Dissertations

Spring 5-30-2018

The Use of Student Feedback in Teacher Development

Lawrence Jarocki
Brandman University, jaro2601@mail.brandman.edu

Follow this and additional works at: https://digitalcommons.umassglobal.edu/edd_dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, Elementary and Middle and Secondary Education Administration Commons, Secondary Education Commons, and the Secondary Education and Teaching Commons

The Use of Student Feedback in Teacher Development

A Dissertation by

Lawrence Jarocki

Brandman University

Irvine, California

School of Education

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Education in Organizational Leadership

September 2018

Committee in charge:

Dr. Laurie Goodman, Ed. D, Committee Chair

Dr. Keith Larick, Ed.D

Dr. Lois Wynne, Ed.D

BRANDMAN UNIVERSITY

Chapman University System

Doctor of Education in Organizational Leadership

The dissertation of Lawrence Jarocki is approved.

_____, Dissertation Chair

<< Dr. Laurie Goodman, Ed.D.

_____, Committee Member

Dr. Keith Larick, Ed.D

_____, Committee Member

Dr. Lois Wynne, Ed.D

_____, Associate Dean

August 2018

The Use of Student Feedback in Teacher Development

Copyright ©2018

by Lawrence A. Jarocki

ACKNOWLEDGEMENTS

I would like to acknowledge everyone who has made this accomplishment possible. Without the help of my friends, family, and colleagues, this dissertation would not have been nearly so successful.

First, I'd like to thank my mother for her constant support in this process. Without her personal example of continual self-improvement, I would not have had the model for the determination necessary for such an undertaking.

My next hearty thanks go out to Brandman University for establishing this program. Through the coursework, projects, immersions, and camaraderie with my cohort and instructors, I have become a more balanced and productive person. Whatever I achieve as a leader in the field of education will be largely due to the knowledge, skills, and connections I have acquired through my doctoral studies.

Of course, I must give thanks to Dr. Laurie Goodman, my cohort mentor, dissertation chair, and self-improvement guru. Laurie, you are a positive inspiration to all that meet you. Your incisiveness, persistence, and kindness buoyed me up when I needed it, keeping me going through this long process.

Finally, and most importantly, I must say thanks to my wife and children. Too often I have been holed up in the office, typing away on the latest draft; thank you for being patient with that. From now on, our trips to Irvine will be in the interest of visiting the Magic Kingdom, not for the latest immersion, I promise.

ABSTRACT

The Use of Student Feedback in Teacher Development

by Lawrence Jarocki

Purpose: The purpose of this study was to explore the perceptions of master teachers, administrators, and teacher trainers about the content of Student Evaluations of Teachers (SET) in California high schools. This study also sought to reach a consensus among experts concerning how SETs can be used both in teacher evaluations and in professional development practices and content at the secondary level.

Methodology: A classical Delphi method was utilized to collect perceptual data from a panel of California master teachers, administrators, and teacher trainers that met specific criteria regarding their education, involvement in their professional communities, and their role training of new and experienced teachers. For the purposes of this Delphi study, an electronic questionnaire was distributed in three rounds to assess the participants' perceptions of the content and use of SETs to inform evaluation and professional development practices.

Findings: Analysis of the mixed methods data indicated a variety of findings. First, a collection of forty-nine potential SET questions were generated and ranked. Next, participants favored using SETs at the secondary level for informing professional development purposes over using them as a weighted factor in teacher evaluations. They also gave higher rankings to questions that addressed a teacher's actions and affect in the classroom over those that dealt with course content and activities. Finally, preference was expressed for twice-yearly implementation, with the resulting data being distributed individually and in aggregated form for subject leads and administrators.

Conclusions: This study supported the use of SETs at the secondary level, particularly to inform professional development processes. It also revealed continued resistance to the use of SETs in teacher evaluations, in part due to the perception that secondary students' biases would influence their ratings.

Recommendations: Further research is recommended to explore the effects of teacher unions on SET acceptance and implementation, the possibility of using SETs with younger students, the effects of SET implementation on student voice, and the potential sources of professional development once specific needs are identified through SET use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

The educational environment in the United States has been changing greatly for the past few decades. According to the California Department of Education, forty-five states have adopted the common core standards since 2010. After decades of autonomous action in classrooms, teachers are being asked to teach a unified curriculum in order to ensure a quality education for all students, regardless of where they are being taught ("What Are," 2012). With common curricula and standards-based assessments, it becomes easier for teachers to collaborate on sequencing and instructional practices (Phillip & Hughes, 2012). At the same time, having common curricula also makes it easier for consumers to make direct comparisons between teachers, schools, districts, and states (Mayer & Phillips, 2012). This, in turn, has led administrations to seek ways of investigating what is going on in individual classrooms in terms of teacher effectiveness (Brown-Easton, 2008; Torff & Sessions, 2009).

In a recent example from the Los Angeles Times, the Los Angeles Unified School District and its teachers union agreed to include student results on standards-based tests as part of the teacher evaluation process ("A New Way," 2012). While the degree to which these scores will be taken into consideration is still up for debate, this was the first and largest school district in California to adopt such a policy. Similar measures are being considered or implemented in most states (Darling-Hammond, Amrein-Beardsley, Hartel, & Rothstein, 2012). As a result, teachers are facing increasing pressure from parents and administrators to show increases in student achievement, with possible financial consequences for failing to do so (Walker, 2014).

For teachers seeking to boost student success rates, a key aspect to improving their practice is effective professional development linked to this achievement and to district goals (Kelleher, 2003). As teachers explore their craft, they would also benefit from immediate and incisive feedback on their classroom experiments (Ball & Cohen, 1999;). One potentially rich and commonly underused source of feedback involves the students themselves (Fisher, Fraser, & Cresswell, 1995). However, many teachers fail to take advantage of this resource, for a variety of reasons.

While some educators actively seek out student feedback on their teaching, others are reluctant to use students as a source of information about their craft (Costin, Greenough, & Menges, 1971; Schmelkin, Spencer, & Gellman, 1997). According to the National Comprehensive Center for Teacher Quality, one source of this reluctance is the perception that students lack the ability to make judgments about the entire teaching context. In their report, "A Practical Guide to Evaluating Teacher Effectiveness," the authors cite teachers' concerns that students would rate them not on their effectiveness as instructors but on their personalities and the rigor of their courses. In particular, teachers worry that students will evaluate instructors based on laxity and friendliness (Elbow & Boice, 1992; Little, Goe, & Bell, 2009). However, several studies have shown that secondary students are not more liable to be more biased than university students (Burniske & Meibaum, 2012; Goe, Bell, & Little, 2008). Students themselves expressed that they tend to have better relationships with those teachers that they view as effective than with those they saw as more nurturing (Grooters, 2008; Kane & Staiger, 2012). A study from the Colorado Legacy Foundation found students well able to evaluate via surveys their teachers' classroom practices (Colorado Legacy Foundation, 2013).

Despite the fact that the perception of bias in reporting is often cited as a significant reason for not incorporating student views into the planning of curriculum and instruction, recent studies have found direct connections between students' perceptions of teacher practices and either teachers' own perceptions or student achievement data (Fisher, Fraser, & Cresswell, 1995; O'Shea, 2006). In other words, in some contexts, there is no appreciable difference between what students are reporting and what the teachers are self-reporting about what goes on in their classrooms in terms of the support that teachers provide to students. What is still unclear, however, is the extent that these perceptions differ in other areas of instruction and classroom management (Goe et al., 2008).

A further factor affecting teacher willingness to elicit feedback from students is the question of how much of a difference an individual teacher can make in a student's learning and achievement. For example, one study of Norwegian high school science classes questioned how much of an effect an individual teacher can have on student performance, attributing learning successes to the cumulative effects of years of learning rather than the influence of an individual teacher (Christopherson, Elstad, & Turmo, 2010). However, this same study also finds that teachers can, over a short time period, significantly influence students' perceptions of science and their study habits, including motivation and self-discipline. The lasting effects of good teachers were also confirmed in a recent Harvard study, which found that students of top teachers (i.e., those in the top five percent of value-added rankings) more often went to college and earned more money, and they were less likely to become teen mothers (Chetty, Friedman, & Rockoff, 2011)

A final concern of teachers is that collecting student surveys will evolve from a feedback process to an evaluative one, where, as has happened with the Los Angeles Unified School District with student performance on standardized tests, this data could be used to affect placement, salary, and tenure decisions (Mayer & Phillips, 2012). Despite these misgivings, secondary teachers are using feedback from students to inform classroom management and curriculum decisions, and the outcomes have been positive (Little, Goe, & Bell, 2009; Stecher, Garet, Holtzman, & Hamilton, 2012). Important questions that need to be answered concern the content of these surveys and the ways that their results can be used in initial and ongoing teacher development.

Problem Background

The general problems confronting teachers and administrators in an environment of high accountability is that too little is known about the nature and quality of instruction in individual classrooms (Ball & Cohen, 1999; Darling-Hammond, Amrein-Beardsley, Hartel, & Rothstein, 2012; Elbow & Boice, 1992). Furthermore, the little that is known about classroom practices is not informing professional development practices (Battey & Franke, 2008; Odden, Archibald, Fermanich, & Gallagher, 2002; Webster-Wright, 2009). In his report on possible reforms to current evaluation processes, Peter Youngs (2013) comments on the importance of educators engaging in professional development devoted to expanding their knowledge of all aspects of teaching in order to help them improve student learning (Ball & Cohen, 1999; Youngs, 2013). Unfortunately, current teacher evaluation processes seem to offer teachers little in the way of concrete steps for professional development (PD), as administrative observations and evaluations of teachers are typically conducted in a cursory fashion, resulting in little lasting effects on

instruction or personal decision-making by teachers (Darling-Hammond, Amrein-Beardsley, Hartel, & Rothstein, 2012; Youngs, 2013). This, in turn, has colored teachers' perceptions of the evaluation/professional development cycle because it creates an environment where evaluations are something to be endured and not an opportunity for improving classroom practice or supporting professional growth (Crow, 2011; Towe, 2012; Webb, 1995).

One underutilized solution for improving the effectiveness and perception of evaluation and professional development initiatives comes in the form of student evaluations of teachers (SETs) (Hanover Research, 2013). This practice, common at the university level since the early 1900s, is used by just 5% of U.S. school districts as a means of studying or evaluating teachers (Fulmer, 2013). Joanne Jezequel points out that this potential source of authentic information about a teacher's practices and effectiveness is being little utilized (Jezequel, 2008). This is in part because of teachers' doubts that students can provide valid feedback on the teaching they are experiencing (Ferguson, 2012; Schmelkin et al., 1997). Besides concerns about students' ability to judge good teaching, there is also debate about what that quality teaching looks like (Darling-Hammond et al., 2012; Williams, Sullivan, & Kohn, 2012). Fulmer (2012) concludes that teacher improvement programs need to identify the instructional practices that comprise good teaching and to support teachers in acquiring those practices through more effective professional development.

