are detailed in Appendix D.

Table 15: Categories of Phase II APSCK Items (found in Appendix A)

	APSCK Measurable Characteristics		
A.P. Statistics Content Sub-Domains	Problem Solving Knowledge	Organizing Content	Student Learning
Sampling and Experimentation (S1)	10*	15*	
Anticipating Patterns (S2)	11*	20	
Statistical Inference (S3)	12	21*	17
Exploring Data (S4)	13	16	

*Revised Items (Items 14, 18 and 19 were deleted

APSPCK Item Analysis

Item-level analysis for the APSPCK began with two different raters using item specific rubrics to rate APSPCK individual responses. If there was a discrepancy, individual item ratings were reconciled between the two raters and a final grade was determined. Cohen's Kappa (Cohen, 1960) was calculated for each item to determine inter-rater reliability. These results are listed in Table 18 below:

Table 16: APSPCK Cohen's Kappa Coefficients

APSPCK Item Number*	N	Cohen's Kappa (K)
1	50	0.87
2	42	0.89
3	27	0.72
4	24	0.81
5	20	0.55
6	17	0.66
7	14	1.00
8	11	0.87
9	9	0.44

* Items found in Appendix A

Cohen's Kappa indicated that raters had high consistency with scoring. Those items with less than the established acceptable level of 0.70 Kappa coefficients were negotiated a second time among raters.

Also due to low response rates and missing data in Phase I, it was decided that the constructed-response APSPCK items be changed into multiple-choice items. APSPCK responses from Phase I were analyzed to produce a workable set of multiple-choice responses placed on a continuum of one to four corresponding to increasing levels of pedagogical content knowledge. This was accomplished for each item by using graded respondent constructed answers, from the initial sample, in combination with the item-specific rubrics.

From here, each APSPCK multiple-choice items there were four answer choices with each answer choice signifying a specific level of APSPCK. Individual responses given a "4" grade for example were combined and synthesized to make-up a level four answer choice. This level-four answer choice signified the highest level of APSPCK possible for that item. Those responses given a "3" grade were synthesized to produce a level-three answer choice and so on. Respondents were asked to choose the multiple-choice response that best represented their personal knowledge of teaching A.P. Statistics.

From this process, a total of five APSPCK items remained in the survey. Specifically, the last four APSPCK initial items were deleted from further analysis due to low response rate and missing data. The characteristic categories of APSPCK items one through five did not change from the initial to revised assessment. Table 17 outlines the

categories of APSPCK items after item-analysis. Please note that item characteristics for all final assessment items are also detailed in Appendix D.

Table 17: Categories of Phase II APSPCK Items (found in Appendices A and B)

	APSPCK Measurable Characteristics		
A.P. Statistics Content Sub-Domains	Adjust curriculum, instruction and assessment strategies	Recognize and interpret errors and misconceptions	Student thinking and learning
Exploring Data		5	
Sampling and Experimentation			4
Anticipating Patterns	1	2	
Statistical Inference	3		

*Items 6-9 were deleted due to low response

Phase II: Measurement and Structural Model Analyses

For Phase II, a second group of 1000 potential teacher respondents were invited to respond to the remaining assessment questions from Phase I. It was thought that shorter, more straightforward items would increase the number of completed cases. It was thought that approximately 300-400 complete surveys would be returned. Only 945 survey invitations reached the intended teacher due to incorrect or invalid email addresses listed on the College Board directory. For Phase II, response rates were approximately 25.60% and data collection resulted in 241 complete survey responses. Three reminder emails were sent to potential respondents to increase response rates, and one reminder email was sent to those who began the survey but did not finish.

The first step of Phase II data analysis was to conduct a CFA using SEM to confirm a measurement model using LISREL v8.80 student edition software. Specifically, the CFA of the measurement model for APSPCK and APSCK, and the structural model for overall APSTK were examined. As explained in Chapter 3, a number of methodological applications were utilized in an effort to evaluate the relationships between the assessment items and APSPCK and APSCK constructs as well as relationships theorized to explain the overarching construct of APSTK.

For this study, the Model Specification step for SEM has already been completed, because the study is using theory to define a hypothesized measurement model (See Chapter 3). Also, a maximum likelihood estimation technique was chosen for the Model Estimation step for this study because the multivariate normality assumption has been met. Consideration of data that was non-normal is outlined in the following discussion.

Measurement Model Results

For this study, the measurement model (presented in Fig. 3) was evaluated to understand the relationships between latent variables, APSPCK and APSCK, and the measured individual survey item responses. The goal of this analysis was to evaluate how well the measured model was being reproduced by the data.

For APSCK items, scored on a continuous scale from 0 to 100%, subscale-score 2 (S2) produced the highest average scores among respondents with 79.25%. Respondents scored on average 74.27% on S1 items, 66.60% on S4 items and 57.45% on S3 items. This means that respondents scored, on average, highest in the Anticipating Patterns category (S2) and lowest in the Statistical Inference category (S3).

Descriptive statistics for the measurement model show that for APSPCK items, respondents scored the highest on APSPCK items one, two and four. Out of a maximum score of 4 and a minimum score of 1, respondents scored on average 2.98, 3.17 and 3.25 respectively. The remaining two APSPCK items three and five had average scores higher than 2.50. This means that respondents scored highest (item four) in the area of Sampling and Experimentation and Student Learning. Respondents also scored well in the area of Anticipating Patterns and Recognizing and Interpreting Errors and Misconceptions (item two) as well as Anticipating Patterns and Adjusting Curriculum, Instruction and Assessment Strategies (item one). Descriptive statistics for the measurement model are outlined in Table 18.

Table 18: Measurement Model Descriptive Statistics

Item	Minimum	Maximum	Mean	Std. Deviation
APSCK				
S1	0.00	1.00	0.74	0.33
S2	0.00	1.00	0.79	0.28
S3	0.00	1.00	0.58	0.31
S4	0.00	1.00	0.66	0.28
APSPCK				
Q1	1.00	4.00	2.98	1.18
Q2	1.00	4.00	3.17	1.27
Q3	1.00	4.00	2.68	0.58
Q4	1.00	4.00	3.25	1.00
Q5	1.00	4.00	2.66	1.43

Valid N = 241 (listwise)

The measurement model produced several model fit statistics that are shown in Table 19 below.

Table 19: Measurement Model Fit Statistics

	Satorra-Bentler Chi-Square	df	RMSEA	NFI	NNFI	GFI	SRMR
Measurement Model	0.04	26	0.00	0.99	1.40	0.98	0.03
Ideal Levels			0.05-0.08	1	1	1	<0.05

Overall, the measurement model produced several fit indices with acceptable levels. The adjusted chi-square statistic of 0.04 is not statistically significant ($df = 26$, $p < 0.05$), so the model does fit the data. The RMSEA value of 0.00 is not an acceptable level; however, the NFI and GFI indices produce values close to one or perfect fit and the SRMR index of 0.03 is within adequate range. Therefore, it is concluded that the theoretical relationships are being accurately reproduced.

Because the measurement model fits the data well, individual parameter estimates were interpreted and examined for statistical significance. Overall results for the individual parameter estimates are summarized in Table 20 and evaluated below.

Table 20: Measurement Model Parameter Estimates

Item	Unstandardized Parameter Estimate	Standardized Parameter Estimate	t-value*	Error Variance	R^2
APSCK					
S1	0.21	0.21	1.77	0.06	0.42
S2	0.11	0.11	1.59	0.06	0.17
S3	0.08	0.08	1.21	0.09	0.08
S4	0.09	0.09	1.26	0.07	0.09
APSPCK					
Q1	0.17	0.17	0.23	1.35	0.02
Q2	0.32	0.32	0.11	1.50	0.07
Q3	0.30	0.30	0.26	0.24	0.27
Q4	0.27	0.27	0.42	0.93	0.07
Q5	-0.06	-0.06	-0.01	2.03	0.02

Statistically significant t-value ($p < 0.05$)

Item-level *t*-values were evaluated to determine which variables were statistically significant. The *t*-statistic critical value for df = 26 is 2.05 ($p < 0.05$), so none of the variables in the measurement model were statistically significant. This means that we fail to reject the null hypothesis, so the model is, in fact, a good fit to the data. Therefore, at the item level, it was concluded that the measurement model is a good representation of the data, because the items fit well within the model.

R^2 coefficients were also analyzed to determine which measurable items were most reliable. The R^2 values also show that latent variable APSCK explains 42.00% of the variance in subscale score one (S1), and about 17.00% of variance in subscale score 2 (S2). This also means that S1 items were the most reliable items and therefore the best indicators of APSCK among all the items. S1 items were designed to measure knowledge of Sampling and Experimentation, Problem Solving Knowledge and knowledge of Organizing Content. Therefore, items focused on these content areas were fairly good predictors of APSCK. Finally, S3 items (designed to measure Statistical

Inference) and S4 items (designed to measure Exploring Data) were neither strong nor reliable indicators of APSCK.

For the APSPCK items, the highest R^2 value came from APSPCK item number three with a value of 27.00%. This means that the latent variable APSPCK is explaining 27.00% of the variance in the score on the item. For the measurement model, this also means that APSPCK item number three was the most reliable item compared to the other APSPCK items. APSPCK item number three was designed to measure knowledge of Statistical Inference and knowledge of Adjusting Curriculum, Instruction and Assessment Strategies. In other words, this item was the best indicator of APSPCK out of all the items.

Item-level standardized parameter estimates show that for APSCK, S1 items had the highest, positive correlation with APSCK at 0.21, and S2 items had the next highest correlation with APSCK at 0.11. This means that, items measuring Knowledge of Sampling and Experimentation, Problem Solving Knowledge and Knowledge of Organizing Content were most correlated with APSCK.

It was also noted that respondents scored the highest on S1 and S2 items. So, S1 and S2 items were not only the best predictors of APSCK, but also the highest scoring items among respondents. This means that these items had the highest average scores among respondents and were also correlated and/or reliable predictors of APSCK. For further analysis, score value frequency distributions for S1 and S2 items can be found in Appendix E.

For APSPCK, items number two had the highest correlation at 0.33 and item number three had the second highest correlation at 0.30. This means that, the item

measuring Knowledge of Anticipating Patterns and Recognizing and Interpreting Errors and Misconceptions was most correlated with APSPCK. It should be noted that, item number three was almost identically correlated with APSPCK as item number two. Finally, it was noted that APSPCK item number five performed poorly when compared to other similar items. Item number five produced negative standardized and unstandardized parameter estimates as well a high level of error variance (compared to other items).

It was also noted that respondents scored the highest on APSPCK items one, two and four. So, even though items two and three were the best predictors of APSPCK, results show that respondents scored highest on items one, two and four. This means that even though items one and four had highest average scores among respondents, they were neither highly correlated nor reliable predictors of APSPCK. For further analysis, score value frequency distributions for APSPCK items two and three can be found in Appendix E.

Overall, the measurement model had acceptable fit, and several individual items had high reliability and were adequate predictors of either APSCK or APSPCK. It was concluded that the measurement model represented the data well and sufficiently supported the viability of a two-factor model.

Validity and Reliability of Latent Factors

Validity and reliability measures were also examined to establish the quality of the latent factors in the measurement model. Specifically, the variance extracted coefficient and Coefficient H construct reliability coefficient were calculated for each

latent variable (APSCK and APSPCK). In addition, the shared covariance between the two latent factors was evaluated.

For APSPCK, the Coefficient H construct reliability coefficient (Hancock & Mueller, 2001) equaled 0.35 which means the construct is fairly stable and reflected in the data by the chosen items designed to measure it. In addition, the variance-extracted coefficient for APSPCK was equal to 0.23. This means that APSPCK explained on average 23.20% of the variance in the items designed to measure it.

For APSCK, the Coefficient H construct reliability coefficient equaled 0.53, which means this construct is more stable than APSPCK and is better reflected in the data by the items designed to measure it. In addition, the variance-extracted coefficient for APSCK was equal to 0.41. This means that APSCK explained on average 41.00% of the variance in the items designed to measure it. This was higher than the variance extracted value for APSPCK.

Finally, the measurement model showed an 83.00% shared covariance between the latent factors APSPCK and APSCK. This was considered to be a high level of covariance and points towards a strong linear relationship between the two latent factors. The exact strength of the relationship between APSPCK and APSCK may need to be examined in the future (see Chapter 5). Even though the measurement model proved to have acceptable fit with two distinct factors, this amount of positive covariance suggests there may be a higher order factor with overlap between the two smaller constructs. Thus, there is a need to investigate a structural model and the possibility of the higher-order factor, APSTK.