Models of Effective Teaching

A number of studies have attempted to delineate what constitutes effective classroom practice. In a study of Colorado schools, elements of good teaching were seen to include the following factors:

- teachers' help in enhancing student understanding of material,

- teacher-student personal relationships, including care shown by teachers and mutual respect,

- teacher content knowledge,

- students' feelings of being prepared for the futures,

- classroom management and instruction,

- grading policies and issues of equity, and

- development of student voice. (Colorado Legacy Foundation, 2013)

In contrast, the Measures of Effective Teaching Project (MET), sponsored by the Bill and Melinda Gates Foundation, focused on seven different factors, ranging from how well teachers clarify complex ideas to how they make lessons more coherent through consolidating learning (Ferguson, 2010; Ferguson, 2012). For Helding, the dimensions included student cohesiveness, teacher support, involvement, investigation, task orientation, cooperation, and equity (Helding & Frasier, 2013). In a meta-analysis conducted in 2005, Keane cites two other studies, each with its own number of factors, seven and nine respectively, among which are subject matter mastery, curriculum development, and instructor enthusiasm (Keane & MacLabhrainn, 2005). While there is much overlap in these lists, a definitive set of classroom procedures and practices has yet

to be established (Goe et al., 2008). There is equal confusion concerning the content of SETs.

The Content and Use of SETs

In constructing SETs, one major consideration is the nature of the questions to include (Desimone, 2011). Where Algozzine argues against the use of single, global rating of teacher effectiveness because it cannot adequately express the multi-dimensional nature of teaching (Algozzine et al., 2004), the University of Michigan's Center for Research on Learning and Teaching believes that such questions should be used because there is a higher correlation between student learning and global ratings (Center for Research on Learning and Teaching [CRLT], 2014). For Jezequel (2008), it is important that SETs be multidimensional, eliciting meaningful data on course content, delivery, pacing, workload, and learning outcomes for a particular course. In discussing the Gates Foundation's MET project, Fulmer (2012) asserts that concentrating on factors such as these misses a valuable aspect of science teaching: whether teachers are implementing model-based or inquiry-oriented practices.

In addition, the nature of the questions can have an impact on future teacher practices. According to Kember and Wong, the focus of the items in surveys determines what is actually being evaluated, with the result that teachers might adopt more conservative (e.g. teacher-centered) models of teaching in order to better match what is being evaluating in the SETs. They argue that SET items asking about teacher-centered practices may lead to more traditional and didactic classroom teaching, thus stifling creativity and experimentation (Kember & Wong, 2000). Others believe that by highlighting desired teaching practices, surveys can potentially lead to positive changes

in instruction (Task Force on Educator Excellence, 2012; Webster-Wright, 2009;

Youngs, 2013).

There is also disagreement concerning how SETs should be used in the

evaluation/PD process. According to Algozzine et al (2004), at the university level, the

original purpose of SETs were to facilitate a private conversation between students and

instructors about relative strengths and weaknesses, but they have now evolved into a

means of providing input for decisions about tenure, salary, and promotions. Keane

(2005) agrees that SETs have become simplistic and decontextualized systems that render

them punitive tools rather than support mechanisms for improving the learning

environment. Youngs (2013) argues for the need for administrators to use SETs to

provide teachers with immediate and useful feedback and to inform professional

development decisions. A report from the New Teacher Project adds that better SETs

can also help hold administrators accountable for providing more targeted and effective

professional development experiences to help teachers improve their practice (New

Teacher Project, 2013). This assumes, however, that teachers are willing to allow survey

data to be collected.

Overcoming Resistance to the Use of SETs

Perhaps the greatest obstacle to widespread use of SETs in teacher development is

the attitude of the teachers themselves. A frequent assertion about SETs is that students

lack the knowledge and experience to evaluate their teachers (Costin et al., 1971; Elbow

& Boice, 1992; Schmelkin et al., 1997). Numerous studies, however, have refuted this

claim (Kane & Staiger, 2012; Rockoff & Speroni, 2010). The Colorado Legacy

Foundation (2013) found that although students spend more time with teachers than any

other group in the educational chain, they are rarely asked to comment on how or what teachers are doing in the classroom.   This is despite the fact that studies have found student response to have validity, reliability, and stability over time (Ferguson, 2012; Kane & Cantrell, 2010). The Bill and Melinda Gates Foundation argues that even though individual students might lack a complete understanding of the classroom context, they do experience a teacher's work over the course of an entire year. Also, their scores are averaged among the entire class, which greatly contrasts with a single observer's limited contact with the person he is evaluating (The Bill and Melinda Gates Foundation, 2012). Michigan's Center for Research on Learning and Teaching (2014) affirms students' abilities to comment effectively and reliably on such items as a teacher's preparedness, enthusiasm, and ability to communicate and stimulate interest; however, they do not believe that students can judge a teacher's content knowledge. Keane (2005) adds that although students cannot effectively evaluate course design or grading practices, they are in the best position to provide feedback on content delivery. Jezequel (2008) concludes that, despite the skepticism of some teachers, data shows that students can evaluate them with accuracy and meaning.

Against this growing body of research viewing SETs as a valid and reliable form of collecting information about classroom practices, teachers are still reluctant to undergo the process (Schmelkin et al., 1997). Keane (2005) believes this is because teachers for whom regularized feedback might be a new experience can harbor suspicions about SETs being used for purposes other than professional development. Youngs (2013) recommends that schools make a concerted effort to help teachers see how valuable student data can be, and that principals be trained to use survey data to provide timely

and relevant feedback to teachers. In addition, districts should support principals in

connecting educators with relevant professional development opportunities based on data

obtained from observation and SETs (Crow, 2011; Youngs, 2013). Another study from

Towe (2012) concludes that the use of SETs is a recent and rare phenomenon, but one

that has the potential to positively affect teacher professional development and to reliably

assess teacher effectiveness. For SETs to be effective, however, they need to be

implemented in a way that incorporates research on the ways that adults learn.

Adult Learning Theory and SETs

Although teachers in the secondary setting are working with adolescents, Adult

Learning Theory tells us that teachers' learning processes are different from those of their

students. Chief among Eduard C. Lindeman's assumptions about adult learning are two

ideas that can have a strong influence on teacher development: that adults are best

motivated to learn when they see a connection between the potential learning and real-

world needs, and that they are oriented towards life-centered learning (Knowles, Holton,

& Swanon, 2005).   SETs align closely with the former in that they allow teachers to

receive feedback on their day-to-day work in the classroom (Brown-Easton, 2008;

Webster-Wright, 2009). Because of this, professional development offered by

administrators that is based on needs identified by SETs will have practical and

immediate importance to teachers (Ball & Cohen, 1999; TFEE, 2012; Webster-Wright,

2009). Ferguson (2010) identifies them as a low-cost and efficient mechanism for

improving teaching by incorporating the experiences and perceptions of large numbers of

students. When Knowles (2005) writes that goal of leadership is to release an

individual's energy for the good of the system and to direct that energy toward goals that

benefit all, he is acknowledging how powerful PD activities can be when they enable teachers to explore and improve their instructional practices. By providing individual feedback to teachers on their own work in the classroom though the use of SETs, leaders release and direct the energy of teachers (Hattie, 2012; Kane & Cantrell, 2010).

For the latter implication, that learning is life-centered, SETs provide a focus for the type of self-directed learning that Lindeman identifies (Knowles et al., 2005). When teachers work with data and ideas generated by their own functioning in the classroom, it leads to the kind of life-long learning that is the foundation of Adult Learning Theory (Ball & Cohen, 1999; Jezequel, 2008; Stecher et al., 2012). In talking about supervisory behaviors that support teachers, Glickman, Gordon, & Ross-Gordon (2010) assert that leaders should engender and not inhibit learning that is self-directed by engaging in behaviors that encourage teachers' impulses towards self-direction (Glickman et al., 2010). Implicit in this is the notion that adult learners will respond to direction that helps them with integrating new ideas with their past experience and adapt them to their current practices (Ball & Cohen, 1999; Glickman et al., 2010). Change can come through professional development, but teachers need to feel that the new ideas are tied to their current classroom practice, a condition that can be facilitated through the use of SETs.

The Benefits of Using SETs in the Secondary Setting

One area where student surveys are transforming instruction is in the training of student teachers. Chalwa and Thurkal find student feedback to aid in improving the general teaching competence (Chalwa & Thurkal, 2011). Shadreck and Isaac advocate for the use of data from student surveys in determining the content and topics covered in teacher training institutions (Shadreck & Isaac, 2012). Cherylann Dozier confirms that

pre-service and veteran teachers alike can benefit from data obtained through the use of

student surveys because they will gain insight into effective teaching from learning what

students consider to be sound teaching practice (Dozier, 2012).

Another benefit accrues through giving students a voice in the educational

process: they feel empowered when they find that their views are considered to be of

value, and this is especially true if they see signs of change as a result of their input

(Gentile & Pisanu, 2014; Lawson, Leach, & Burrows, 2012).   In a study on "The Impact

of Evaluation of Teachers on Teacher Practices in a Secondary School," Joanne L.

Jezequel (2008) found that eliciting feedback from students has a positive effect on

student motivation through the establishment of mutual trust and an atmosphere that is

flexible and values student involvement. In addition, using SETs can change student

attitudes about their role in the educational process and the world through teaching them

social responsibility and involving them in democratic processes (Hattie, 2012; Williams,

Sullivan, & Kohn, 2012; Worrell & Dey, 2008).   In general, when students are involved

in evaluating the classroom environment, it increases their agency in managing their own

lives in ways that are both personally meaningful and socially acceptable (Jezequel,

2008; Quaglia & Corso, 2014; Worrell & Dey, 2008;).

A further advantage from the use of SETs comes in the form of increased student

engagement. Jezequel (2008) contends that unmotivated secondary students can become

more engaged if they find that their teachers give credence to their opinions on topics like

teacher effectiveness and performance.   Not only can SETs give important feedback to

teachers, they can also provide students with motivation and a voice (Cook-Sather, 2006;

Quaglia & Corso, 2014).

Perhaps the most important benefit of incorporating student surveys is in the improvement of teachers' practice and understanding of the art and craft of teaching (Ball & Cohen, 1999; TFEE, 2012). When teachers stop using the "autopilot" approach to teaching, they can become transformational practitioners of their profession, ones who invest time in eliciting the opinions of the very customers they are trying to serve (Cook-Sather, 2006; Fenwick, 2006; Hattie, 2012; Thiessen, 2006; Webster-Wright, 2009).

Gaps in Current Knowledge

While long a fixture in university classrooms, SETs are still far from the norm in the secondary school environment. Jezequel (2008) documents that the vast majority of literature on SETs is devoted to their use at the collegiate level. As such, much is still not known about how to create and implement them effectively at the secondary level. Kember and Wong (2000) highlight a dearth of research on the impact of SETs on teachers' beliefs about their craft. They also discuss the need for longitudinal studies to see the effects of SETs on teacher practices over time. Algozzine (2004) calls for more studies on the changes that faculty make in response to student ratings. Ferguson (2012) calls for increased study of the impact of SETs on the implementation of model-based instruction when teachers engage in professional development that takes student views into account.   Finally, in light of the disagreement over the content and practice of SETs (e.g., global versus specific indicators of teacher practice, private versus public dissemination of results), more research is needed to determine what they should contain and how the use of these items can be increased in order to inform teacher reflection and administrative implementation of professional development practices. Questions still remain about how best to elicit student feedback. Keane (2005) argues for the

development of an evaluation system which will then be used appropriately and with general agreement concerning its purpose. Once such decisions have been made, however, putting in place a system of feedback that involves giving voice to students has the potential to enact transformational change in the ways teachers and students interact and perform in our nation's secondary school classrooms.