Structural Model Results

After estimating a measurement model in CFA, the next step was to assess how well the higher-order model matched the observed data by evaluating a structural model. The structural model was also used to indicate how the latent variables APSCK and APSPCK were related to APSTK. Structural model fit indices as well as parameter estimates, *t*-values and R^2 values were analyzed to identify which variables (APSCK or APSPCK) had the most impact on overall APSTK.

The structural model metric was established by setting the path of S1 equal to one. In addition, a Heywood case (Schumacker and Lomax, 2010) appeared during this analysis when the error variance of APSPCK was found to be negative. This was corrected for model specification purposes by setting the error variance of APSPCK equal to zero. Structural model fit indices are summarized in Table 21.

Table 21: Structural Model Fit Indices

	Adjusted Chi-square	df	RMSEA	NFI	NNFI	GFI	SRMR
Structural Model	0.04	26	0.00	0.99	1.11	0.98	0.04
Ideal Level			0.05-0.08	1	≥1	1	<0.05

* Statistically significant ($p > 0.05$)

The structural model produced an adjusted chi-squared statistic of 0.04. This value is not statistically significant (df = 26, $p < 0.05$) against the null hypothesis, so the model fits the data. In other words, there was no difference between the patterns observed in the data and the model as it was specified. Based on the chi-squared statistic, it was concluded that the APSTK structural model represented the measured data well.

In addition, model fit indices were evaluated to determine if the structural model produced acceptable fit. Several model fit indices show acceptable levels of fit for the structural model. For example, the SRMR value at 0.04 is less than the suggested 0.05 and the NFI, NNFI and GFI indices are all close to one, which implies perfect model fit. Therefore, it was concluded that the structural model did an acceptable job of predicting total APSTK. In other words, the hypothesized model has been supported by the sample variance-covariance data. Please see the structural model path diagram in Appendix 3.

In addition to overall model fit, relationships among latent variables were examined to see if the latent variable APSTK explained any of the variance in APSPCK and APSCK. Latent variable parameter estimates, *t*-values and R^2 coefficients are summarized in Table 22.

Table 22: Structural Model Parameter Estimates

Construct	Unstandardized Parameter Estimate	Standardized Parameter Estimate	*t* value*	Error Variance	R^2
APSCK	0.10	0.83	0.26	0.00	0.69
APSPCK	0.32	0.56	0.11	1.00	0.31

* Statistically significant t-value ($p < 0.05$)

The unstandardized parameter estimate for APSCK was 0.10. This means that the latent variable APSCK had a 10.00% direct effect on, or correlation with APSTK. The unstandardized parameter estimate for APSPCK was 0.32. This means that the latent variable APSPCK had a 32.00% direct effect, or correlation with APSTK. The *t*-values for both APSCK and APSPCK were not statistically significant ($p < 0.05$). This means that there was not sufficient evidence against the null hypothesis, so the model is indeed a good representation of the data.

In addition to path estimates and t-values, R^2 values were evaluated for the structural model. The R^2 value for APSCK was approximately 69.00%. This indicates that APSTK explained 69.00% of the variance in APSCK. The R^2 value for APSPCK was approximately 31.00%. This means that the latent variable APSPCK had 31.00% of its' variance explained by APSTK. Therefore, it was concluded that APSPCK was a stronger predictor of APSTK, but APSCK was a more reliable predictor of APSTK.

Overall, the structural model produced a non-significant chi-squared statistic and several model fit indices that were in acceptable range. It was concluded that the model as a whole did an adequate job of predicting APSTK. Relationships among latent variables were also examined using parameter estimates, *t*-values and R^2 values. According to the parameter estimates, APSPCK had more of a direct effect on APSTK than APSCK. Both APSCK and APSPCK had non-significant *t*-values and the R^2 values point towards a relationship between APSCK, APSPCK and overall APSTK. According to the R^2 coefficient, APSCK is a more reliable predictor of APSTK than APSPCK.

Calculating Latent Variable Scores

This study calculated individual latent variable scores (LVS) for APSPCK items and APSCK items as well as overall LVS for each respondent. Unlike factor scores, which contain systematic error, latent variables scores resulting from a structural model are unbiased estimates, and therefore are error-free estimates. Post-modeling with latent variable scores generated from the SEM model provides more accurate values with which to examine the effect of exogenous variables, such as APSTK assessment items. For a point of reference, the mean LVS of -0.00 equated to a raw score of 72.00%. This was the

average score for all participants. The minimum raw score was 24.00% and the maximum raw score was 98.00%. Table 23 shows the latent variable score descriptive statistics.

Table 23: Descriptive Statistics for Latent Variable Scores

	N	Min	Max	Mean	Std. Deviation
Latent Variable Score	241	-7.32	2.83	-0.00	1.92
Valid N (listwise)	241				

Phase III: Exploring Teacher Characteristic Variables

In this phase, individual teacher's LVS were compared to one another in order to establish a range, average and level of variability in APSTK among those who responded to the survey. A multiple regression analysis was conducted to evaluate how well predictor variables such as "Number of Years Teaching A.P. Statistics", "Number of College-level Statistics Courses Completed", and "Having a Degree in Statistics or Not" predict the criterion variable of APSTK LVS.

To begin, the sample multiple correlation coefficient was analyzed to determine the relative strength of individual predictors. This coefficient also indicated how much variance in the dependent variable (APSTK LVS) each independent variable explains or predicts. In addition, the F critical value and F test statistic were examined to determine if individual predictor variables had a statistically significant and/or practically significant impact on the APSTK score.

For this analysis, the dependent variable consisted of individual respondent LVS on the APSTK assessment. The independent variables or predictors of LVS are "Having

a degree in Statistics", "Number of college-level Statistics Classes Completed" (assuming one class is worth 3 college credits), "Number of years Teaching A.P. Statistics" and a self-reported measure of "Student Success on the AP exam". Descriptive statistics for the teacher characteristic variables are outlined in Table 24.

Table 24: Descriptive Statistics for Teacher Characteristic Variables

	N	Range	Min	Max	Mean	Std. Deviation
Degree	240	1	0	1	0.08	0.28
Classes	240	19	0	19	2.35	2.59
Years	240	15	1	16	5.90	3.95
Success	240	100	0	100	62.70	25.00
LVS	240	6.31	-4.62	1.69	-0.01	1.18
Valid N (listwise)	240					

From a sample of 240 respondents, the average LVS was -0.01% with the highest score being 1.69 and the lowest being -4.62. Those respondents who have a degree in Statistics were coded "1" and those who do not were coded "0". Of the 240 total respondents, seventeen respondents (or .07 percent) have a degree in Statistics while 223 do not. The average number of college-level Statistics classes taken by respondents was 2.35 with a maximum of 19 classes and a minimum of none. The average number of years teaching AP Statistics among respondents is 5.90 or roughly 6 years. Finally, the average Percentage of Students Passing the AP Exam reported by respondents is 62.70 percent.

Results of Multiple Regression Analysis

A Multiple Regression analysis was conducted to evaluate how well the predictor teacher characteristic variables "Having a degree in Statistics", "Number of college-level

Statistics courses completed", "Number of years teaching A.P. Statistics" and "Student success on the A.P. exam" predicted the criterion (or dependent) variable LVS.

The alternative hypothesis states that at least one variable belongs in the model. In addition, the following regression equation was obtained:

Y = 64.08 + -6.18 (degree) + 0.21 (classes) + 0.14 (years) + 0.11 (success)

Regression output also produced the following correlation matrix of the data.

Table 25: Correlation Matrix of Characteristic Variables

Pearson Correlation	LVS	Degree	Classes	Years	Success
LVS	1.00				
Degree	-0.10	1.00			
Classes	-0.05	0.73	1.00		
Years	0.12	0.11	0.21	1.00	
Success	0.25	0.08	0.09	0.29	1.00

The correlation matrix shows that the variable most positively correlated with APSTK survey score was "Success" or "Student success on the AP Statistics exam" at 25.10%. In fact, variables "Having a Degree in Statistics, and "Number of College-Level Statistics Classes" completed, have a negative correlation to overall APSTK survey score. The table shows that the highest predictive relationship among variables was between "Classes" and "Degree" at 73.00%. There was also a 28.70% correlation between the variables "Years teaching A.P. Statistics", "Years", and "Success".

Table 26 shows the linear combination of teacher characteristic variables was significantly related to the total survey score, $F(4, 235) = 5.29$, $p = .005$. The F critical value for this model was 2.41 ($p < .05$), so the null hypothesis was rejected and the overall regression equation was a significant predictor of APSTK survey score. In addition, the *p*-value associated with the F statistic was also significant therefore

confirming the fact that at least one predictor variable was statistically significant in predicting the APSTK LVS.

Table 26: ANOVA Table

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	27.61	4	6.90	5.29	.005
Residual	306.07	235	1.30		
Total	333.68	239			

The sample multiple correlation coefficient for the model was 0.29, indicating that approximately 6.70% of the variance in the LVS in the sample can be accounted for by the linear combination of teacher characteristic variables. This result is shown in Table 27.

Table 27: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.29	0.07	0.07	1.14

Table 28 shows the indices indicating the relative strength of the individual predictors in predicting total APSTK LVS. If the *p*-value for the variable was less than 0.05, the null hypothesis was rejected, so the predictor did not significantly contribute to the total APSTK LVS. Of all the predictors, only "Percentage of Student Success on the A.P. Exam" resulted in being a significant predictor above and beyond the other three predictors of APSTK survey score with $t(240) = 3.71$ ($p < 0.05$).

Table 28: Regression Parameter Estimates

Model	Understandized Coefficients		Standardized Coefficients		
	β	Std. Error	β	t*	*p*-value
Constant	-0.81	0.22		-3.75	0.00
Degree	-0.61	0.39	-0.14	-1.55	0.12
Classes	0.01	0.04	0.02	0.19	0.85
Years	0.02	0.02	0.07	0.99	0.33
Success	0.01	0.00	0.24	3.71*	0.00

* Statistically Significant ($p < 0.05$)

In fact, "Percentage of Student Success on the AP Exam" explained 4.20% (see Table 29) of the variance in APSTK survey score holding the variables "Having a Degree in Statistics", "Number of Statistics Courses Taken" and "Number of Years Teaching A.P. Statistics" constant. Also, the "Percentage of Student Success on the AP Exam" predictor explains 4.10% of the variance in APSTK survey score given that the other three independent variables were included in the model.

"Having a degree in Statistics" was very close to being a significant predictor of APSTK LVS above and beyond the other predictors $t(240) = -1.55$, $p < 0.05$. "Having a degree in Statistics" explained only 0.97% of the variance in APSTK survey score holding the other three predictors constant. Also, "Having a degree in Statistics" explained only 1.10% of the variance in APSTK survey score given that the other three predictors were included in the model.

The "Number of College-level Statistics Courses Completed" variable was not a significant predictor of APSTK survey score above and beyond the other three predictors, $t(240) = 0.19$, $p < 0.05$. In fact, "Number of College-level Statistics Courses Completed" only explained 0.01% of the variance in APSTK survey score holding the other three predictors constant and 0.01% with other three predictors included in the model.

The "Number of Years Teaching AP Statistics" variable was not a significant predictor of APSTK survey score above and beyond the other three predictors, $t(240) = 0.99$, $p < 0.05$. "Number of Years Teaching AP Statistics" only explained 0.40% of the variance in APSTK survey score holding the other three predictors constant, and 0.40% including the other predictors in the model.

Table 29: Regression Coefficients

Model	Zero-Order	Partial	Semi-Partial
(Constant)			
Degree	-0.10	-0.10	-0.09
Classes	-0.05	0.01	0.01
Years	0.12	0.06	0.06
Success	0.25	0.24	0.23

Overall, the linear combination of teacher characteristic variables was a significant predictor of APSTK LVS. In addition, the *p*-value associated with the F statistic confirmed the fact that at least one predictor variable was statistically significant in predicting the APSTK LVS. Furthermore, the independent variables validated 6.70% explanation of variance in the dependent variable with the all predictors included in the model.

Individually, the "Percentage of Student Success on the AP Exam" variable was statistically significant, but only contributed 4.20% to the level of variance in the total APSTK LVS. It was thought that new, or more specific predictors be added to improve the overall quality of the model in predicting APSTK survey score, because of the low level of contribution from any of the current predictors. This will be discussed further in Chapter 5.

Brief Summary of Results

This study resulted in several significant findings pertaining to item development and analysis, measurement and structural model fit and indicators of APSTK LVS. To begin, items designed to measure APSCK and APSPCK fit well within a measurement model, and a measurement model produced good fit. In addition, a structural model produced good fit, and evidence pointed towards APSPCK having a stronger overall effect on APSTK; however, APSCK was a more reliable predictor of APSTK. Finally, the linear combination of teacher characteristic variables was a significant predictor of APSTK LVS. A more detailed discussion of results can be found in Chapter 5.

Chapter 5

Implications and Recommendations

Re-Statement of the Problem

Increasing student enrollment in high-school level Advanced Placement (AP) Statistics courses necessitates the need for teachers who are knowledgeable in the subject-area. However, recent research has shown that we do not know much about statistics teachers (Mills, 2007), particularly at the secondary level. In fact, the majority of recent research on statistics teaching and learning calls for more research into the pedagogic knowledge necessary for teaching secondary-level statistics (Casey, 2010; Groth & Bergner, 2005; Franklin, 2011; Hassad, 2006; Leavy 2010; Liu & Thompson, 2007; Phannkuch, 2006; Rossman, 2006). Though a need exists, there is no established benchmark for the amount or types of content-specific teaching knowledge that is required of AP Statistics teachers.

This study is based upon the idea that a teacher may possess enough teaching knowledge for mathematics but be deficient in the subject-specific teaching knowledge necessary to teach AP Statistics. Therefore, this study aimed to establish a measure of AP Statistics teaching knowledge by creating an A.P. Statistics Teaching Knowledge (APSTK) assessment. This assessment was developed to investigate the content and pedagogical knowledge of secondary-level, in-service AP Statistics teachers.

Overview of Findings

The methodological strategies for this study consisted of three primary phases designed to answer key research questions. Phase I was entitled Item Development and

Analyses, Phase II was entitled Measurement and Structural Model Analyses and Phase II was entitled Exploring Teacher Characteristic Variables.

As a brief overview, Phase I consisted of an item-level analysis using scores obtained from the initial data collection. Completion of this phase resulted in a modified assessment with fourteen multiple-choice items. Phase II consisted of completing a CFA to analyze the results of a measurement model and structural model using SEM. Phase III consisted of a multiple regression analysis to determine which teacher characteristic indicator variables predicted overall APSTK LVS. A detailed overview of findings from all three phases is outlined below.

Phase I Results

Phase I data were analyzed to establish measurable subscales of APSCK and item level scores of APSPCK items. Based on difficulty and discrimination coefficients, four APSCK items were revised and three items were deleted from the original survey. Due to low response rates, it was decided that the constructed-response APSPCK items be changed into multiple-choice items. APSPCK responses were analyzed to produce a continuum of one to four corresponding to increasing levels of pedagogical content knowledge. At the end of the analysis, there were nine remaining APSCK items and five remaining APSPCK items.

Phase II Results

Measurement model results.

In order to evaluate the measurement model, a chi-squared statistic and model fit indices were examined as well as descriptive statistics, item-level parameter estimates, *t*-values and R^2 values. Based on a variety of measurement model fit indices, it was

concluded that theoretical relationships were being accurately reproduced with the measurement model. The model was a good fit to the data. The adjusted chi-square statistic of 0.04 is not statistically significant ($df = 26$, $p < 0.05$), thus signifying good model fit.

Because the measurement model fits the data well, individual parameter estimates, *t*-values and R^2 values were interpreted and examined for statistical significance. None of the measured variables in the measurement model had a statistically significant *t*-value. Therefore, the *t*-values suggest that all the items fit well within the model.

Item-level standardized parameter estimates show that for APSCK, S1 items had the highest, positive correlation with APSCK at 0.21, and S2 items had the next highest correlation with APSCK at 0.11. This means that, items measuring Knowledge of Sampling and Experimentation, Problem Solving Knowledge and Knowledge of Organizing Content were most correlated with APSCK.

For APSPCK, item number two had the highest correlation at 0.33 and item number three had the second highest correlation at 0.30. This means that, the item measuring Knowledge of Anticipating Patterns and Recognizing and Interpreting Errors and Misconceptions was most correlated with APSPCK. It should be noted that, item number three was almost identically correlated with APSPCK as item number two.

Finally, R^2 values were evaluated to determine the amount of variance in the measured items explained by the latent variables APSCK and APSPCK. Specifically, APSCK explained the highest level of variance among items (42.00%) for APSCK subscale score one (S1). Items in S1 were designed to measure knowledge of Sampling

and Experimentation, Problem Solving Knowledge and knowledge of Organizing Content. This means that APSCK subscale score one (S1) items were the most significant indicator of APSCK out of the items designed to measure it, and contained the most reliable items.

Comparatively, S1 (Sampling and Experimentation, Problem-Solving Knowledge and Organizing Content) items were the most reliable, but respondents actually performed better on S2 items designed to measure knowledge of Anticipating Patterns, Problem-Solving and Organizing Content. This tells us that S1 items were well written and more effective at measuring APSCK, but were slightly more difficult for respondents to answer. Also, respondents were most successful at answering S2 items, but those items contributed less to APSCK and were less reliable in the measurement model.

Finally, S3 items (Statistical Inference, Problem-Solving Knowledge, Organizing Content and Student Learning) were not as significant or reliable as others items, and were also the most difficult for respondents to answer. It is concluded that these items may need revision, but also point towards a potentially weak area in AP Statistics content knowledge. For more insight into overall APSTK assessment scores compared to APSCK item-level scores, please see Appendix F.

In addition, APSPCK explained the highest level of variance among items (27.00%) for APSPCK item number three which was designed to measure knowledge of Statistical Inference and Knowledge of Adjusting Curriculum, Instruction and Assessment Strategies. In other words, this item was the most significant indicator of APSPCK out of all the items designed to measure it, and was the most reliable APSPCK item. Finally, it was noted that APSPCK item number five performed poorly when

compared to other similar items. Item number five produced negative standardized and unstandardized parameter estimates as well a high level of error variance (compared to other items).

Comparatively, APSPCK item number three was the most reliable and significant predictor of APSPCK, but respondents performed better on items one, two and four. This tells us that item number three (designed to measure knowledge of Statistical Inference, Adjusting Curriculum, Instruction and Assessment Strategies) was written well but one of the most difficult items for respondents to answer. Respondents were most successful at answering APSPCK item number four (designed to measure Sampling/Experimentation and Student Thinking), but this item did not have a significant impact on overall APSPCK. Finally, APSPCK item number five (designed to measure knowledge of Recognizing Errors and Misconceptions and Exploring Data) performed poorly in the measurement model, and was also the most difficult for respondents to answer. This item will need to be examined and possibly revised in the future, but may point towards a potentially weak area in AP Statistics pedagogical content knowledge. For more insight into overall APSTK assessment scores compared to APSPCK item-level scores, please see Appendix F.

Finally, validity and reliability measures were examined to establish the quality of the latent factors (APSPCK and APSCK) in the measurement model. Based on the variance extracted coefficient and Coefficient H values, the construct APSCK was slightly more stable and more reliable than APSPCK, and therefore better reflected by the items designed to measure it.

Structural model results.

In order to evaluate the structural model and relationships among latent variables, a chi-squared statistic and model fit indices were examined as well as item-level parameter estimates, *t*-values and R^2 values. The structural model produced a non-significant chi-squared value (df = 26, $p < 0.05$), thus signifying a model that was well represented by the observed data. In addition, the structural model produced acceptable fit indices, so it was also concluded that the hypothesized model did an acceptable job of predicting APSTK.

Parameter estimates indicated that the latent variable APSPCK had a 32.00% direct effect on, or correlation with APSTK while APSCK only had a 10.00% direct effect on, or correlation with APSTK. The *t*-values for both APSCK and APSPCK were non-significant ($p < 0.05$), suggesting that the model was a good representation of the data.

The R^2 value for APSPCK was approximately 31.00% while the R^2 value for APSCK was comparably larger at approximately 69.00%. In other words, APSCK explained 69.00% of variance in APSTK. Thus signifying that APSCK items were considerably more reliable at predicting APSTK than APSPCK items. Based on these mixed results, it was concluded that APSPCK was a better predictor of APSTK, while APSCK was the more reliable construct.

Finally, LVSs were calculated for each respondent. The minimum LVS was -7.319 and the maximum was 2.83 producing a range of -4.49. The average LVS was -0.000 equating to a raw score of 72.50%. These APSTK scores were used as the dependent variable in the following multiple regression analysis.

Phase III Results

Phase III was the completion of a multiple regression analysis to determine which teacher characteristic variables predicted overall APSTK LVS. To start, the linear combination of teacher characteristic variables was a significant predictor of APSTK LVS, and the *p*-value associated with the F-statistic confirmed that at least one predictor variable was statistically significant in predicting the APSTK LVS. Overall, the multiple regression analysis validated only a 6.70% explanation of variance in the dependent variable with the all predictors included in the model.

Individually, the "Student Success on the A.P. Exam" variable was statistically significant, but only contributed 4.20% to the level of variance in the total APSTK LVS. Therefore, not providing very much practical significance for predicting total APSTK survey score. It should also be noted that the variable, "Having a Degree in Statistics" was extremely close to being a significant predictor of LVS score. This variable needs more investigation in the future.

Implications of Findings

This study sought to fill a gap in the research regarding secondary-level, in-service Statistics teaching knowledge. The findings of this study suggest several implications for the current APSTK model, as well statistics education in general, statistics teacher preparation and professional development, student learning and licensure requirements.

To begin, items developed for the APSTK assessment point towards specific sub-components of content and pedagogical content knowledge to teach AP Statistics. Furthermore, CFA findings provided evidence for a two-factor model of teaching

knowledge for AP Statistics, as well as evidence of a higher-order factor called APSTK that encompasses some combination of APSCK and APSPCK. Finally, results of the multiple regression analyses suggest a correlation between student achievement on the A.P. Statistics exam and overall LVS on the APSTK assessment.

Considering the results of the CFA, it seemed that APSCK was a more stable and reliable construct than APSPCK. However, APSPCK was a stronger predictor of APSTK than APSCK. In addition, there was significant overlap or correlation between APSCK and APSPCK (83.00%). This suggests that content knowledge alone may not be enough to define APSTK as overall teaching knowledge for the subject, and pedagogical content knowledge also plays an important role.

Furthermore, items designed to measure APSCK were proven to be most reliable. These items may be useful in future efforts to measure content knowledge for Statistics. In addition, several APSPCK items were found to be significant indicators of that construct as well. Specifically, APSCK S1 and S2 items and APSPCK item numbers one, two and four adequately measured what they were designed to measured. This supported the idea that many items developed for the assessment were well-written and reliable indicators of APSCK and APSPCK.

Impact on Statistics Education

Past research efforts in Statistics education have uncovered the idea that statistics is a distinct, separate discipline from Mathematics. Specifically, Cobb and Moore (2007) found that statistics requires a different kind of thinking from teachers and students. Therefore, this study aimed to provide compelling evidence that AP Statistics requires teachers to possess a unique and distinct type of teaching knowledge. Developing and

validating the APSTK assessment accomplished this by working to differentiate those teachers who possess measurable characteristics of APSTK from those who do not.

Furthermore, this study intended to expand the current research base that exists for secondary-level statistics teaching knowledge. Of particular importance, Groth's (2007, 2013) framework for Statistical Knowledge for Teaching (SKT) was acknowledged. In comparison, both the APSTK and SKT frameworks used the previously mentioned GAISE framework (Franklin et al., 2007) as a foundational structure from which to investigate Statistics teaching knowledge. In fact, both the APSTK and SKT frameworks outlined several of the same content knowledge components for teaching Statistics such as Sampling and Experimentation (Collecting Data) and Exploring Data (Analyzing Data) for example.

In contrast, the APSTK framework did not utilize the Learning Mathematics for Teaching (LMT) project (Hill & Ball, 2004) to frame the APSTK construct such as Groth (2007) suggests. Furthermore, the current study did not differentiate between Common Knowledge and Specialized Knowledge as was done within the SKT framework. These differences and similarities (as well as others not mentioned) were attributed to the task of characterizing teaching knowledge specific to AP Statistics and not meant to dispute previous and on-going research in the field.