Problem Statement

In an era of ever-increasing visibility and accountability, teachers are feeling pressured to improve student achievement (Walker, 2014). When seeking ways to enhance classroom practice, they are faced with infrequent administrative evaluation (Hibler & Snyder, 2015), usually tied to employment, and limited or ineffective professional development opportunities (Odden et al., 2002; Shulman, 1986; Webster-Wright, 2009). At the same time, a rich potential source of feedback on their classroom practices remains largely ignored. Though collected routinely at the university level, student feedback is rarely used as a teacher development tool in secondary schools (Hanover Research, 2013). This is partly because of resistance from teacher unions, who fear that such feedback could be used for evaluations (Mayer & Phillips, 2012). Teachers themselves, however, are also resistant for a variety of reasons, and recent calls for the use of feedback (Ferguson, 2012) have fallen on deaf ears. Because the use of student feedback is such a rare and politically-charged practice, few studies of its implementation have been conducted. And yet, who is in a better position to comment on a teacher's work in the classroom than the very students who experience it for hours each week (Jezequel, 2008; Kane & Staiger, 2012)? In order to understand the complexities of

student feedback, further research is needed concerning the content and implementation of SETS.

Purpose of the Study

The purpose of this Delphi study was to identify the most important elements for SETs (Student Evaluation of Teachers) at the high school level as perceived by a panel of expert master teachers, administrators, and teacher trainers. In addition, it was the purpose to determine how the results of SETs can best be used by teacher trainers and administrators to inform evaluation and professional development practices for secondary teachers.

Research Questions

The study sought to answer the following research questions:

1. What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

2. How do the panel of master teachers, administrators and teacher trainers rank the importance of the elements of SETs?

3. What do a panel of master teachers, administrators, and teacher trainers identify as strategies for using the data from SETs to inform evaluation and professional development for secondary teachers?

Significance

In an environment of ever-increasing scrutiny through standardized testing and digital reporting options, teachers are being held accountable for gains in student learning. At the same time, current evaluation practices show little ability in identifying

15

good or bad performance or providing information on professional development needs (Weisberg, Sexton, Mulhern, & Keeling, 2009). In contrast, SETs offer a valuable and cost-effective way to provide information about teachers' practice and behavior, and this information can be used to provide more targeted and effective professional development (Desimone, 2011). Especially for beginning teachers, whose classroom practices are still developing, there is an urgent need for the kind of feedback and subsequent professional development experiences that SETs can provide (Chalwa & Thurkal, 2011).

Unfortunately, the use of SETs at the secondary level is still limited, despite their proven efficacy (Hanover Research, 2013). This is in part because of the perceived limitations in secondary students' ability to provide useful and unbiased information (Jezequel, 2008). Although a growing body of research showing that secondary students can and want to give teachers feedback, teachers are still reluctant to use students' evaluations of their teaching practices to inform their work (Elbow & Boice, 1992). Chalwa (2011) reports that student feedback can be effectively used to improve the performance and general teaching competence of student teachers, while other researchers assert the validity and reliability of the perceptions of secondary students (Helding & Frasier, 2013; Jezequel, 2008).

Further compounding the problem is the lack of consensus about the content and use of SETs. Where some researchers favor holistic scoring systems (CRLT, 2014), others call for the use of multiple categories (Algozzine et al., 2004; Jezequel, 2008; Thorne, 1980). Among those advocating for the use of SETs in informing teacher development, some want the results to be distributed to individuals for reflection (Elbow & Boice, 1992; Jezequel, 2008). In addition, the use of SET data in making system-wide

16

professional development decisions is also encouraged (Darling-Hammond et al., 2012).

Finally, the frequency of such evaluations is under debate (Pallas, 2011; Ramsdell, 2011).

While the effectiveness of SETs has been proven, still in question at this time are the

content of SETs, the uses of the results, and the frequency of their administration.

A study that investigates both the content and the implementation of SETs in the

secondary school context is needed because of the dearth of information on how they can

be used in a professional development model to aid new and experienced teachers

improve classroom practice (Jezequel, 2008). Current evaluation practices provide little

useful feedback for teachers (TFEE, 2012). The lack of uniformity in existing SETs

confirms that there is still debate regarding their content, and clear guidelines regarding

the forum and form for disseminating survey results shows how little is known about

their usage. Even less is known about how the use of SETs changes classroom practices

over time. All of these factors contribute to the need for a Delphi panel discussion of the

construction and use of SETs in professional development initiatives at the secondary

level.

Definition of Terms

Theoretical Definitions.

For the purposes of this research, understandings of the following theoretical

definitions for reference are below:

- *Andragogy.* "A set of core adult learning principles that apply to all adult

  learning situations" (Knowles, 2005, p. 2)

- *Delphi Technique.* "A widely used and accepted method for gathering data from

  respondents within their domain of expertise. The technique is designed as a

group communication process which aims to achieve a convergence of opinion on a specific real-world issue" (Hsu, 2007).

- *Student Evaluation of Teachers (SETs).* "A subjective form which can be quantitative or qualitative in nature, sometimes a combination of the two, that students will independently and anonymously fill out, assessing their teachers' performance and effectiveness. Depending on the makeup of the SETs form, students will sometimes self-assess their own learning outcomes in the class on this evaluation. This form is most often completed at the end of a course, although there are instances where SETs are distributed at the midpoint of a semester, and again at the conclusion of the course" (Jezequel, 6).

- *Self-Directed Learning*. Learning that is characterized by "free choice of subject matter and free choice in determining outcomes" and "individual, critical thinking" (Knowles, 2005, p. 43).

Operational Definitions

For the purpose of this research, operational definitions of major variables and best practice terms are described below:

- *Administrator.* For the purposes of this study, *administrator* is defined as a principal, assistant principal, or learning director with responsibility for conducting evaluations and professional development for secondary instructors.

- *Professional Developm*ent. For the purposes of this study, *professional development* is defined as "the advancement of skills or expertise to succeed

in a particular profession, esp. through continued education"

(Dictionary.com, 2014)

- *School Climate*. For the purposes of this study, *school climate* is defined as "The emotional and social aspects of school environment. A measure of the quality of school climate is students' feelings of safety and connectedness to their school" (Connecticut State Department of Education, 2007).

- *Secondary School*. For the purposes of this study, *secondary school* is defined as "A school intermediate between elementary school and college and usually offering general, technical, vocational, or college-preparatory courses" (Merriam-Webster, 2014).

- *Teacher Training Institution*. For the purposes of this study, a *teacher training institution* is defined as a higher education institution that specializes in the training of new teachers.

- *Veteran Teacher*. For the purposes of this study, a teacher with over five years of experience will be considered a veteran teacher.

Delimitations

The following is the delimitation of this study:

1. The participants of this study were delimited to teachers, teacher trainers, and administrators working in secondary schools or in training institutions for secondary school teachers in California meeting specific criteria for inclusion (see Table 1).

Organization of the Study

The study is organized into five chapters, a bibliography, and associated appendices. Chapter 2 focuses on reviewing the available literature related to the content

and use of SETs. In chapter 3, the methodology and design of the study are outlined, as well as the instruments used to gather data and the composition of the study panel. Chapter 4 features a presentation and analysis of study findings. Chapter 5 comprises a summary of the study, resulting conclusions, and recommendations for further study. Following that are the bibliography and appendices.

CHAPTER II: REVIEW OF THE LITERATURE

Chapter II of this study reviews the professional literature and research related to student evaluation of teachers (SET) at both the tertiary and secondary levels, including their historical and current usages. It also focuses on past and present professional development (PD) models, including the use of student data in making decisions about PD practices and content. Theories concerning andragogy (adult learning) are applied to the use of SETs in determining and facilitating PD practices. Finally, research related to the benefits of eliciting student voice is presented.

The History and Principles of Andragogy in Learning Theory

The history of pedagogy stretches back millennia. From arguments about the use of Socratic seminars to discussion of the practices of Socrates himself, mankind has always engaged in spirited debate concerning the best ways to pass on learning to successive generations. In particular, the past fifty years have seen fundamental changes in the way that this debate has been framed, with one major point of contention concerning the nature of learners themselves.

In its traditional sense, pedagogy concerns how all humans learn. The word itself first appeared in the English language in 1623, and, until the last century, the primary meaning of "the art, occupation, or practice of teaching" remained constant ("Oxford English Dictionary," 2016, para. 3). Etymology, however, tells a different story, with the sixteenth-century French word *pédagogie* coming from the Latin word paidagogia, or "education, attendance on boys" ("Etymology online," 2016, para. 1). Until recently, the main theories of instruction grouped all humans in the same category and assumed that what was effective in teaching children would also serve to educate adult learners.

In the 1970s, Malcolm Knowles began to question this assumption. He, along with American researcher R. M. Smith and British researcher Peter Jarvis, theorized that humans of various stages of development learn in different ways, and that what works for one age group would not necessarily work well for another (Zemyov, 1998). Pedagogy, according to Knowles, focuses primarily upon the material to be learned and the attitudes and actions of the instructor, and in doing so, it ignores what the student brings to the learning environment in favor of content that has been predetermined (Forrest & Peterson, 2006). In a worldwide survey of adult education practices, Zemyov (1998) found that continuing to use the principles of pedagogy with adult learners results in poor efficiency. A new paradigm was needed, one focusing on the unique factors that concern adult learners. This area field of study was named andragogy, "the art and science of teaching adults" (Forrest & Peterson, 2006, p. 114).

One major difference between andragogy and traditional pedagogy lies in the role of the learner. Traditional pedagogical methods, developed in the 7th century as a means of preparing young boys for the priesthood, had students playing the role of passive recipients of an established curriculum. In this conception of the education process, the teacher is the sole determiner of the subject, method, timing, and assessment of learning (Knowles et al., 2005; Rada & Knowles, 1980). His job is to transmit knowledge to the students, whose minds serve as , in John Locke's terms, a 'tabula rasa' to be inscribed through direct instruction ("Pioneers in our field," 2016). While centuries of scholarship have refined and redefined this view and have provided for more agency on the part of students in the learning process, the idea of a top-down curriculum to be developed and administered by teachers to students is still more often the rule than the exception, even

when dealing with adult learners (Webster-Wright, 2009). As Newton suggests, despite decades of advancement in pedagogy, educators still have trouble recognizing that an adult is not merely an oversized child (Newton, 1977). Andragogy challenges this view, putting the focus squarely on the adult learner, with the ultimate goal of instruction being the development of self-sufficient, adaptive learners engaged in free inquiry (Forrest & Peterson, 2006).