Impact on Statistics Teacher Preparation

The results of study further expose the need for more teaching methods courses specific to secondary-level statistics content knowledge and pedagogy. According to Rossman (2006), a distinct type of instructional preparation is needed for statistics teachers because students react differently to statistics than to Mathematics. Also, we

know that most mathematics teachers are not comfortable teaching statistics (Franklin, 2010). This matter needs to be addressed in teacher preparation programs. It may not be enough to cover a small unit of statistics content and/or pedagogy in the conventional mathematics teaching methods course. The APSTK assessment provided evidence in support of knowledge constructs specific to AP Statistics above and beyond Mathematics in general. So, there may be a need to create more, or expand upon, teacher preparation courses specific to secondary-level statistics.

Specifically, respondents did not perform well on content knowledge items measuring College Board defined content areas of Exploring Data and Statistical Inference. Furthermore, respondents did not perform well on pedagogical content knowledge items designed to measure the same content areas. Therefore, Mathematics teaching methods courses (or statistics methods courses) need to provide more focus on these specific content areas. In addition, methods courses need to provide more instruction in the pedagogical areas of Adjusting Curriculum, Instruction and Assessment Strategies and Recognizing Errors and Misconceptions. Overall, respondents had difficulty answering assessment items focused in these areas.

Finally, there are minimal opportunities for in-service statistics teachers to gain content and pedagogical content knowledge through professional development opportunities. Oftentimes, a teacher who is licensed in secondary mathematics is asked to teach AP Statistics, but may not be knowledgeable in the subject-area. In-service teachers need more opportunities to receive training and guidance on instructional strategies and curriculum options that support student learning. This study addresses this

effort by beginning to define and detail exactly what content knowledge and pedagogical content knowledge look like for secondary-level statistics teaching.

Impact on Mathematics Licensure Requirements

By outlining a set of measurable characteristics and creating knowledge constructs for secondary-level statistics teaching knowledge, this study also impacts pre-service statistics teacher licensure exams and requirements. Information obtained from this study, could improve testing agencies' ability to better assess pre-service Statistics teachers' knowledge of content and pedagogy. Mathematics licensure exams might contain better, enhanced items specifically designed to measure knowledge for teaching statistics. This study has confirmed that while specific content knowledge for teaching statistics is needed (and potentially deficient in some areas), pedagogical content knowledge also plays a significant role in overall APSTK. Mathematics licensure exams should reflect this by more accurately targeting and measuring both knowledge types.

Impact on Student Learning of Statistics

The idea of student learning was included in the conceptualization of APSCK and APSPCK as representing a possible link between knowledge of content and knowledge of pedagogy. In fact, Statistics teaching knowledge is closely tied to student learning, and research in the field has called for more focus on teaching knowledge as it relates to student learning (Ball, et al., 2008; Fennema and Frank, 1992).

Now, the focus has progressed to include understanding how teachers use their mathematical knowledge to make informed pedagogical decisions affecting student learning in the context of a particular class, such as AP Statistics. For example, respondents to the APSTK assessment did not perform well on items (S4) measuring

knowledge of Statistical Inference and Student Learning. This confirms the need for more research to determine how teachers' use of content knowledge affects student learning, and their own decision making with regard to curriculum, instruction and assessment choices. Development of the APSTK assessment helped to not only understand these relationships, but also measure them.

Recommendations for The College Board

Because this study focused specifically on AP Statistics teaching knowledge, there are also several key recommendations for the College Board. Specifically, AP Statistics teachers need more training and/or educational opportunities in the content areas of Exploring Data and Statistical Inference. From a pedagogical perspective, AP Statistics teachers need more training and exposure to strategies that will help them recognize errors and misconceptions, adjust curriculum, instruction and assessment strategies as well as understand student learning.

Also, it would benefit the College Board to ascertain a representative profile of what a current AP Statistics teacher with an average-level of APSTK looks like in terms of characteristic data. Table 30 details the average characteristics of those who scored the mean APSTK score of 72.50% on the assessment.

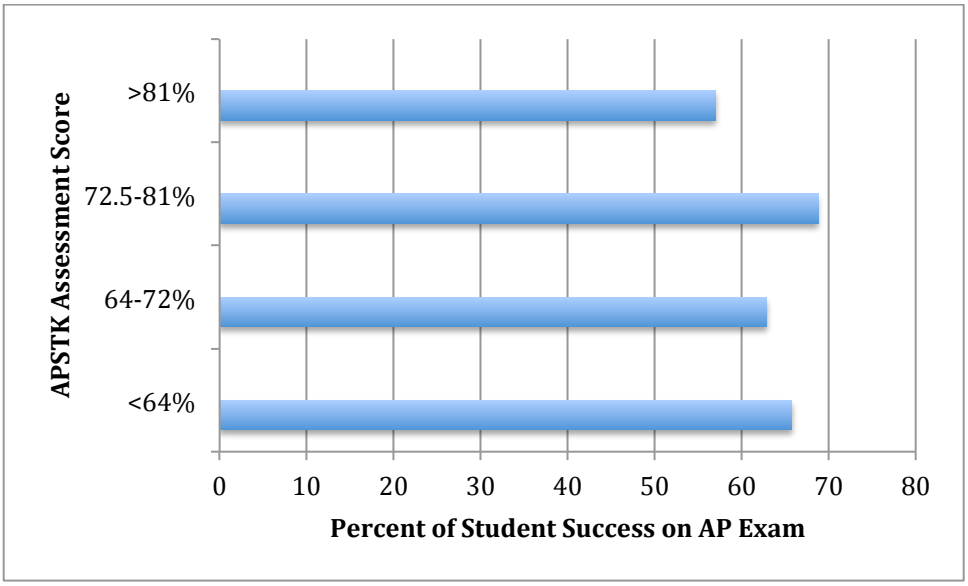
Table 30: Profile Characteristics of Respondent with Average Score on Assessment

APSTK Assessment Average Score	Degree	3-Credit Classes Completed	Years Teaching AP Statistics	Percent Student Success on AP Exam
72.50%	0.05% or 1 person	1.78	6.05	66.16

These respondents reported that they, for the most part, did not have a degree in statistics, and have only completed one to two (3-credit) college-level courses in statistics. The number of those with a degree in statistics was fairly consistent with the overall sample, 0.05 compared to 0.07, respectively. The College Board may benefit from knowing how many current AP Statistics teachers do not have a degree in the subject. Specifically, current teachers have often completed little more than one college-level course in statistics. There is cause for concern if teachers potentially do not understand or have not been exposed to content beyond the basics. AP Statistics teachers are expected to not only understand the content, but teach it as well, and this requires content knowledge beyond what she expects her students to know.

The College Board may also want to consider how a self-reported measure of student success on the AP Statistics exam compares to overall APSTK assessment score. Figure 5 was provided to help visualize the relationship between the percent of students who scored a 3 or higher on the AP Exam and APSTK assessment score.

Figure 5: APSTK Assessment Score Compared to Student Success on AP Exam



It was interesting to note that those who scored below average on the assessment reported a higher student success rate than those who scored above average. In fact, those who scored highest on the assessment reported only a 55.00% student success rate on the AP Statistics exam, while those scoring on the lowest quartile reported nearly 65.00% student success. Because this was a self-reported measure, this result could be due to several factors influenced by teacher perspective or student characteristics. For example, lower performing respondents might have more knowledge of strategies that achieve higher rates of student success on the AP exam, but are lacking in related content and pedagogical content knowledge when compared to their peers.

Alternatively, higher performing respondents may have a better understanding of the content and pedagogical content knowledge necessary to teach the subject, but struggle with providing effective strategies that achieve higher rates of student success on the AP exam. In this case, discrepancies might have existed because respondents had difficulty differentiating between knowledge of grading procedures and/or strategies for passing the exam and teaching knowledge required to successfully teach the course. There may also exist other significant relationships for the College Board to consider between APSTK assessment score and teacher characteristic variables, as well as further item development in specific areas. Some of these recommendations are highlighted and discussed below.

Recommendations for Future Research

Phase I Future Research

Each phase of this study indicated potentially more evaluation and research, some of which were previously noted in the Implications of Findings sections. Phase I

consisted of item development and analysis. It is recommended that future research include more item development of both APSCK and APSPCK items. After the initial item-level analysis, the APSTK assessment was left with nine APSCK items and five APSPCK items. There were two APSCK items left in S1, S2 and S4 and three items left in S3.

Also, there were two APSPCK items left per sub-component (Adjusting Curriculum, Instruction and Assessment Strategies and Identifying Error and Misconceptions). However, there was only one item for the sub-component entitled Student Thinking and Learning. This is an area that needs to be addressed in future research. There is a need for more items to measure this component, because recent research calls for increased emphasis on student learning of statistics.

Finally, an APSTK conceptual framework based on two primary constructs APSCK and APSPCK was created in Phase I. A two-factor model of APSTK was verified because both constructs were proven to be important contributors to overall APSTK. Furthermore, examination of the structural model showed that these two constructs explained a significant amount of variance in overall APSTK scores. So, it was concluded that the conceptual framework described a complete picture of teaching knowledge for AP Statistics. It was recommended that future research focus on exploring different relationships that might exist between APSCK and APSPCK. This is discussed in more detail below.

Phase II Future Research

Results of the measurement model analysis concluded that the model adequately fit the data, and several assessment items were found to be significant indicators of

APSCK and APSPCK. In addition, considerable overlap was discovered between APSCK and APSPCK. The measurement model analysis uncovered a high, positive correlation (83%) between the two constructs. This relationship needs to be examined further in future research. Perhaps seeking answers to questions such as, to what extent does APSCK affect APSPCK or vice versa? How much content knowledge is contained within the APSPCK construct?

Moreover, the structural model produced a non-significant chi-squared value and acceptable model fit indices. However, mixed results showed that APSCK was a more reliable construct than APSPCK, but APSPCK was a stronger predictor of APSTK. It was thought that different model identification and modification strategies may uncover better information about the relationships among latent variables in future research. In addition, there may be a need to further investigate high error variance cases and negative parameter estimates to determine the cause or source. Schumacker and Lomax (2010) point out that parameter estimates are sensitive to model identification. So, cases like this may call for better model identification and modification strategies, or suggest a larger flaw in the measurement or structural model that needs to be examined.

Phase III Future Research

Teacher characteristic variables were analyzed in Phase III to gather descriptive information about current AP Statistics teachers, and examine relationships among these characteristics and APSTK assessment scores. The analysis showed that the variable "Student Success on the AP Statistics Exam" was the only statistically significant predictor of APSTK LVS. However, the variable did not have much practical significance and was a self-reported measure. Future endeavors should include a closer

look at how student success and learning in AP Statistics is related to teacher knowledge. Development of the APSTK assessment helped to provide a measurable starting point for this analysis.

Finally, the other teacher characteristic variables included in this study need to be explored in more detail. It was noted that the variable "Having a Degree in Statistics" performed only slightly below the "Student Success on AP Statistics Exam" variable in predicting APSTK score, and was deemed not statistically significant by a narrow margin. This variable, in particular, should be examined in future research. Finally, there may be other teacher characteristic variables that better predict APSTK assessment score such as knowledge of AP grading procedures or number of professional development hours completed specific to AP Statistics.

References

Anderson, R. D., & Rubin, H. (1956). Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability, 5, 111-150.

Ball, D. L. (1988). Research on Teaching Mathematics: Making Subject Matter Knowledge Part of the Equation. In J. Brophey (Ed.) *Advances in Research on Teaching: Vol 2. Teacher's subject matter knowledge and classroom instruction* (pp 6-67). Greenwich, CT: JAI Press.

Ball, D.L. & Bass, H. (2003). Towards a Practice-based Theory of Mathematical Knowledge for Teaching. In B. Davis & E. Simmt (Eds.), Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group, (pp. 3-14). Edmonton, AB: CMESG/GCEDM.

Ball, D.L., Hoover, M., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education, 59(1)*, 389-407.

Bessant, K.C., & MacPherson, E. D. (2002). Thoughts on the Origins, Concepts, and Pedagogy of Statistics as a 'Separate Discipline'. *The American Statistician*, *50(1)*, 22-29.

Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Callingham and Watson (2011) found in C. Batanero, G. Burrill, and C. Reading (eds.), Teaching Statistics in School 283 Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study, DOI 10.1007/978-94-007-1131-0_28, © Springer pg 283-293.

Casey, S.A. (2008). *Subject Matter Knowledge for Teaching Statistical Association* (Doctoral Dissertation). Retrieved from: http://search.proquest.com.proxygw.wrlc.org/pqdtft/advanced?ip?accountid=11243

Casey, S.A. (2010). Subject Matter Knowledge For Teaching Statistical Association. *Statistics Education Research Journal*, *9(2)*, 50-68.

Chance, B., & Rossman, A. (2006). Investigating statistical concepts, applications, and methods. Belmont, CA: Duxbury.

Cobb, G.W., & Moore, D.S. (1997). Mathematics, Statistics and Teaching. *The American Mathematical Monthly, 104(9),* 801-823.

Common Core State Standards Initiative. (2010). Common Core Mathematics Standards. *National Governor's Association*. Retrieved from http://www.corestandards.org/

Cohen, J., (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 1,* 37-46.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd.ed). Lawrence Eribaum Associates, Inc Mahwah, NJ

Common Core State Standards Initiative. (2010). Common Core Mathematics Standards. *National Governor's Association*. Retrieved from http://www.corestandards.org/

Cook, C., Heath, F., & Thompson, R.L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement, 60(6),* 821-836.

Crocker, L. & Algina, J. (2008). *Introduction to Classical and Modern Test Theory.* Mason, OH: Cengage Learning.

Cummings, M.B., (2010). *Rethinking Teacher Preparation: Conceptualizing Skills and Knowledge of Novice Teachers of Secondary Mathematics.* University of Utah

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates incovariance structure analysis: Comments on the causes of and solutions to Heywood cases. Psychological Bulletin, 101(1), 126-135.

DeBoer, G.E., Abell, C.A., & Gogos, A. (2007). *Assessment Linked to Science Learning Goals: Probing Student Thinking During Item Development.* Paper presented at the AAAS Project 2061 Symposium on Assessment National Association for Research on Science Teaching Annual Conference, New Orleans.

Education Testing Service. (2012) State Requirements. *The Praxis Series*. Retrieved from http://www.ets.org/praxis/states

Education Testing Service. (2012) Mathematics Content Knowledge Test at a Glance. *The Praxis Series*. Retrieved from http://www.ets.org/Media/Tests/PRAXIS/pdf/0061.pdf

Ferrini-Mundy, J., & Senk, S., (2006, April). *Knowledge of Algebra for Teaching: Framework, Item Development, and Pilot Results*. Paper presented at the NCTM Research Symposium, St. Louis.

Franklin, C., Kader, G., Mewborn, D.S., Moreno, J., Peck, R., Perry M., & Schaeffer, R. (2005). *A Curriculum Framework for PreK-12 Statistics Education*. Retrieved from American Statistical Association website: http://www.amstat.org/education/gaise/GAISEPreK-12.htm

Franklin, C., & Mewborn, D. (2006). The statistical education of grades pre-K–12 teachers: A shared responsibility. In G. Burrill (Ed.), Thinking and reasoning with data and chance: 68th NCTM yearbook (pp. 335-344). Reston, VA: National Council of Teachers of Mathematics.

Franklin, C., & Garfield, J. (2006), "The Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project: Developing Statistics Education Guidelines for Pre K-12 and College Courses," In G.F. Burrill, (Ed.),Thinking and Reasoning about Data and Chance: Sixty-eighth NCTM Yearbook (pp. 345-375). Reston, VA: National Council of Teachers of Mathematics.

Franklin, C., (2010) Are Our Teachers Prepared to Provide Instruction in Statistics at the K–12 Levels? *National Council of Teachers of Mathematics.* Retrieved from http://www.nctm.org/resources/content.aspx?id=1776

Franklin, C., Hartlaub, B., Peck, R., Scheaffer, R., Thiel, D., & Freier, K.T., (2011). AP Statistics: Building Bridges Between High School and College Statistics Education. *The American Statistician*, *65(3)*, 177-182.

Frey, B., (2005). Item-Writing Rules: Collective Wisdom. *Teaching and Teacher Education, 21(4)*, 357-364.

Garfield, J., & Everson, M. (2009). Preparing Teachers of Statistics: A Graduate Course for Future Teachers. *Journal of Statistics Education, 17(2),* Retrieved from www.amstat.org/publications/jse/v17n2/garfield.html.

Garfield, J., and Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. *International Statistical Review*, *75(3)*, 372–396.

Gelman A., (2005). A Course on Teaching Statistics at the University Level. *The American Statistician*, *59(1)*, 4-7.

Groth, R.E., & Bergner, J.A. (2005). Pre-Service Elementary Teachers' Metaphors For The Concept Of Statistical Sample. *Statistics Education Research Journal*, *4(2),* 27-42.

Groth, R.E. (2007). Towards a Conceptualization of Statistical Knowledge for Teaching. *Journal for Research in Mathematics Education, 38(5)*, 427-437.

Groth, R.E. (2008). Assessing Teachers' Discourse About The Pre-K-12 Guidelines For Assessment And Instruction In Statistics Education (GAISE). *Statistics Education Research Journal*, *7(1),* 16-39.

Groth, R.E. (2013). Characterizing Key Developmental Understandings and Pedagogically Powerful Ideas Within a Statistical Knowledge for Teaching Framework. *Mathematical Thinking and Learning, 15*, 121-145.

Grossman, P., Schoenfeld, A., & Lee, C. (2005). In: Preparing teachers for a changing world: What teachers should learn and be able to do. L. Darling-Hammond & Bransford, J (Eds.). (pp. 201-231). San Francisco, CA : Jossey-Bass.

Gould, R., & Peck, R. (2004). Preparing Secondary Mathematics Educators to Teach Statistics. *Curricular Development in Statistics Education*, 244-255.

Haladyna, T.M., & Downing, S.M. (1989). A Taxonomy of Multiple-Choice Item Writing Rules. *Applied Measurement in Education, 2(1),* 37-50.

Hancock, G. R., & Mueller, R. O. (2006). *Structural equation modeling: A second course.* Greenwich, CT: Information Age Publishing, Inc.

Harkness, W.L., & Rosenberger, J.L. (2005). Training Graduate Students at Penn State University in Teaching Statistics. *The American Statistician*, *59(1)*, 11-13.

Hassad, R. A. (2006, July). *Reflection On Training, Experience, and Introductory Statistics A Mini-Survey Of Tertiary Level Statistics Instructors*. Paper presented at the International Association for Statistical Education (IASE) International Conference on the Teaching of Statistics, Salvador, Bahia, Brazil.

Hawkins, A. (1990). Training Teachers to Teach Statistics. *Teaching Statistics, 11(3),* 83-84.

Hill, H., Schilling, S., & Ball, D. L. (2004). Developing measures of teacher's mathematics knowledge for teaching. *Elementary School Journal, 105(1),* 11–30.

Hill, H., Rowen, B., & Ball, D.L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42(2)*, 371-406.

Hill, H., Ball, D.L., & Schilling, S.G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal of Research in Mathematics Education, 39(4),* 372-400.

Izsák, A., Orrill, C.H., Cohen, A.S., & Brown, R.E. (2010). Measuring Middle Grades Teachers' Understanding of Rational Numbers with the Mixture Rasch Model. *The Elementary School Journal, 110(3)*, 279-300.

Krauss, S., Brunner, M., Kunter, M., & Baumert, J. (2008). Pedagogical Content Knowledge and Content Knowledge of Secondary Mathematics Teachers. *Journal of Educational Psychology*, *100(3)*, 716-725.

Krosnick, J. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company.

Leavy, A. (2006). Using Data Comparisons To Support A Focus On Distribution: Examining PreService Teachers' Understandings Of Distribution When Engaged In Statistical Inquiry. *Statistics Education Research Journal*, 5(2), 89-114.

Leavy, A.M. (2010). The Challenge Of Preparing PreService Teachers To Teach Informal Inferential Reasoning. *Statistics Education Research Journal*, *9(1)*, 46-67.

Liu, Y., & Thompson, P. (2007). Teachers' Understanding Of Probability. *Cognition and Instruction*, *25(2)*, 113-160.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Manizade, A.B., & Mason, M. (2010). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics, 76*, 183-207.

Mills, J.D. (2007). Teacher Perceptions and Attitudes About Teaching Statistics in P-12 Education. *Educational Research Quarterly*, *30(4)*, 16-34.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teachers' knowledge. *Teachers College Record, 108(6),* 1017–1054.

Myer, J.P. (2009). jMetrik (Version 2.1.0) (Software). Available from http://www.itemanalysis.com/index.php

National Board of Professional Teaching Standards (2010) Mathematics Standards for teachers of students ages 11-18+. *NBPTS*. Retrieved from (http://www.nbpts.org/userfiles/file/aya_math_standards.pdf).

National Council of Teachers of Mathematics. (2000). Data Analysis and Probability Standard for Grades 9-12. *NCTM*. Retrieved from http://standards.nctm.org/document/chapter7/data.htm

Peters, S.A. (2009). *Developing an Understanding of Variation: A.P. Statistics Teachers' Perceptions and Recollections of Critical Moments* (Doctoral Dissertation). Retrieved from: http://search.proquest.com.proxygw.wrlc.org/pqdtft/advanced?ip?accountid=11243

Pfannkuch, M. (2006). Comparing Box Plots Distributions: A Teacher's Reasoning. *Statistics Education Research Journal*, *5(2)*, 27-45.

Roberts, R., Schaeffer, R., & Watkins, A. (1999). Advanced Placement Statistics- Past, Present and Future. *The American Statistician, 53(4)*, 307-320.

Schaeffer, R.L. (2006). Statistics and Mathematics: On Making a Happy Marriage. In G.F. Burrill (Ed.) *Thinking and Reasoning with Data and Chance Sixty-eighth Yearbook*. (pp. 309-322). Reston, VA: NCTM.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando, FL: Academic Press.

Schumacker, R.E., & Lomax, R.G. (2010). *A Beginner's Guide to Structural Equation Modeling*. New York, NY: Routledge.

Sharma, S. (2007). Exploring Pre-Service Teachers' Understanding Of Statistical Variation: Implications For Teaching And Research. *Australian Senior Mathematics Journal*, *21(2)*, 31-43.

Schmidt, D.A., Baran, E., & Thompson, A.D. (2011). Technological Pedagogical Content Knowledge (TPACK): The Development and Validation of an Assessment Instrument for Preservice Teachers. *Journal of Research on Technology in Education, 42(2),* 123–149.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15(2)*, 4-14.

Tchoshanov, M.A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics, 76*, 141-164.

Van Driel, J.H. & Berry A. (2012). Teacher Professional Development Focusing on Pedagogical Content Knowledge. *Educational Researcher, 41(1)*, 26-27.

Watson, Callingham and Donne (2008) Establishing PCK for Teaching Statistics. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.) *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Statistical Institute and International Commission of Mathematical Instruction.

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

Appendix A

Advanced Placement Statistics Teaching Knowledge Survey
© Copyright 2014 by Brenna J. Haines
All rights reserved

Thank you for agreeing to take the A.P. Statistics Teaching Knowledge Survey!

Your responses to the following questions are very important, and will be combined with many other A.P. Statistics' teacher responses to gather information about the types and amounts of teaching knowledge A.P. Statistics teachers use in the classroom.

There is no time limit for completing this survey, and you are free to stop and start again at any time. There are 25 total questions: 12 multiple-choice, 9 constructed response, and 4 short questions about your teaching background. It is expected that it will take approximately 45 minutes to complete the survey.

Please read the attached sheet for more information regarding purpose, risks, benefits and confidentiality issues pertaining to participating in this study.

I have read the attached information sheet and understand that by logging in I accept the risks and benefits associated with completing this survey. BOX CHECK

If you have any questions or comments, please contact Brenna Haines at bhaines@gwu.edu.

Section 1: Multiple-Choice

In this section, there are 12 multiple choice content knowledge questions covering 4 content areas: exploring data, sampling and experimentation, anticipating patterns, and statistical inference. These 4 content areas are then cross-referenced with 3 characteristics of content knowledge that is specific to teaching A.P. Statistics: problem-solving knowledge, organizing content, and student learning.

This section will take approximately 15 minutes to complete. Please feel free to use a graphing calculator when needed. Assume $\alpha = 0.05$ when applicable.

(1). A.P. Statistics teachers know appropriate sampling techniques to estimate population characteristics.

To obtain a sample of 35 students, 500 students are present in a given school day. Every 25th student waiting in line to attend a required assembly in the gym is selected. Will this procedure result in a simple random sample?

A. No
B. Yes

(2). A.P. Statistics teachers anticipate patterns of random variables.

Which of the following statements is false?

A. The standard deviation of a random variable is always positive.
B. The standard deviation of a random variable is zero if the random variable takes on a single value in every case.
C. The standard deviation of a random variable will be negative if the values of the random variable are negative.

(3). A.P. Statistics teachers use statistical inference to draw conclusions.

Assume the average daily ownership cost of an automobile is $15.35. This includes depreciation, finance charges, general maintenance, gasoline and insurance, and assumes 12,000 miles driven annually. A researcher found that a random sample of 34 college students who own cars have an average daily ownership cost of $11.85 with a sample standard deviation of $6.21. Is the population of college students' average daily ownership cost less than the national average?

A. Yes
B. No

(4). A.P. Statistics teachers use expected value to make statistical decisions.

A high school student government is sponsoring a spring break Hawaiian cruise raffle. The proceeds from the raffle will be donated to a local animal shelter. A local travel agency donated the cruise, valued at $5000. Students sold a total of 2852 raffle tickets at $5.00 per ticket. One particular student wants to contribute at least $20 to the shelter regardless of whether he wins the cruise or not. Is this possible if he buys six tickets?

A. Yes
B. No

(5). A.P. Statistics teachers know prerequisite skills for A.P. Statistics' students.

Which of the following skills/topics would you consider to be the most related to the skills a student needs to master prior to taking A.P. Statistics?

A. Unit Circle
B. Slope
C. Integrals
D. Quadratic Functions

(6). A.P. Statistics teachers explain connections between statistics topics within the A.P. curriculum.

A student asks her teacher to explain how sampling technique choices can affect trials in a binomial experiment? The teacher answers by saying, "In a binomial distribution, the trials must be independent. This means that the probability of success must be the same in every trial irrespective of what happened in the previous trial."

What important piece of information did the teacher leave out?

A. Sample size affects independence among trials
B. Independence among trials affects sample size
C. Sample size does not affect population size

(7). A.P. Statistics teachers explain connections between statistics topics and other mathematics topics/procedures.

The formula for the variance of a variable "x" is defined as $\delta^2 = \Sigma (x_i - \mu_x)^2 p_i$ where x is the independent variable, μ is the mean of x and p is the probability of x. What pre-requisite math skill does an A.P. Statistics student need to master prior to being able to successfully perform this calculation?

A. Simplify Square Roots
B. Multiplying Exponents
C. Order of Operations
D. Dividing Fractions

(8). A.P. Statistics teachers apply the College Board standard of grading to general classroom practice.

An assessment question that is making a claim about a population parameter is asking for a hypothesis test to be performed. However, one of your A.P. Statistics' students answers a hypothesis test question by constructing a confidence interval estimate of a parameter to determine if the parameter value is contained within the interval found.

If you are scoring her answer using the same criteria as an A.P. grader, would you give her full credit, partial credit, or no credit?

A. Full credit
B. Partial credit
C. No credit

(9). A.P. Statistics teachers understand problems students experience while learning conditional probabilities.

You give an A.P. Statistics student the following problem to solve:

A psychologist is interested in right-handedness verses left-handedness and IQ scores. The psychologist collected the following data from a random sample of 2000 high school students.

	Right Handed	Left Handed	Total
High IQ	190	10	200
Normal IQ	1710	90	1800
Total	1900	100	2000

What is the probability that a student has a high IQ; given that she is left-handed? Does the occurrence of High IQ affect Left-Handedness?

The student answers P (high IQ | left-handed) = 0.1, and the events are independent because the occurrence of one event does not affect the other.

Does this student understand conditional probability?

A. Yes
B. No

(10). A.P. Statistics teachers know if students can apply procedures outside of the context in which they were learned.

You give your A. P. Statistics students the following problem:

Company I manufactures out-door lights that last an average of 50 minutes with a standard deviation of 10 minutes, while Company II advertises out-door lights that last an average of 55 minutes with a standard deviation of 5 minutes. Which company's lights are more likely to last at least 1 hour?

Which one of the following A.P. Statistics' procedures must students know in order to obtain the correct solution?

A. Read a chi-squared distribution table
B. Standardize a sample statistic
C. Control for randomization
D. Plan a statistical experiment

(11). A.P. Statistics teachers provide students the opportunity to evaluate recently published statistical results.

You ask your A.P. Statistics students to evaluate the following hypothetical claims from three newspaper articles:

1. Juveniles who watch more violence on TV are more likely to exhibit antisocial behavior.
2. TV violence is positively associated with antisocial behavior.
3. If juveniles were to watch less violence on TV, they would exhibit less antisocial behavior.

What statistical concept would you teach prior to having students evaluate these articles?

A. Causation
B. Simulation
C. Observation
D. Replication

(12). A.P. Statistics teachers help students learn basic statistical facts and procedures over time.

Which of the following formulas might students have to memorize and correctly use in order to do well on the A.P. Statistics exam?

A. Mean of a Sampling Distribution
B. Mean of a Geometric Distribution
C. Mean of a Binomial Distribution
D. Mean of a Random Variable

Section 2: Constructed Response

In this section, there are 9 constructed response pedagogical content knowledge questions covering the same 4 content areas as above, but also cross-referenced with different characteristics of pedagogical content knowledge for teaching A.P. Statistics: adjusting curriculum, assessment and instruction strategies; recognizing and interpreting student errors and misconceptions; and student thinking and learning.

This section will take approximately 30 minutes to complete. Please feel free to use a graphing calculator when needed. Assume $\alpha = 0.05$ when applicable.

(13). A.P. Statistics teachers assess the statistical accuracy of information in an A.P. Statistics textbook.

An A.P. statistics textbook explains independence among random variables by saying: "if two random variables are independent their variances add algebraically, but their standard deviations do not".

Is this statement accurate for teaching A.P. Statistics?

Why or why not?

(14). A.P. Statistics teachers must be able to recognize errors in student understanding of a particular concept.

You provide the following test question to your A.P. Statistics students:

Assume there is no overlap between 56% of the adult population who wear glasses and the 4.5% who wear contacts. If 55% of those who wear glasses are women and 63% of those who wear contacts are women, what is the probability that a person chosen from a random sample will be a man who wears glasses?

A student answers by writing the following:

(.37)(.045) = 0.017 is the probability that a person chosen will be a man who wears glasses

Explain the error in the student's calculation, and then show the correct calculation.

(15). A.P. Statistics teachers evaluate and adjust assessment items based on student results.

On a recent assessment, you ask students the following question:

What type of statistical test has an H_0: $\mu_{diff} = 0$?

Due to many incorrect student responses, you decide to revise the question. Write a new question that addresses the original concept you were trying to assess.

(16). A.P. Statistics teachers provide students appropriate feedback.

One of your A. P. Statistics students states, "sampling error and bias are two things that must be avoided while collecting data".

Please describe feedback you might give this student.

(17). A.P. Statistics teachers distinguish misconceptions from reasoning on student work.

You ask your A.P. Statistics students the following question:

A random sample of farms in Kansas shows a mean yield of 29 acres of wheat with a standard deviation of 3.2 acres. If all the measurements are converted from acres to square yards, what are the new mean and standard deviation? (Note: 1 acre = 4840 square yards)

A student answers by writing the following:

First, I need to multiply the current mean, 29 by 4840, so the new mean is equal to 29*4840 = 140, 360 square yards. Multiplying by a constant, 4840 in this case, changes the mean but does not change the distance between the values. Therefore, the standard deviation stays the same.

Which answer choice best explains this student's misconception?

(18). A.P. Statistics teachers must be able to assist students in extending their knowledge of basic statistical concepts to include more advanced A.P. Statistics concepts and skills.

Please state a question you would include on an assessment to check students' understanding of independence tests. Then, state a question you would include on an assessment that tests students' knowledge of the chi-squared goodness-of-fit test.

How would you explain to students the difference between the concepts in the two questions you pose?

(19). AP Statistics teacher recognize when students need to learn more of a particular concept.

You give an AP Statistics student the following question to answer:

A candidate for political office claims to have the support of 65% of the people. A researcher plans to gather data from a simple random sample to test the claim. You ask your students the following question: When will the null hypothesis be rejected?

The student answers by saying "when the sample proportion is less than 60%, or based on the power of the test".

Does this student need more instruction in hypothesis testing?

Why or why not?

(20). A.P. Statistics teachers show students how theoretical calculations often differ from actual results.

You ask an A.P. Statistics student the following question: In general, how does tripling sample size affect confidence interval size?

The student answers the question by saying that the interval size will also be tripled.

You know that this answer is incorrect.

Describe how you would explain to the student that his answer is incorrect.

(21). A.P. Statistics teachers extend students' thinking beyond memorization and calculations.

You have recently discovered that your A.P. Statistics students can recognize normally distributed and skewed data from looking at histograms.

Describe a lesson you would use to help deepen students' understanding of normal and skewed data.

Section 3: Teacher Characteristic Questions:

Please take a moment to answer to following questions to help us further understand characteristics of current A.P. Statistics teachers.

(22). Do you have a college-level degree in statistics?

Yes No

(23). How many college-level statistics courses have you completed?

<1
2-3
3+

(24). How many years have you been teaching A.P. Statistics?

0-5
6-10
11-15
15+

(25). If you taught A.P. Statistics last year, what percent of your students scored a 3 or higher on the A.P. Statistics exam?

_____

Appendix B

Advanced Placement Statistics Teaching Knowledge Survey
© Copyright 2014 by Brenna J. Haines
All rights reserved

Thank you for agreeing to take the A.P. Statistics Teaching Knowledge Survey!

Your responses to the following questions are very important, and will be combined with many other A.P. Statistics teachers' responses to gather information about the types and amounts of teaching knowledge A.P. Statistics teachers use in the classroom.

There is no time limit for completing this survey, and you are free to stop and start again at any time. There are 14 multiple-choice questions and 4 short answer questions about your teaching background. It is expected that it will take approximately 20 minutes to complete the survey.

Please read the attached sheet for more information regarding purpose, risks, benefits and confidentiality issues pertaining to participating in this study.

I have read the attached information sheet and understand that by logging in I accept the risks and benefits associated with completing this survey. BOX CHECK

There is a section at the end of the survey to provide additional comments or feedback. If you have any other questions or comments, please contact Brenna Haines at bhaines@gwu.edu.

Section 1: Pedagogical Knowledge

In this section, there are 5 questions designed to assess typical A.P. Statistics pedagogical knowledge. Each question provides sample-teaching scenarios followed by recent, authentic teacher responses placed on a continuum. For each question, please check the response that best represents your own knowledge of teaching A.P. Statistics.

This section will take approximately 10 minutes to complete. Feel free to use a graphing calculator if needed. Assume $\alpha = 0.05$ if applicable.

(1). A.P. Statistics pedagogical knowledge includes assessing the statistical accuracy of information in an A.P. Statistics textbook.

An A.P. statistics textbook explains independence among random variables by saying: "if two random variables are independent their variances add algebraically, but their standard deviations do not".

Is this statement accurate? Check the box next to the response that best represents your thinking.

_____Yes, variances add algebraically and standard deviations add geometrically
_____Yes, but independence should be defined in terms of probability
_____Yes, but this is only one characteristic of independent variables
_____Yes, but this is not an acceptable explanation for the A.P. Statistics course

(2). A.P. Statistics pedagogical knowledge includes recognizing errors in student understanding of a particular concept.

The following test question is provided to A.P. Statistics students:

56% of the adult population wears only glasses and 4.5% wears only contacts. If 55% of those who wear only glasses are women and 63% of those who wear only contacts are women, what is the probability that a person chosen from a random sample will be a man who wears only glasses?

A student answers by writing the following:

The probability that a person chosen will be a man who wears glasses is (.37)(.045) = 0.017.

What error did the student make? Check the box next to the response that best represents your thinking.

_____The student failed to realize the events are not independent.
_____The student calculated the probability of a man wearing only contacts.
_____The student multiplied incorrectly due to a decimal place error.
_____The student calculated the probability of a woman wearing only contacts.