Fundamental to this development is the acknowledgement of what an adult learner brings to the educational setting. Knowles proposes five principles about the adult learner: they are self-directed; they come to the educational setting with a wealth of prior experience; they learn in response to real-world problems; they come motivated to learn; and this motivation is internal (Jensen, Sonnemann, Roberts-Hall, & Hunter, 2016). In this model, the adult learner is a problem-solving, self-directing repository of accumulated experience who learns in order to fulfill his societal role (Forrest & Peterson, 2006; Newton, 1977).

The first principle, that learners are self-directed, contends that adults need to feel in   control of their own learning. As people mature, they become less dependent on others for orchestrating their own learning (Zemyov, 1998). Newton (1977) asserts that this realization of independence is what defines adulthood, while Gehring (2000) suggests that meaning in life is found in the goals that humans set for themselves and that a person achieves adulthood to the extent that he perceives himself and is perceived by others to be self-directing. Because man's deepest need is to be treated as a self-directing individual, one deserving of respect, the theory of andragogy argues for a change in the dynamics of the learning environment (Weingand, 1996). The relationship between

teacher and student should be one of guide and traveler, with the teacher giving direction but encouraging adult learners to make use of their own experiences as they explore new ideas (Forrest & Peterson, 2006).

The second principle of andragogy involves what adults bring to an educational setting. They come to learning and development with a lifetime of accumulated experience as parents, spouses, workers, and students, and this experience needs to be taken into consideration when planning how to help them develop (Forrest & Peterson, 2006). According to both Lev Vygotsky and John Dewey, attempting to use an atomistic approach to instruction, one that tries to isolate the content to be imparted from the learner and the learning context, fails to acknowledge the complexity of the whole environment (Webster-Wright, 2009). Designers of development programs for adults need to recognize learners both as recipients of instruction and as valuable assets to be exploited as they incorporate the variety of different viewpoints, life stages, and values they embody (Forrest & Peterson, 2006; Zemke & Zemke, 1984).

A third difference between the child and the adult learner lies in their reasons for being in the classroom. Children study something because the teacher tells them to, while adults learn what they feel they need to know (Boulton-Lewis, Wilss, & Mutch, 1996). Adults seek out learning because they desire to be empowered to solve real-world problems and become more effective in their various societal roles, and their enthusiasm for a learning activity will be in direct proportion to how efficacious they perceive it to be for their own development (Darling-Hammond, Wise, & Pease, 1983; Zemke & Zemke, 1984). Coming to the classroom with their problem-solving mindset, adults seek out information and skills that they can apply immediately to making improvements in their

work (Forrest & Peterson, 2006).   They also prefer learning through collaboration and activity rather than passive reception (Battey & Franke, 2008; Jensen et al., 2016). Those designing learning experiences for adults must incorporate these social and pragmatic tendencies into their planning if they want engaged and motivated participants (Fogerty & Pete, 2009).

The final two principles of andragogy involve adults' motivations for learning: they come to an educational situation already wanting to learn, and this motivation is largely internal. Where children often require external motivation to increase their enthusiasm for study, factors such as raising self-esteem, increasing job satisfaction and performance, and improving one's quality of life all contribute to an adult's motivation to engage in learning (Boulton-Lewis et al., 1996). The adult focus on the real-world application of learning also shifts the temporal perspective from a delayed need for knowledge to the desire for immediate utilization (Zemyov, 1998). Adding to the complexity is the fact that because the motivation for adults is primarily internal, involving them in decision-making about their learning will increase their investment in the process (Zemke & Zemke, 1984). When they are made to feel that proposed changes in behavior are suggestions for consideration rather than rules for action, the resulting feelings of empowerment are a spur towards compliance with and acceptance of them (Darling-Hammond et al., 2012).

A Brief History of the Use of Student Evaluation of Teachers in Education

The process of surveying students about their instructors' performance has been conducted in the US from the late 19th century, with students from Iowa first providing input on the effectiveness of their instructors (Hanover Research, 2013). By the 1920's,

formal evaluation of faculty by students was becoming more commonplace in American institutions of higher learning, including published student ratings of teachers at Harvard and Purdue (Mertler, 1999). By 1960, forty percent of such institutions were having students evaluate their instructors (McKeachie & McKeachie, 1957; Rodin & Rodin, 1972). Currently, SETs are in use in a majority of higher level institutions in the United Stated (Schmelkin, et al., 1997).

At the tertiary level, surveys have three main purposes: as a tool to provide feedback on instructional practices; as a factor in making personnel decisions, including those involving promotion and tenure; and as a guide for students as they choose their courses (Schmelkin et al., 1997). Temple University, for example, ensures the quality of its instructors by having each evaluated every semester in order to help instructors evaluate the effectiveness of their instructional practices and materials, and provide data for administrators and instructions in matters of promotion, merit, and tenure ("Temple," 2016). Of the purposes of using SETs, it is the evaluative one that has received the most attention from and raises the most concerns for educators (Algozzine et al., 2004; Elbow & Boice, 1992; Schmelkin et al., 1997; White, 1976).

The use of student surveys for evaluative purposes is seen as suspect by university instructors for a variety of reasons. Primary among these are the perceptions that students are too immature and too ignorant about pedagogy, that they tend to respond positively to a teacher who is entertaining, and that inter-student rating reliability is low (Costin, Greenough, & Menges, 1971; Elbow & Boice, 1992; Schmelkin et al., 1997). Another potential problem involves the halo effect, through which students give higher ratings to those educators giving higher grades (Costin, Greenough, & Menges, 1971).

26

Finally, Schmelkin et al. (1997) cited other concerns regarding SETs:

[They] are affected by various extraneous factors including course characteristics (e.g., class size, subject matter, level of course, whether it is a required course or not, time of day, if first-time course is being taught, if innovations are introduced), instructor characteristics (e.g., sex, rank, grading pattern, personality), and student characteristics (e.g., age, sex, student level, major/nonmajor, interest in course). (p. 576)

Given all the factors that are seen to influence students' responses to SETs, the often-cited resistance to them is reasonable, but ill-founded (Elbow & Boice, 1992; Schmelkin et al., 1997). While the reality of SETs might be quite different from the perceptions of those tertiary instructors who object to them, these concerns do raise two important questions: Are the results of SETS reliable and valid, and can they be used effectively for another of their main purposes, as tools for teacher development?

Perceptions of SETS—Validity and Reliability

In terms of face validity, the research is divided on whether instructors give credence to the views of students expressed in SETs. That instructors hold the results of SETs to be suspect is often cited in studies (Ferguson, 2012; Little, Goe, & Bell, 2009; McKeachie & McKeachie, 1957; Towe, 2012; Youngs, 2013). At the same time, one study targeting this specific phenomenon found that instances of instructor resistance were largely anecdotal and that there was actually little resistance to the use of SETs in summative and formative evaluations (Schmelkin et al., 1997).

Despite anecdotal evidence of instructors' reservations about the reliability and validity of SETs at the tertiary level, studies have found them to be a valid indicator of

teacher performance in the classroom. Regarding the halo effect, the perception that teachers who give better grades get better evaluations, Scheurich (1983) found that course marks do not significantly influence the ratings teachers receive. In Costin's assessment, any correlations between higher evaluations and course grades result from students with higher grades having more interest in the course, not from any halo effect (Costin, 1971). Another study concluded SETs to be highly effective as a means for collecting data for both formative and summative evaluations (Johnson, 2012). For this last statement to be true, however, SETs must be, in fact, reliable and valid.

In response to reservations about the use of SETs in evaluations due to perceived problems with validity (Darling-Hammond et al., 2012), numerous studies have shown that SETs can provide educators and administrators with consistent and accurate information about what is happening in a classroom (Rockoff & Speroni, 2010). The MET Project (2012), found data obtained from SETs to be more reliable than other measures, such as either administrator evaluations and value-added measures alone. This was in part because students experience a teacher's work over months and are not evaluating individual (and potentially variable) lessons, and because SET scores are averaged over entire groups of students rather than relying on a few observers. In a review of research, Dillon (2010) found that teachers who scored high in student ratings on maintaining order, focusing instruction, and providing effective remediation also had significant gains in their students' scores on standardized tests. According to Schmelkin (1997), averaged student ratings offer stable, reliable, and multidimensional assessments of a teacher's work in the classroom, and these assessments target the teacher, not the course taught, in a manner relatively unbiased by the hypothesized variables. This

assumes, of course, that the SETs themselves are asking students the right questions; exactly what should be elicited from students is another question that has yet to be resolved.

The Content of SETs

While the reliability and validity of SETs in helping to determine teacher effectiveness has been generally accepted by researchers and educators, the content of SETs themselves is still in dispute. If the purpose of using SETs in the classroom is to inform personnel decisions, then more global questions (e.g. "Overall, this is a good instructor.") might be used, as higher scores on such questions have a high correlation with student learning (CRLT, 2014). At the same time, Algozzine argues that a single score cannot encompass all that comprises effective teaching (Algozzine et al., 2004). In particular, general or holistic questions do not provide specific details about what is being done poorly or well, or what the evaluator was looking for when conducting the evaluation (Elbow & Boice, 1992). If the purpose of SET implementation is to help teachers and administrators make decisions about training and classroom practice, assessments targeting multiple and more specific traits would be more useful (Darling-Hammond et al., 2012; McKeachie & McKeachie, 1957; Schmelkin et al., 1997). As a tool for informing teachers' understanding of their work, SETs containing a variety of questions targeting the multidimensional aspects of classroom activity appear to be preferred (Jezequel, 2008; Thorne, 1980; Youngs, 2013).

When teaching is acknowledged to be a multidimensional activity, the discussion of what to include in SETs centers around which dimensions can and should be explored in order to improve instruction. Darling-Hammond (2012) posits that teaching can be

seen as both labor and art, with each conception resulting in different aspects of classroom practice being investigated. In the former view, teacher activity should be organized into rational, programmatically uniform routines that can be taught and evaluated by administrators. As long as the proper routines are established, proper learning will result. The latter view, however, holds that teaching effectively requires more than just opening a toolbox of techniques. Instead, teachers must use sound professional judgment as they apply a repertoire of procedures in a dynamic and ever-changing environment, one in which many factors are beyond their immediate control (Darling-Hammond et al., 2012). Seen in this light, SETs could focus on both established routines and a teacher's ongoing responses to the classroom environment. In the MET survey, for example, some of the instrument's categories of questions deal with classroom routines (Control, Clarify, Consolidate), while others elicit responses about affective factors (Care, Captivate, Confer), with the result that a more complete picture of a teacher's craft in the classroom is considered (Ferguson, 2010). Both the logistics of classroom activity (amount of work, quality/quantity of teacher feedback, topics covered) and the environment that the teacher establishes (quality of activities/discussions, teacher accessibility) could be explored in surveys (Elbow & Boice, 1992). They could also elicit information on student perceptions of teacher content knowledge, enthusiasm, and preparedness (Schmelkin et al., 1997).