(3). A.P. Statistics pedagogical knowledge includes evaluating and adjusting assessment items based on student learning.

On a recent assessment, students are asked the following question:

What type of statistical test has an H_0: $\mu_{diff} = 0$?

Due to many incorrect student responses, you decide to write a new question to assess the same concept. Check the box next to the question that best represents your new question or use the space provided to write a new question.

_____What is the null hypothesis for a two-sample t-test for difference of population means?
_____What is the null hypothesis for a matched pairs z-test for difference of population proportions?

_____What is the null hypothesis for a matched pairs t-test for difference of population means?
_____Write a new question: _____

(4). A.P. Statistics pedagogical knowledge includes providing students appropriate feedback.

An A. P. Statistics student states, "sampling error and bias are two things that must be avoided while collecting data".

Please describe the most appropriate feedback you would give this student. Check the box next to the response that best represents your thinking.

_____Sampling error cannot occur during data collection.
_____Sampling error is a form of bias that can distort results.
_____Sampling error cannot be avoided but bias can be eliminated.
_____Sampling error cannot be avoided, but bias should be minimized.

(5). A.P. Statistics pedagogical knowledge includes identifying misconceptions on student work.

A.P. Statistics students are asked the following question:

A random sample of farms in Kansas shows a mean yield of 29 acres of wheat with a standard deviation of 3.2 acres. If all the measurements are converted from acres to square yards, what are the new mean and standard deviation? (Note: 1 acre = 4840 square yards)

A student answers by writing the following:

First, I need to multiply the current mean, 29 by 4840, so the new mean is equal to 29*4840 = 140, 360 square yards. Multiplying by a constant, 4840 in this case, changes the mean but does not change the relative distance between the values. Therefore, the standard deviation stays the same.

What misconception does the student have? Check the box next to the response that best represents your thinking.

_____Standard deviation needs to be added to the constant.
_____Variance needs to be multiplied by two times the constant.
_____Standard deviation needs to be multiplied by the constant.
_____Relative distance between values change when multiplied by the constant.

Section 2: Content Knowledge

In this section, there are 9 multiple choice content knowledge questions. Please choose the most correct answer from the choices given.

This section will take approximately 10 minutes to complete. Please feel free to use a graphing calculator when needed. Assume $\alpha = 0.05$ when applicable.

(6). A.P. Statistics teachers know appropriate sampling techniques to estimate population characteristics.

To obtain a sample of 35 students, 500 students are present in a given school day. Every 25th student waiting in line to attend a required assembly in the gym is selected.

Will this procedure result in a simple random sample?

C. No
D. Yes

(7). A.P. Statistics teachers anticipate patterns of random variables.

Which of the following statements is false?

D. The standard deviation of a random variable is always nonnegative.
E. The standard deviation of a random variable is zero if the random variable takes on a single value in every case.
F. The standard deviation of a random variable will be negative if the values of the random variable are negative.

(8). A.P. Statistics teachers use statistical inference to draw conclusions.

Assume the average daily ownership cost of an automobile is $15.35. This includes depreciation, finance charges, general maintenance, gasoline and insurance, and assumes 12,000 miles driven annually. A researcher found that a random sample of 34 college students who own cars have an average daily ownership cost of $11.85 with a sample standard deviation of $6.21.

Is the population of college students' average daily ownership cost less than the national average?

C. Yes
D. No

(9). A.P. Statistics teachers use expected value to make statistical decisions.

A high school student government is sponsoring a spring break Hawaiian cruise raffle. The proceeds from the raffle will be donated to a local animal shelter. A local travel agency donated the cruise, valued at $5000. Students sold a total of 2852 raffle tickets at $5.00 per ticket. One particular student wants to contribute at least $20 to the shelter regardless of whether he wins the cruise or not.

Is this possible if he buys six tickets?

C. Yes
D. No

(10). A.P. Statistics teachers explain connections between statistics topics within the A.P. curriculum.

A student asks her teacher to explain how sampling technique choices can affect trials in a binomial experiment? The teacher answers by saying, "In a binomial distribution, the trials must be independent. This means that the probability of success must be the same in every trial irrespective of what happened in the previous trial."

What important piece of information did the teacher leave out?

D. Sample size affects independence among trials
E. Independence among trials affects sample size
F. Sample size does not affect population size

(11). A.P. Statistics teachers explain connections between statistics topics and other mathematics topics/procedures.

The formula for the variance of a variable "x" is defined as $\delta^2 = \Sigma (x_i - \mu_x)^2 p_i$ where x is the independent variable, μ is the mean of x and p is the probability of x.

What pre-requisite math skill does an A.P. Statistics student need to master prior to being able to successfully perform this calculation?

E. Simplify Square Roots
F. Multiplying Exponents
G. Order of Operations
H. Dividing Fractions

(12). A.P. Statistics teachers apply the College Board standard of grading to general classroom practice.

An assessment question that is making a claim about a population parameter is asking for a hypothesis test to be performed. However, one of your A.P. Statistics students answers a hypothesis test question by constructing a confidence interval estimate of a parameter to determine if the parameter value is contained within the interval found.

If you were scoring her answer using the same criteria as an A.P. grader, would you give her full credit, partial credit, or no credit?

D. Full credit
E. Partial credit
F. No credit

(13). A.P. Statistics teachers provide students the opportunity to evaluate recently published statistical results.

You ask your A.P. Statistics students to evaluate the following hypothetical claims from three newspaper articles:

1. Juveniles who watch more violence on TV are more likely to exhibit antisocial behavior.
2. TV violence is positively associated with antisocial behavior.
3. If juveniles were to watch less violence on TV, they would exhibit less antisocial behavior.

What statistical concept would you teach prior to having students evaluate these articles?

E. Causation
F. Simulation
G. Observation
H. Replication

(14). A.P. Statistics teachers help students learn basic statistical facts and procedures over time.

Which of the following formulas must students memorize in order to prepare for the A.P. Statistics exam?

E. Mean of a Sampling Distribution
F. Mean of a Geometric Distribution
G. Mean of a Binomial Distribution
H. Mean of a Random Variable

Section 3: Teacher Characteristic Questions:

Please take a moment to answer to following questions to help us further understand characteristics of current A.P. Statistics teachers.

(1). Do you have a college-level degree in statistics?

Yes No

(2). How many college-level statistics courses have you completed?

(3). How many years have you been teaching A.P. Statistics?

(4). In what year did you last teach AP Statistics? _____

In that year, what percent of your students scored a 3 or higher on the AP Statistics exam? _____

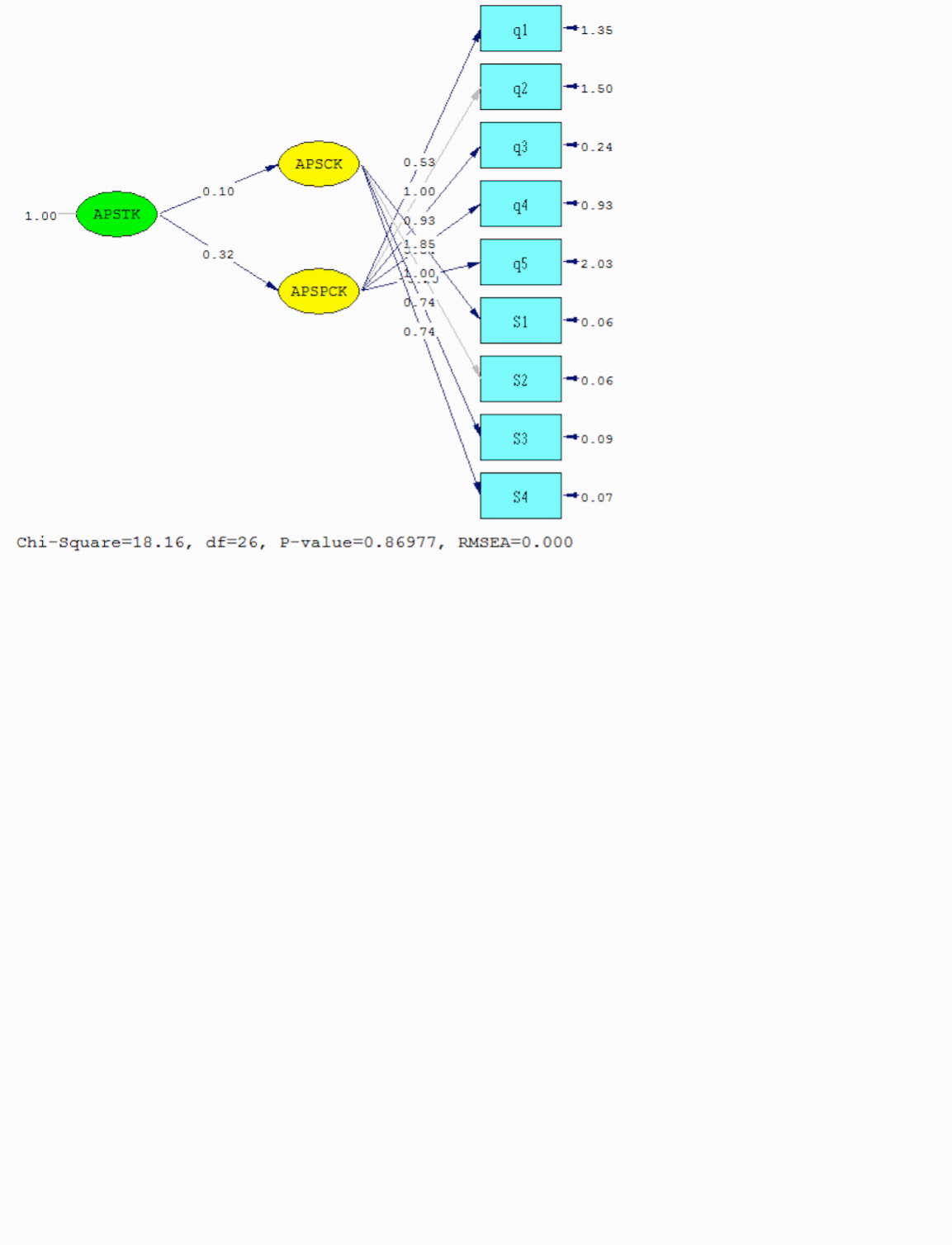
Feedback or Comments:

Don't forget to enter your email address here for a chance to win a $50 gift card to Amazon.com!

A summary of these survey results will be available after December 1st. Please feel free to email me at bhaines@gwu.edu if you would like to see the summary report.

Appendix C

Structural Model Path Diagram



Appendix D

Category Descriptions of Final APSCK Items

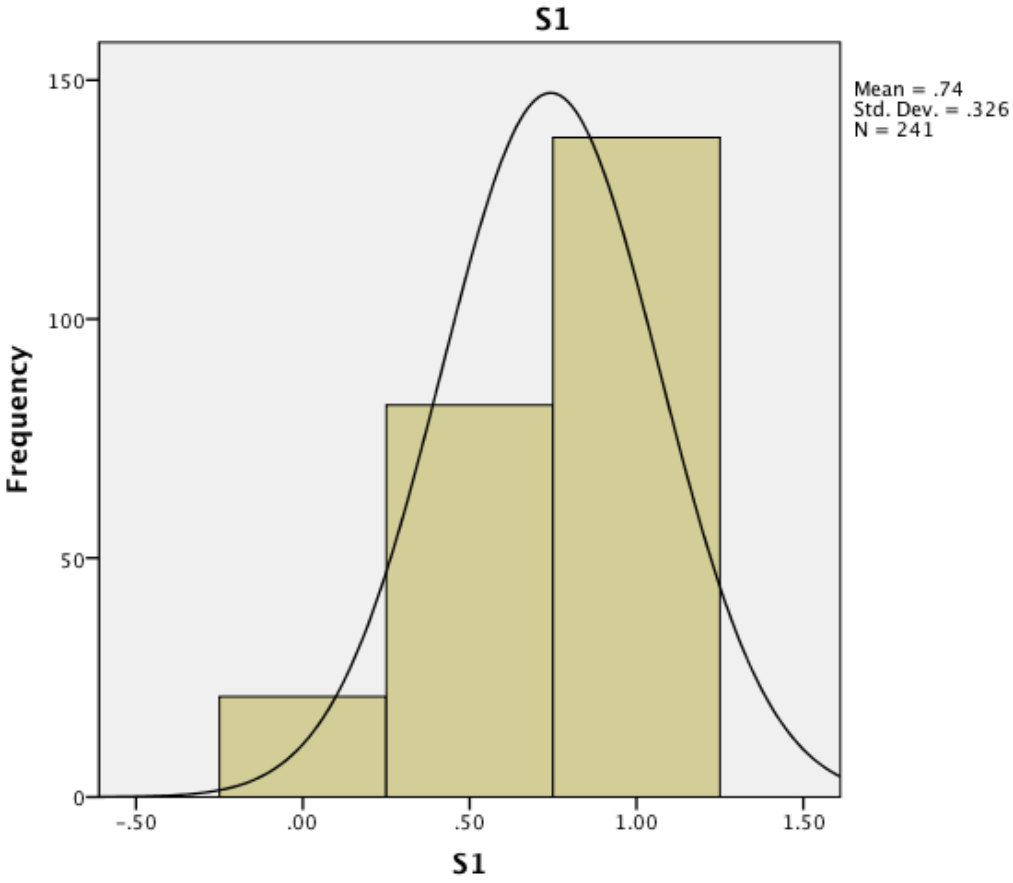
	APSCK Measurable Characteristics		
A.P. Statistics Content Sub-Domains	Problem Solving Knowledge	Organizing Content	Student Learning
Sampling and Experimentation (S1)	6	10	
Anticipating Patterns (S2)	7	13	
Statistical Inference (S3)	8	14	12
Exploring Data (S4)	9	11	