Also still in question are frequency of administration of SETs and the manner in which data obtained from them are used. Pallas argues for frequent administration of evaluations for beginning teachers, as their performance can change greatly as they grow in their practice (Pallas, 2011). Both Jezequel (2008) and Weisberg, Sexton, Mulhern, &

Keeling (2009) call for them to be used often, with detailed feedback being given to teachers, while the MET project calls for various measures over many years to allay teachers' fears about the high-stakes nature of one-off evaluations and to provide a more complete picture of classroom practice for administrators (Ramsdell, 2011). For Jezequel (2008), surveys are most effective when used mainly for reflective purposes. In contrast, Darling-Hammond et al. (2012) call for evaluations to include feedback for teachers and to be used to inform the professional development that they are given.

Despite a growing appreciation of the value of eliciting information about a teacher's work through surveys (Desimone, 2011), questions around the content and use of SETs at the secondary level remain unanswered. Although SETs have proven effective and reliable in providing effective feedback for instructors at the tertiary level , other forms of evaluation are more common at the secondary one (Hanover Research, 2013; Kane & Cantrell, 2010).

Common Evaluation Practices at the Secondary Level

According to current California Education Code, each district is charged with determining evaluation procedures for its teachers, and these can vary both between districts and between individual campuses within them ("California Education Code," 2005). The frequency of these evaluations can depend on years of experience, with probationary teachers undergoing evaluation every year, teachers with permanent status every two years, and highly qualified teachers (as defined by the No Child Left Behind Act) with ten years in a district every five years. Each teacher's performance is evaluated in terms of expected student achievement on established standards, instructional strategies, compliance with curricular objectives, and the establishment of a suitable

learning environment. Beyond that, each district has leeway in the content, frequency, and application of evaluations ("California Education Code," 2005).

This leeway in frequency that districts have in evaluation practices can affect the effectiveness of professional development practices. According to the School Staffing Report, for the 2010-2011 school year, 95% of the 9,400 K-12 educational institutions reported that at least 80% of their teachers were NCLB compliant, and 72.7% claimed 100% compliance with NCLB from their teachers (California Department of Education, 2010). While the number of teachers within these percentages having ten or more years of experience was not reported, that still leaves a large group of teachers with the potential to undergo formal evaluation only once every five years. If one of the purposes of formal evaluation is to provide teachers with feedback geared towards professional development (Darling-Hammond et al., 2012; "California Education Code," 2005), current California education code leaves the potential for teachers to receive this feedback once every five years. This also assumes that the feedback is effective.

Current evaluation practices at the secondary level have been described as "haphazard" (Hibler & Snyder, 2015, p. 42) and centered on enforcing policies rather than improving instruction (Webb, 1995), with the primary tool used being infrequent administrative observation of teacher practices. According to California's Task Force for Educator Excellence (TFEE, 2012), current evaluation systems fail to provide either instructors or administrators the necessary support and feedback to enhance instruction or properly inform decisions regarding personnel. At the same time, the consequences of these haphazard evaluations can be extreme, as at least nine states require that evaluations be used as a factor in tenure decisions (Lacireno-Paquet, Bocala, & Bailey, 2016).

With so much at stake, this system of teacher evaluation often comes under fire from educators and researchers for many reasons. First of all, the instruments used to determine a teacher's effectiveness rarely discriminate between levels of ability among teachers. A US Department of Education study found that 97% of teachers surveyed for the 2012-13 school year received either an 'effective' (60.3%) or 'satisfactory' (36.8%) rating (Lacireno-Paquet, Bocala, & Bailey, 2016). In another survey, over 99% of teachers received a 'satisfactory' rating during administrative evaluations, and, as former chancellor of DC schools Michelle Rhee points out, while 95% of all DC schoolteachers received good evaluations, less than 10% of eighth graders scored at or above grade level in math (Hibler & Snyder, 2015).

Beyond this discrepancy between teacher ratings and student achievement, when all teachers are grouped in the same category, outstanding teachers are not identified, with the consequence that administrators lack the necessary data to inform retention decisions or identify high-quality teachers that could help their colleagues to develop and improve (Weisberg et al., 2009). According to the author of the Measures of Effective Teaching (MET) study, Thomas Kane, we are coming to the understanding that teacher evaluation systems are broken, with little benefit coming from rating 98% of all teachers as 'satisfactory' (Kane & Staiger, 2012). In other words, because schools are unable to identify excellence or support the improvement of middle-range teachers, districts end up seeing and treating all teachers as equal in terms of both capability and need for professional development (Lacireno-Paquet et al., 2016; Weisberg et al., 2009).

Second, what is being addressed in administrative evaluations does little to help teachers because better practitioners of their craft. Fielding describes a system where

administrators mark boxes on a sheet, rating items that rarely mean anything of value to all involved (Fielding, 2004). Instead, the evaluation process concentrates on executing prescribed policies and maintaining properly-timed paper trails, rather than valuing the interactions and slowly-nurtured relationships between teachers and students (Webb, 1995). As a result, teachers are held to a discredited and punitive system of accountability of dubious rigor (Fielding, 2004) that ignores non-technical variables that can have a profound effect on teaching (Darling-Hammond et al., 2012). Youngs (2013) cites the lack of variability in teacher evaluations as a problem, along with the fact that the instruments used rarely address factors such as content and pedagogical knowledge and their effects on how students learn, with the result that administrative observations have little long-term effect on classroom practices. In one study, fewer than half of teachers evaluated in their first four years in the classroom had any development areas targeted, and neither proactive nor regular feedback was given outside of the evaluation process (Weisberg et al., 2009).

Further compounding the difficulties in effectively evaluating teachers through classroom observations is the fact that they are often done by a principal who, due to staff size, rarely observes the teachers being evaluated. Such a principal might see the teacher four times a year and have to rely on outside sources for information about the teacher in question (Hibler & Snyder, 2015). These conditions are in part due to principals' need for a process that combines efficiency with maintaining staff morale and that is objective, time-effective, and feasible within the confines of their organizational structure (Darling-Hammond, Wise, & Pease, 1983). These tensions for principals (i.e., their need to promote both teacher development and staff morale) can lead to artificially-inflated

assessments of teacher competency, with the result that the ability of evaluations to determine the effectiveness of teaching is weakened by a principal's need to motivate his staff (Pallas, 2011). These potentially conflicting demands have led to a superficial system of evaluation based on oversimplified criteria that does little to determine what teachers need to do in order to refine their craft (Milanowski, 2004).

An additional criticism of administrator observations at the secondary level involves the qualifications of the person doing the observation. High school administrators work with teachers from a variety of subjects, and the observer can lack the subject-specific knowledge necessary to effectively evaluate a single-subject classroom, particularly when it concerns disciplinary-specific concepts and knowledge (Pallas, 2011). Goe, Bell, & Little (2008) argue that a deep assessment of content knowledge might be better conducted by a content expert or peer than an administrator, who may lack the specialized knowledge for such an evaluation, while also noting that in their study, less than ten percent of administrators made mention of the training of evaluators as a part of their procedures for teacher evaluation (Goe et al., 2008). In particular, a survey of California principals found them to be less likely to engage with teachers on issues such as classroom practices, curriculum development, professional development, or data usage, partially due to limited training and a lack of support for principals in California in these areas (TFEE, 2012). Secondary administrators, then, might be asked to evaluate the teaching in a classroom for which they lack both the necessary subject-matter knowledge and instrument training in order to do so effectively. In response to this, the Task Force for Educator Excellence recommends that any team of

evaluators should include subject specific experts, especially when tenure or renewal

decisions are being affected (TFEE, 2012).

Finally, observation criteria often do not take into account that teachers alter their

classroom conduct to fit the given context, particularly regarding distinct teacher actions,

and the behaviors an administrator notes in a few observations may not be indicative of a

teacher's usual classroom practice (Darling-Hammond et al., 1983).

Administrator evaluation, especially when used as the sole means of determining

the effectiveness of secondary teachers, can be a system that fails to differentiate ability

and knowledge among teachers (Youngs, 2013), provides little support for individual

professional development (Dresel & Rindermann, 2011; Milanowski, 2004), and is

conducted by an administrator who rarely sees the teacher in action and can be

undertrained and lacking the content knowledge necessary to provide effective feedback

(Hibler & Snyder, 2015; Mertler, 1999). One response to these conditions, the use of

SETs as a component of the evaluation process, has become more prevalent in the last

decade.

The Use of SETs in Determining Teacher Effectiveness of Secondary Teachers

Although SETs have been used in formative and summative evaluations at the

tertiary level for over a century, their use in the K-12 setting has until recently been the

exception to the rule (Hanover Research, 2013), with only a few districts using them for

evaluating programs or instructors (Dillon, 2010; Johnson, 2012; Stecher, Garet,

Holtzman, & Hamilton, 2012). This is in part due to the resistance of teachers and

teacher unions. As in colleges and universities, there are instructors at the secondary

level that view SETs as little more than popularity contests and an encouragement to give

higher grades, this despite the fact that SETs are only infrequently used in matters of retention or salary decisions at the secondary level (Jezequel, 2008). Still, there is a recent and growing trend toward using SETs as a factor in teacher evaluation, with districts in Illinois, Pennsylvania, Arizona, and Georgia planning some form of implementation in the near future (Hanover Research, 2013), and Massachusetts requiring the collection of student evaluation data from 2013 onward (Massachusetts Department of Elementary & Secondary Education, 2013).

While SETs are not currently being used on a large-scale basis in California (Hanover Research, 2013), another student survey has been in use for decades. In California, all students in the seventh, ninth, and eleventh grades are regularly administered the Healthy Kids Survey biannually, and while this survey deals with school environment rather than the work of individual teachers, the processes for giving surveys to California middle- and high-school students are already in place statewide ("Healthy Schools," 2016).

A prominent example of the current trend toward using SETs in teacher evaluation is seen in the work of Measures of Effective Teaching (MET) study, funded by the Bill and Melinda Gates Foundation and begun in 2008. This study, based on the work of Ron Ferguson of Harvard University, sought to "improve the quality of information about teaching effectiveness available to education professionals… [and] help them build fair and reliable systems for measuring teacher effectiveness that can be used for a variety of purposes, including feedback, development, and continuous improvement" (Kane & Cantrell, 2010, p. 2).