Category Descriptions of Final APSPCK Items

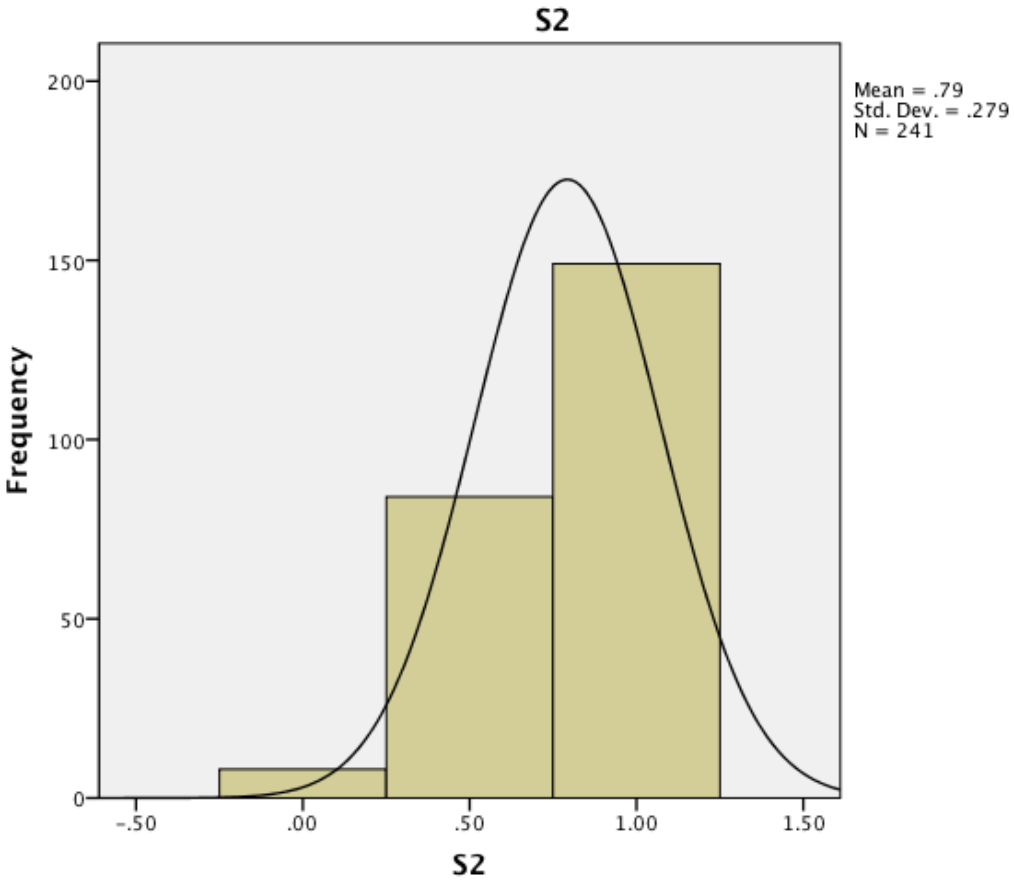
	APSPCK Measurable Characteristics		
A.P. Statistics Content Sub-Domains	Adjust curriculum, instruction and assessment strategies	Recognize and interpret errors and misconceptions	Student thinking and learning
Exploring Data		5	
Sampling and Experimentation			4
Anticipating Patterns	1	2	
Statistical Inference	3		

Appendix E

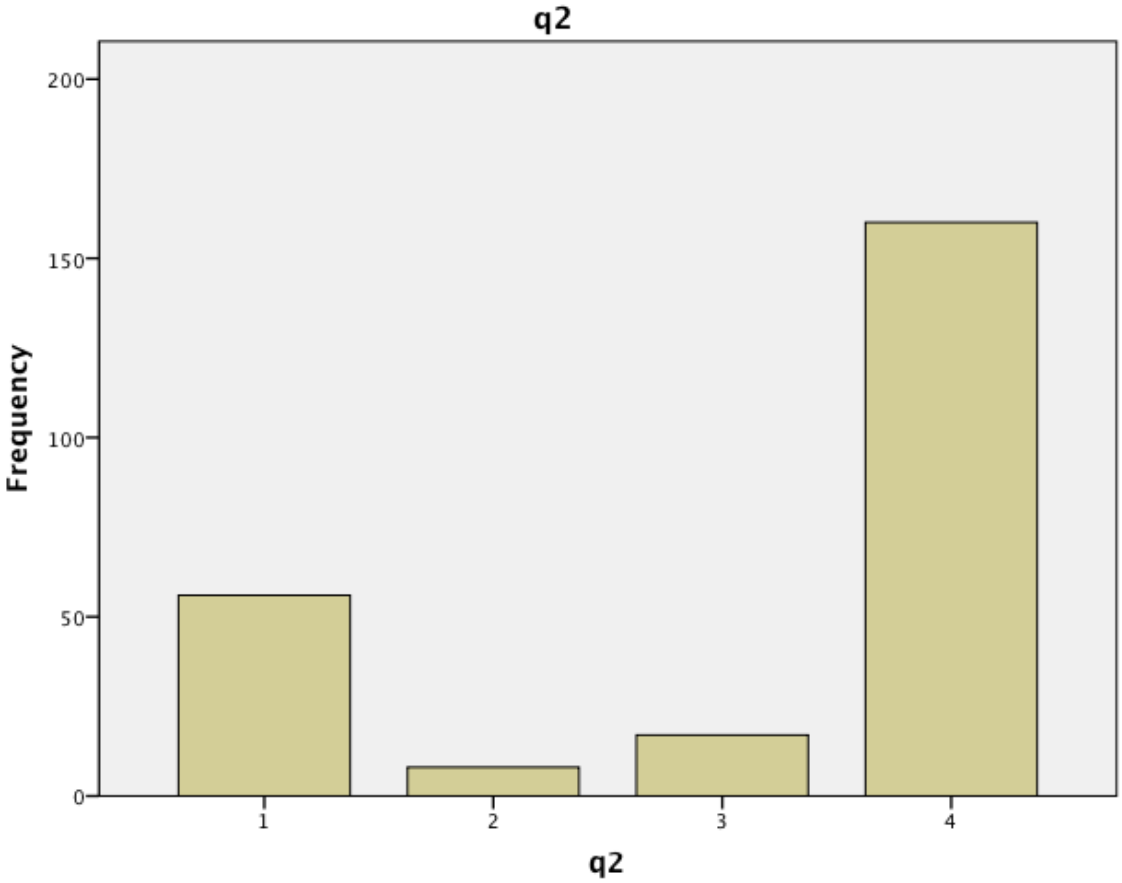
Frequency Distribution of Score Values for Subscale 1 (S1) APSCK Items: Exploring Data, Problem Solving Knowledge and Organizing Content



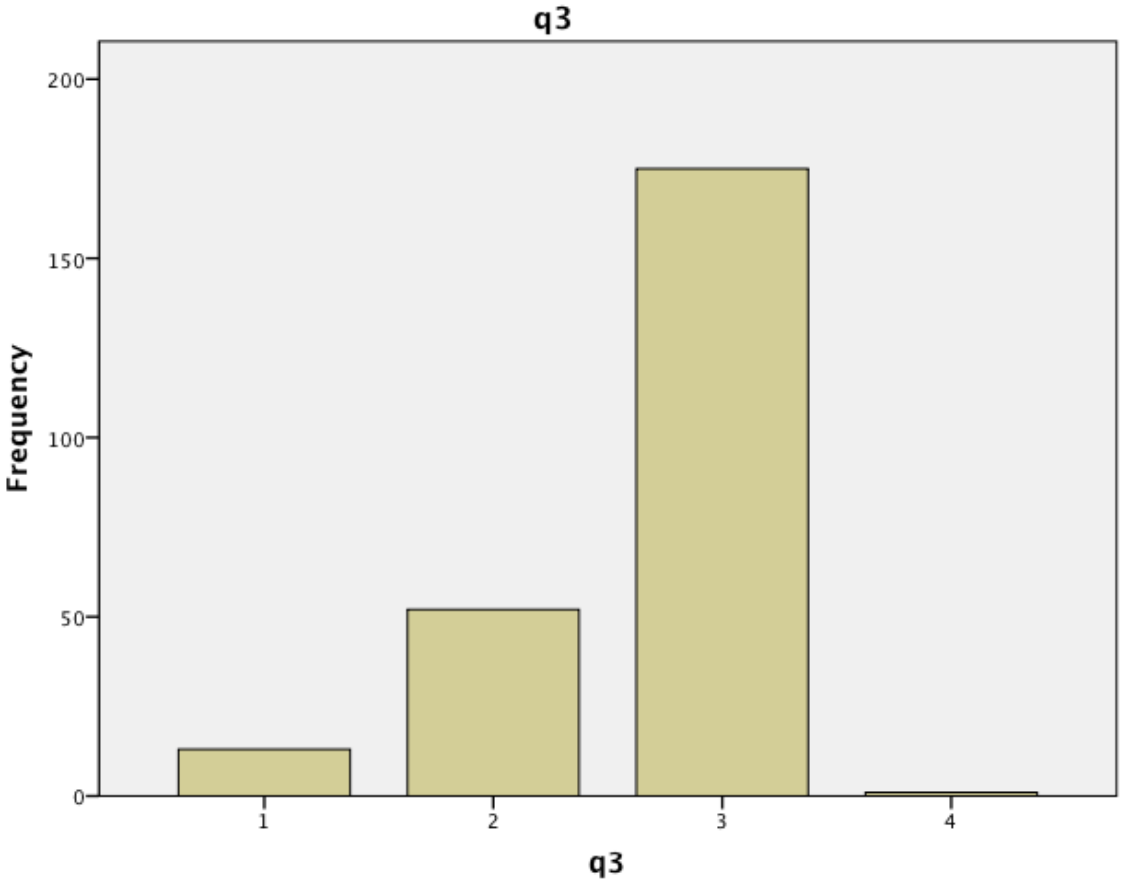
Frequency Distribution of Score Values for Subscale 2 (S2) APSCK Items: Sampling and Experimentation, Problem Solving Knowledge and Organizing Content



Frequency Distribution of Score Values for APSPCK Item Number Two: Anticipating Patterns and Recognizing/Interpreting Errors and Misconceptions



Frequency Distribution of Score Values for APSPCK Item Number Three: Statistical Inference and Adjusting Curriculum, Instruction and Assessment Strategies



Appendix F

APSTK Assessment Score Data Compared to Item-Level Scores

Assessment Score by Quartile	N	APSTK Cut-Off Score	APSPCK q2 average score	APSPCK q3 average score	APSCK S1 average score	APSCK S2 average score
Q1	67	<64.00%	60.00	61.00	49.00	63.00
Q2	52	64.00%	85.00	66.00	75.00	76.00
Q3	56	72.50%	82.00	71.00	82.00	91.00
Q4	63	>81.00%	96.00	73.00	98.00	93.00