Over the course of five years, the MET project's goal was establish a useful and reliable system for providing input to teachers on improving their practice and to administrators when making decisions about personnel through the triangulation of data obtained using a combination of teacher evaluation, analysis of value-added measures (VAMs), and student surveys (Kane & Cantrell, 2010). The study found that each of these measures, used separately, only provided a part of the picture. For example, in an analysis of MET project data, Rothstein found that VAMs were not effective in controlling for the influence of the differences in students that teachers have in their classes each year (Rothstein, 2011). Another study found that only adding student test data in evaluations also lowered teacher satisfaction with the process by a factor of 250%. (Lacireno-Paquet et al., 2016). However, the MET project found that combining the data from all three sources (observations, VAMs, and SETs) allowed them to rank teachers in a statistically significant way and to provide them with targeted feedback designed to improve teaching practice. According to the report's authors, the use of both VAMs and SETs make it possible to give teacher who are interested in improving their practice feedback that is both targeted and indicative of their current level of effectiveness (Kane & Cantrell, 2010). Furthermore, by combining multiple measures of evaluation, school leaders gain a more thorough measure of a teacher's practice, giving them insight into a teacher's effectiveness and allowing them to furnish targeted diagnostic feedback (Hattie, 2012; Jezequel, 2008; Kane & Cantrell, 2010). Regarding SETs themselves, Ferguson found that if SETs are well constructed, they can provide administrators with effective direction in choosing professional development foci and conduction evaluations of a teacher's work in the classroom (Ferguson, 2012). However, because the main goal of

the MET Project was to devise a way to determine each teacher's effectiveness in terms of student learning data, the project stops short of prescribing how VAMs, observation, and SET data could be used in informing professional development (PD) for individual teachers, departments, and schools. This is in part because PD processes in the US are far from uniform and often potentially ineffective.

The Current State of Professional Development in the US

Professional development in the US is a multi-billion-dollar industry, with federal and state governments providing large sums of money to districts in order to facilitate better instruction. The exact amount is difficult to determine because of a lack of uniformity in accounting codes, the use of differing frameworks for accounting, and the fact that expenditures can be supplemented by individual school sites, whose contributions are not included in many districts' assessments (Odden, Archibald, Fermanich, & Gallagher, 2002). Still, a number of researchers' estimates place the annual expenditure from $3,800 to $6,900 a year (Odden, Archibald, Fermanich, & Gallagher, 2002) up to Sawchuk's (2010) calculation of between $6,000 to $8,000 a year per teacher. California comes in at the high end of the scale, with a study by Little et al. placing the annual expenditure for professional development at $6,973 per teacher in 2000 dollars (Odden et al., 2002). What is also unclear, however, is just how efficiently and effectively these funds are being used.

Odden defines effective PD as "professional development that produces change in teachers' classroom-based instructional practice, which can be linked to improvements in student learning" (Odden et al., 2002, p. 53). In 2001, the National Staff Development Council, in conjunction with forty professional learning associations, promulgated a set

of staff development standards, outlining how to determine and ensure the effectiveness of professional development programs (TFEE, 2012). By 2007, they had been adopted by twenty-five states (Hirsh, 2007). Based on these standards, the Association for Supervision and Curriculum Development (ACSD) outline what is needed in order to make PD successful at the secondary level. According to their standards, effective professional development has the following characteristics:

• Directly focused on helping to achieve student learning goals and supporting student learning needs

• A collaborative endeavor- teachers and administrators work together in planning and implementation

• School-based and job-embedded

• A long-term commitment

• Differentiated

• Tied to the district goals. (Association for Supervision and Curriculum Development [ASCD], 2005, p. 1)

Unfortunately, current professional development practices fail to meet these criteria for effectiveness.

Concerns about Professional Development at the Secondary Level

Unfortunately, much of what constitutes professional development at the secondary level fails to fulfill these criteria. In the six criteria established by the ASCD, much of what constitutes current PD practices falls short of prescribed norms. Each criterion will be addressed individually.

A concentration on reaching student learning goals and supporting their needs

Typical PD programs rely on content that fails to address the specific learning goals and needs of teachers and students. As a result, they have little impact on what teachers do and how students learn in the classroom (Odden et al., 2002). According to Jensen, teachers find much professional development done in districts to vary in quality, be unsuited to their needs, and not connected to their teaching. High-performing systems, however, concentrate on using PD practices that evidence has shown to have significant effects on a teacher effectiveness and student improvement (Jensen, Sonnemann, Roberts-Hall, & Hunter, 2016). The content of PD also needs to be tied to long-term programs that recognize that teaching and learning are not discrete episodes, but rather a complex and long-term process (Ball & Cohen, 1999; Odden et al., 2002; Webster-Wright, 2009). At the same time, learning opportunities need to be linked to the specific weaknesses of the teachers for which they are provided (Hill, 2009). In order to do this, however, those that determine what comprises PD in a district need access to data on what teachers need to know, and SETs can play a vital role in those determinations (Ball & Cohen, 1999; Kelleher, 2003).

Beyond their current role in informing teacher evaluation, SETs can be used to target and monitor professional development goals and results at the district, school, and classroom levels (Burniske & Meibaum, 2012). Failing to do so leaves system leaders unable to assess how effectively their professional development dollars are being spent (Archer, Kerr, & Pianta, 2014). One 2009 study of teacher evaluations found that 73 percent of teachers surveyed were not given any specific feedback concerning developmental needs, and of those that did, only 45 percent reported getting useful

support to improve (Weisberg et al., 2009). Administrators need the means to determine how PD dollars are being spent and if this spending is aligned to the institution's strategic goals for improving instruction (Sawchuk, 2010). To this end, the Commission on Effective Teachers and Teaching calls for PD that is of high quality, student centered, and selected based on needs identified in evaluations and assessments (Commission on Effective Teachers and Teaching [CETT], 2012). SETs can be one form of evaluation in this process.

Collaboration between teachers and administrators

There is some disagreement concerning whether teachers or administrators should be determining the focus and content of PD. On the teachers' side, Zimmerman notes that excluding teachers from the decision-making process regarding PD is one of the factors that keeps them in a blue-collar mentality and deters their efforts at professionalism (Zimmerman & Jackson-May, 2003). Webster-Wright argues that this transmission model, where teachers who lack knowledge are trained by a presenter who possesses it, thus undervaluing the teachers knowledge of local contexts (Webster-Wright, 2009). Lester argues that secondary teachers will give credence to PD efforts about which they feel that their opinions and desires are being considered (Lester, 2003), while in a survey of studies on PD models, Guskey found that those programs deemed most effective contained an element of teacher discretion in content (Guskey & Yoon, 2009).

At the same time, Guskey also notes that site-based PD decisions lead to school staff paying lip service to research and tending to cleave to designs and research that confirm their existing practices. If given the choice of what PD to attend, teachers often

pick training that reinforces their existing knowledge rather than broadening it (Hill, 2009). As a result, the decentralizing of the process might actually serve to undermine the acquisition of new skills and knowledge rather than support it (Guskey & Yoon, 2009). While the question over who should determine the content of PD at the secondary level is still unanswered, Hill notes that instead of arguing about the value of teachers or districts making PD decisions, researchers should recognize that these decisions are rarely based on considerations of the specific deficiencies and needs of teachers " (Hill, 2009). Assessing those deficits and needs is one of the primary functions of SETs.

A focus on specific sites and jobs

Although teachers desire PD opportunities that are focused on the needs of their particular students and on improving their daily work, there is often a disconnect between what is presented to them off-site and what is being done on-site (CETT, 2012; Webster-Wright, 2009). Guskey corroborates the findings of the National Staff Development Council that potent PD results from adapting a 'best practices' model to the particular elements of the local context (Guskey & Yoon, 2009). The decontextualization resulting from PD experiences that are not tied to the needs and conditions of the particular participants reinforces a divide between what is taught in a course and what is actually done in the classroom; for at least a decade, research has shown that wholesale transfer of a discrete package is not possible, no matter how well-intentioned or -designed (Webster-Wright, 2009). When PD is based on the collaboration of teachers and school leaders working within local contexts, it increases the prospects that teachers will be more responsible for the continued learning of themselves and their colleagues (Jensen et al., 2016; Lester, 2003).

43

A long-term undertaking

Another concern with current PD practices is that they tend to be one-off events, built upon the assumption that learning can be fostered through processes with a defined beginning and end (Yoon, Duncan, Lee, Scarloss, & Shipley, 2007). Because of this, districts continue to engage in ineffective and isolated events, even though studies show that the most effective PD is "sustained, content-embedded, collegial and connected to practice; focused on student learning; and aligned with school improvement efforts," (TFEE, 2012, p. 16). For example, one review of studies on PD suggests that effective PD activities should last at least a semester and comprise at least twenty hours of contact time (Desimone, 2011). In other words, the best development for teachers is long-term and embedded in a supportive learning community (Webster-Wright, 2009), while much of what is done in the real world is disorganized and cursory, without a firm and sustained expectation of effective PD (Ball & Cohen, 1999).

One factor contributing to the short and isolated nature of PD is the fact that teachers have little incentive to engage in long-term professional learning. With states typically requiring only a few days of PD each year, and with much of that being poorly designed, teachers are discouraged from engaging in challenging and sustained learning (Hill, 2009). At the same time, Guskey's analysis of PD research shows that almost all of the studies showed a correlation between student learning and follow-up to PD activities that was sustained and structured (Guskey & Yoon, 2009). In contrast, Webster-Wright found that a lack of long-term commitment in PD programs led to professionals considering them almost worthless, with associated learnings quickly forgotten and the whole process being seen as merely complying with local requirements (Webster-Wright,

2009). In an environment where sound pedagogical knowledge is necessary in order to

effectively implement the Common Core Standards, there is a strong need for sustained

PD programs (Youngs, 2013).

Differentiation for the needs and strengths of participants

One common complaint about current PD programs is that they fail to take into

account the specific needs and strengths of individual teachers and, as a result, have little

lasting impact (Kane & Staiger, 2012). A typical "one size fits all" approach does not

yield improvements in classroom practice and student learning (TFEE, 2012). Darling-

Hammond notes that shifting the focus from individuals to groups in determining what

comprises professional learning leads to an atmosphere where participants involuntarily

comply with externally-initiated programs (Darling-Hammond et al., 2012). In contrast,

research shows that while PD should be focused on standards, it should also lead to

collaborative and sustained inquiry into problems of practice and should take into account

the various stages of participants' development, building logical pathways from initial

training to skillful practice (TFEE, 2012). It should also recognize that what teachers

need in PD can vary, depending on their professional experience, the results of their

evaluations, and the conditions of their particular schools, and that educators might be a

leader is some PD activities and a participant in other, depending on their level of

expertise in the given topic (CETT, 2012). In one study, the percentage of teachers

reporting seeing a link between teacher evaluations and PD opportunities ranged from

25% to 45% (Stecher, Garet, Holtzman, & Hamilton, 2012). Ironically, the type of

differentiation typically embraced when working with students according to their abilities

and knowledge levels in the classroom is less commonly seen when learning is applied to the teachers themselves (Hill, 2009).

Alignment with district goals

In identifying six structural features of effective programs, Odden et al. emphasizes the need to promote coherence by aligning PD to established standards and school and district goals. When mixed messages are received from various policy-making levels (district, state, federal), these contradictions can result in programs that are not coordinated with student learning objectives, resulting in PD that is overly varied, unproductive, and lacking proper focus. (Odden et al., 2002).

Much of what constitutes professional development fails to live up to each of the guidelines promulgated by the ASCD (2005). Correcting these failures becomes even more important in light of recent changes to funding procedures for California schools.

Local Control Funding Formula (LCFF)

Enacted during the 2013-14 school year, the Local Control Funding Formula established a new way of funding school programs in California (California Department of Education, 2016). This new system replaced a forty-year-old system of centralized decision-making regarding how resources were allocated and utilized in California schools. One fundamental change involves how decisions about the use of funds are made: under the LCFF, all stakeholders (teachers, parents, school personnel, pupils, parents, and bargaining units) have a voice in the allocation of funds from the state (School Services of California, Inc., 2016). In order for the stakeholders to make informed decisions regarding the allocation of funds for local programs, they need to conduct robust assessments of all aspects of the local learning environment and then

analyze this data to identify significant areas of need ("LCFF Overview," 2016). Data from SETs can inform these needs assessments in the areas of both academics and culture and climate in a rapid and timely way, ensuring that all stakeholders have a better understanding of local conditions when making decisions about funding professional development initiatives in a district.

Student Voice

A final factor when considering the use of SETs in informing the content and practice of professional development involves its effects on student voice. When students are given the opportunity to express their feelings and insights about what is being done in their classrooms, this can have broad repercussions on school culture, with improvements in both learning and teaching (Williams et al., 2012). Unfortunately, the potential benefits from eliciting students' views are largely being ignored in the current educational environment (Jackson, 2004; Jezequel, 2008; Quaglia & Corso, 2014). Still, in cases where student voices have been elicited and utilized, a number of benefits have accrued.

When students are given a voice, it encourages discussions about ways to improve the educational environment (Fielding, 2004). Although there may be some dissonance as long-held views are challenged (Williams et al., 2012), teachers can become more receptive to experimenting with new ideas (Cook-Sather, 2006), and they can gain a better and more immediate understanding of the effects of these experiments on students and their learning (Worrell & Dey, 2008). They begin to see students as active discussants rather than passive recipients of instruction (Fielding, 2004), and students begin to see that their opinions can make a difference in how and what teachers teach

(Mertler, 1999). One of the most prominent themes in studies of student voices is that their opinions help to develop a more complete understanding of what is happening in the classroom, including issues of pedagogy, factors helping or hindering learning, and school norms and policies (Jackson, 2004; Quaglia & Corso, 2014; Thiessen, 2006; Worrell & Dey, 2008).

The presence or lack of a forum for student voices also has a significant effect on the students themselves. Including the voices of secondary students democratizes them and instills in them civic responsibility (Fielding, 2004; Jezequel, 2008; Williams et al., 2012). Their opinions gain legitimacy (Jackson, 2004), with resulting higher levels of self-esteem (Worrell & Dey, 2008) and increased student motivation (Jezequel, 2008). When voices are not elicited, however, negative consequences can also result. Under-utilizing this intellectual capital (Fielding, 2004) leads to students feeling frustrated and detached regarding their education (Worrell & Dey, 2008). This is especially true of students whose learning path strays outside more traditional channels (e.g., vocational education or school-to-work programs), where students might already feel marginalized by mainstream policies and expectations (Cook-Sather, 2006; Fenwick, 2006). They, like other students, need to feel that their voices are being heard.

One final condition determining the efficacy of eliciting student voice is that the students need to feel that their voices are actually being attended to (McKeachie & McKeachie, 1957). Elbow and Boice (1992) argue that the process should be thoughtful and reflective rather than just mechanical. When students see that their opinions are actually being utilized in making decisions about teaching and learning, they become more positive about their educational experience (Mertler, 1999; Williams et al., 2012;

48

Worrell & Dey, 2008). The end result of this process—"students at the centre of the educational process; the main focus: the development of their strengths and talents; in open and interested learning environments, where everyone can experience a sense of personal worth and belonging to a community of people" (Gentile & Pisanu, 2014, p. 22).

Conclusions

Current theories of andragogy highlight the differences between how children and adults learn (Zemyov, 1998). Where children benefit from being guided through learning by their instructors, treating them as passive recipients of a prescribed curriculum, andragogy advances that the purpose of adult learning is to develop self-sufficient, adaptive learners engaged in free inquiry. It acknowledges the wealth of experience that adult learners bring to the learning situation, and it also understands their internal motivation to engage in study with personal and real-world application.

Unfortunately, the current systems of student evaluation of teachers and professional development at the secondary level in the US fail to take into account many of these factors, and one possible remedy for this situation is the use of SETs as a factor in teacher evaluation. In use for over a century at the tertiary level, SETs are still rare at the secondary one (Hanover Research, 2013). One reason for this is the perception that there is resistance to their usage, though evidence for this view is largely only anecdotal (Schmelkin et al., 1997). Despite concerns about the validity of data elicited from students (Darling-Hammond et al., 2012, Schmelkin et al., 1997), multiple studies have shown such data to be as valid and reliable as data about teacher performance acquired through other means (Costin et al., 1971; Johnson, 2012; Scheurich, Graham, & Drolette, 1983). Researchers at the MET Project have found that evaluations based on the

combination of administrator evaluations, SETS, and student test performance data have a high degree of validity and reliability concerning teacher effectiveness (MET Project, 2012).

In place of the use of SETs is an ineffective system of relying primarily on administrator evaluation of teachers (Hibler & Snyder, 2015). The system is flawed because it often relies on untrained administrators (Mertler, 1999) using instruments that fail to differentiate among the abilities and practices of teachers (Kane & Staiger, 2012; Youngs, 2013). It also has very little impact on what should be its primary function: professional development practices and programs for teachers (Stecher et al., 2012; TFEE, 2012).

Instead of being informed by teacher evaluations, many professional development programs are being run counter to the principles outlined by the ASCD (2005). These programs are rarely tied to specific and site-based student learning goals (CETT, 2012; Odden et al., 2002). They are seldom chosen in collaboration with participants (Zimmerman & Jackson-May, 2003). They fail to account for differences among the professional development needs of individual teachers (Kane & Staiger, 2012). Finally, they tend to be short-term programs, without vision for a long-term teacher development (Yoon et al., 2007).

Fortunately, many of the flaws of current professional development practices can be corrected through the use of SETs (Burniske & Meibaum, 2012; Jezequel, 2008). California schools already have the means to elicit data from students ("Healthy Schools," 2016), and the recent adoption of LCAP procedures makes the collection of

this data necessary and useful as districts decide where to allocate funds (School Services of California, Inc., 2016).

The use of SETs in teacher evaluation also accrues the benefit of increasing student voice in the educational process. This accrues three distinct benefits. First of all, giving students increased voice can result in greater student motivation (Jezequel, 2008). It can also lead to students having self-esteem (Worrell & Dey, 2008). Finally, giving students increased voice through the use of SETs can engender in them a greater sense of civic responsibility (Fielding, 2004; Williams et al., 2012).

At the present moment, there is increased interest in the use of SETs in teacher evaluations, with a few states adopting evaluation systems that in some way incorporate teacher performance data elicited from students, and others experimenting with pilot studies of their effectiveness (Hanover Research, 2013). California, however, is not among those few states. Recent changes in school funding in California, the most prominent being the adoption of the LCFF system for determining the use of funds at the local level, have created an important opportunity for ensuring that professional development funds are being used in a manner that best informs decisions about the content and form of the professional development of teachers. Unfortunately, although there is much agreement that SETs can provide useful information in determining what teachers need to improve in their teaching, there is little consensus about the content of those surveys or how information obtained from students should be used to inform professional development decisions. This study seeks to remedy that situation by conducting a Delphi study to elicit the opinions of experts on the subject.

This literature review chapter outlined research on andragogy, teacher evaluation and professional development systems, SETs, and student voice factors. Chapter III outlines the methodology to be used in this study. In chapter IV, the results of the Delphi study are presented, along with an analysis of its findings. Chapter V will feature a summary, findings, conclusions, and recommendations for further research.

CHAPTER III: METHODOLOGY

According to the literature review, current evaluation practices at the secondary level provide little data concerning teacher performance, with the result that schools have difficulty assessing classroom instruction or providing targeted professional development (Weisberg et al., 2009). Though commonly used as a means of judging teacher performance at the tertiary level, student evaluations of teachers (SETs) are still a relatively new phenomenon in high schools (Hanover Research, 2013). This study seeks to understand how they could be used to inform and improve professional development and evaluation practices.

Overview

This chapter comprises a description of the methodology of the study and a presentation of the procedures used to conduct it. It starts with the purpose statement and research questions and then continues with details of the research design. Included in the description of the methodology are information about the study population and sample, the instrument to be used, instrument validation through field tests, and the data-collection process. The chapter ends with an explanation of the data analysis procedures and a description of study limitations.

Purpose of the Study

The purpose of this Delphi study was to identify the most important elements for SETs (Student Evaluation of Teachers) at the high school level as perceived by a panel of expert master teachers, administrators, and teacher trainers. In addition, it was the purpose to determine how the results of SETs could best be used by teacher trainers and

administrators to inform evaluation and professional development practices for secondary teachers.

Research Questions

The following questions were investigated to address the purpose of the study:

1. What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

2. How do the panel of master teachers, administrators and teacher trainers rank the elements of SET?

3. What do a panel of master teachers, administrators, and teacher trainers identify as strategies for using the data from SETs to inform evaluation and professional development for secondary teachers?

Research Design

This study used a non-experimental design, one that investigates phenomena and relationships without directly manipulating conditions (McMillan & Schumacher, 2010). In particular, a survey research design involving prospective policy analysis was used, which entailed engaging in an iterative process of surveying experts in various fields about a proposal, with the feedback informing each successive round of surveys (Patton, 2002).

The Delphi technique was used to elicit data on the formulation and use of SETs. As is typical of graduate research using the Delphi method, the study began with qualitative analysis, which then fed into quantitative analysis of Likert-style questions in subsequent rounds of surveys (Skulmoski et al., 2007). This technique was used because

it is an effective method of building consensus among a panel of experts from related subjects, particularly in an educational setting (Hsu & Sanford, 2007; Yousuf, 2007).

Methodology

The Delphi method was utilized in order to gather perceptual data from an expert panel of administrators, teacher trainers, and master teachers selected according to specific criteria. With the dearth of research on the use of SETs at the secondary level (Jezequel, 2008), more study was needed into the opinions of such experts regarding the construction and use of SETs. A Delphi study is a systematic tool that allows for these informed opinions to collected, exchanged, and analyzed (Rayens & Hahn, 2000). The Delphi format was chosen over others (e.g. Nominal Group Technique) because this technique allows the research to be conducted when face-to-face meetings pose a logistical problem, and research has shown that Delphi and Nominal Group techniques result in similar levels of accuracy and quality (Rowe & Wright, 1999), without the requirement of the Nominal Group technique that all participants be physically present (Yousuf, 2007). Delphi studies are also particularly useful in improving understanding of problems and solutions, especially when such problems could benefit from considering the subjective views of experts (Skulmoski et al., 2007). Finally, Delphi studies allow a panel to engage in a multifaceted process that allows for group interaction, feedback, and exploration anonymously, with the end result being consensus regarding policy issues (Rayens & Hahn, 2000).

This study comprised a Classical Delphi study. Because of variations in how Delphi studies are conducted, Skulmoski, Hartman, and Krahn (2007) suggest that a Delphi study only be named a *Classical Delphi* study if it adheres to specific criteria:

- Anonymity: the study maintains the anonymity of its participants through the use of questionnaires, which frees group members from negative social pressures and ensures that participants consider ideas based on their merit;

- Iteration: by maintaining anonymity through each iteration, participants can change opinions without losing status among the group;

- Controlled feedback: statistical summaries of round results and, on occasion, specific arguments of individual members are distributed, providing participants with the judgments and opinions of the entire group and not just the loudest voices;

- Statistical aggregation: at the end of the final cycle, the final judgment of the group is determined from the statistical average of the last round of responses (Rowe & Wright, 1999).

In order to adhere to these criteria, the following three-round Delphi process was used to conduct the study:
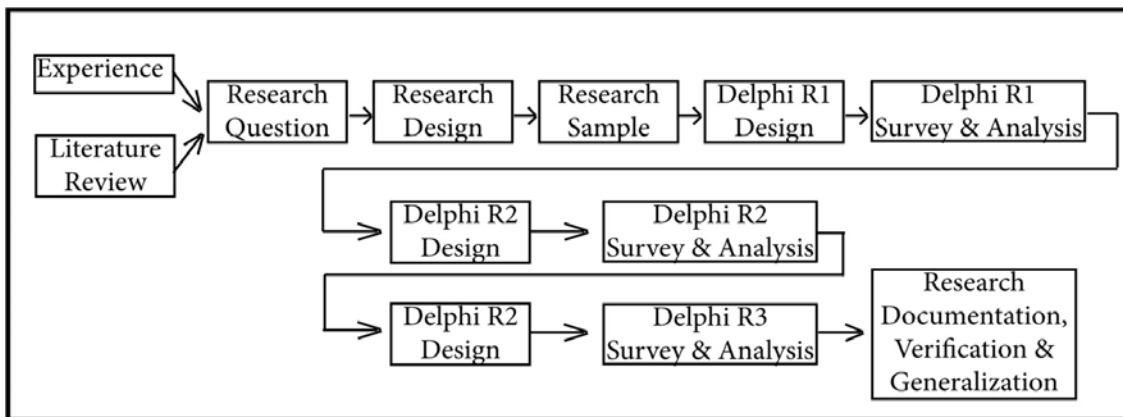
Figure 1. Delphi study methodology. Three sequential rounds of mixed-method survey instruments. Adapted from Skulmoski et al., 2007.

For the purposes of the study, the panel's perceptions were assessed using an electronic questionnaire. As noted in figure 1, these perceptions were elicited in three

rounds of surveying and analysis, which is typical for a Delphi study (Yousuf, 2007). The anonymity of the participants was assured throughout the three rounds of surveying using electronic collection of data, and names were not used when reporting out after each round of surveys.

Survey questions were formulated to elicit the experts' perceptions about the composition and use of SETs, both in terms of their effect on professional development initiatives and in evaluating teacher effectiveness. As is typical of the first round in a Delphi study, the initial questions were open-ended, so that the full range of the panel's perceptions could be elicited (Hsu & Sanford, 2007).

Population

The population of a study is the group about which the researcher wishes the results of the study to generalize (Gay & Airasian, 1996). For this study, the intended populations were administrators and teacher trainers involved in pre- and in-service training of secondary teachers. In an ideal situation, all members of a population would be studied; however, feasibility becomes a factor when dealing with large groups that are spread out geographically (Roberts, 2010, p. 149). In California alone, there are 79,944 teachers working at the secondary level for the 2014-15 school year ("Fingertip Facts," 2016). There are also approximately 1,020 administrators involved in curriculum and instruction, spread out over fifty-eight counties ("Membership Trends," 2013). Added to this is the fact that standards for teacher training and certification differ from state to state, making nationwide generalizations difficult (Kelly, 2015). Consequently, the population was limited to administrators and teacher trainers working in California.

Sample

The participants of a Delphi study should be people who are actively involved in a topic and capable of contributing current and practical knowledge (Hsu & Sanford, 2007). Therefore, for quality assurance considerations, specific criteria needed to be used in selecting panel members (Patton, 2002, p. 238). Participants were solicited from a number of professional and instructional organization forums, including the California Writing Project (CWP), the California Association of Teachers of English (CATE), the Association of California School Administrators (ACSA), and the Beginning Teacher Support and Assessment (BTSA), as well as through direct correspondence with directors of teacher training institutions (e.g. the education departments of local California State Universities, private universities, and local Offices of Education). For purposes of the study, participants were divided into three groups: master teachers, teacher training instructors, and administrators. Master teachers and teacher training instructors were differentiated by where their work lay in the training process (i.e., those involved in the professional development of current teachers and those involved in the instruction of future teachers). Panelists for this Delphi study were selected based on their conformity to separate criteria for each of the three groups, with panel inclusion requiring that a minimum of three standards be met (see Table 1). The initial set of panelists comprised thirty members: eleven master teachers, nine teacher training instructors, and ten administrators. These panel members came from institutions throughout California, representing six secondary schools, six school districts, and five public and private teacher training institutions.

Table 1

Criteria for inclusion in the Delphi Study

Criteria for inclusion in the Delphi study		
Master Teachers	Teacher Training Instructors	Administrators
Five years of teaching experience	Five years of teaching experience	Five years of administrative work
Mentoring/teacher leadership experience (e.g. Curriculum Council, curriculum committee, BTSA mentor)	Direct contact with teachers in a coaching/support role (e.g. pre- and in-service teacher training, a Teacher On Special Assignment role)	Direct contact with teachers in a coaching/support role
Department head	Belong to a professional organization (NCTE, ACSA, CTA, CATE, ACSD, CWP, etc.)	Experience with data analysis (Healthy kids survey, PE tests, performance data)
A level of professional development through conference attendance or participation in formal professional development trainings (PLC, EDI, Kagan, etc.)	Professional development in coaching or supporting new and experienced teachers	Classroom evaluation experience
Advanced Degree	Advanced Degree	Advanced Degree

Typically, ten to thirty experts are employed in a policy Delphi study (Rayens & Hahn, 2000), with a higher number for non-homogenous groups (Hsu & Sanford, 2007). In order to ensure that each group's opinions were sufficiently sampled, a minimum of nine panelists from each group were engaged in the study (Isaac & Michael, 1981). In this study, each category initially contained at least nine members also as a defense against any potential attrition over the course of the surveys.

Instrumentation

In a Delphi study, the first round of surveys typically comprises open-ended questions so as to elicit the widest range of opinions on the questions from participants (Skulmoski et al., 2007; Yousuf, 2007). For this study, the following open-ended questions were distributed to the participants:

- In a Student Evaluation of Teachers (SET) survey to be used for evaluation and professional development purposes, what specific aspects of a teacher's classroom practice should be addressed?

- How can the results of these SETs best be incorporated in the evaluation process?

- How should the results of these SETs be used to inform professional development practices?

The responses of master teachers, teacher trainers, and administrators to round-one electronic surveys were aggregated and used to form the basis of the round-two questions, where participants were asked to rank the importance of each item on a Likert scale and also provide rationale for their decisions (Rayens & Hahn, 2000). These results were again analyzed and used to form the basis for round three, where panelists were

60

asked to rank the items and comment on their decisions, including comments on why

they continued to remain outside the consensus. If sufficient consensus had not yet been

reached, a fourth and final round of surveys would have been implemented (Hsu &

Sanford, 2007; Yousuf, 2007). Following the final round, the researcher verified and

documented the results, then reported these in the form of a dissertation (Skulmoski et al.,

2007).

Instrument Field Tests/Validity

To increase the reliability and validity of the survey instruments, prior to the start

of the first round of questioning, round one questions were subjected to a field test by

three experts, each meeting the criteria for one of the categories. Feedback about the

structure and language of the questions was gathered, and these questions were revised as

a result of the experts' input, where necessary.

Data Collection

Once IRB approval was secured, the directors of various teacher training and

teacher and administrator support organizations (e.g. BTSA, ACSA, CWP, CATE,

TCOE) were contacted to request the name of an organization designee to act as the

contact point for the study. With the approval of the directors, the designees were asked

to distribute via electronic means (e.g. blog, email newsletter, listserv) the invitation to

participate in the study (see Appendix A). Participation on the panel was limited to those

meeting the criteria. When at least five members from each category were identified, the

researcher sent an email outlining the purpose and processes of the study and to obtain

consent for participation. The email also outlined confidentiality procedures and the use

of responses. Throughout the study, confidentiality was maintained, and the results did

not contain any information regarding names or work affiliations. When informed

consent was confirmed by receipt of a signed form (see Appendix B), participants were

sent an electronic link enabling them to provide input in Round One of the process. This

electronic link led participants to a site containing an introduction to the process,

instructions on how to complete the survey, relevant definitions and terms, and a deadline

for survey completion. This information was included in each round of the survey.

Data Analysis

Following each round of surveys, data were analyzed and used as the basis for the

next round. Qualitative data from the first round of surveys was coded and compiled, and

the results were used to create the Likert-scale questions for the second round of surveys.

The quantitative data from responses to the second round were averaged and used to rank

the survey items.   Panel participants were also asked to provide a rationale for their

ratings of survey items. The ratings and comments from round two formed the basis for

the third round of surveys, where participants were asked to revise their rankings and,

where applicable, specify why their responses remain outside the consensus. If sufficient

consensus was not reached by the end of the third round, a fourth round would be

conducted in the same fashion as the third (Hsu & Sanford, 2007). Following the final

round of surveys, the resulting data and comments were analyzed, and the results were

published as a doctoral dissertation.

Limitations

The following were limitations of this study:The study was conducted using

teacher trainers, master teachers, and administrators working in California. The

population may not have been representative of all such individuals outside state borders.

The sample was limited to individuals meeting at least three of the five criteria for inclusion for each group in the study. Results may not be generalizable to a population not meeting these criteria.

All participants were volunteers, which may have skewed the results, as individuals with strong views might have been overrepresented ("Bias in Survey," 2015). At the same time, bias was controlled for through a number of measures, among them the use of anonymous surveys to eliminate dominance bias, and the inclusion of feedback of reasons along with numerical survey data in each round, which has been found to increase the accuracy of data obtained (Hallowell, 2009). Reliability and validity were also confirmed through the use of iteration, the redistribution of surveys with controlled feedback (Hallowell, 2009).

The study relied on a survey instrument whose reliability was not measured over a wide range of contexts.

Summary

The contents of Chapter III include the purpose of the study, research questions, and a presentation of the methodology to be used, which consists of information about the population and sample, instruments, data collection and analysis procedures, and study limitations.

CHAPTER IV: RESEARCH, DATA COLLECTION, AND FINDINGS

Chapter IV begins with a brief introduction providing the reader with a frame of reference and understanding of the material to be covered in this chapter. The introduction includes the major categories of the chapter and serves as a simplified overview of chapter content. The purpose, research questions, methodology, data collection procedures, and population and sample are summarized prior to the presentation of data. Chapter IV should include a detailed report of the findings of the research study as clearly and succinctly as possible.

Overview

For this study, Chapter I featured background information about the current educational environment and the use of SETS in evaluation and professional development. Chapter II reviewed the literature concerning the use of SETS, current evaluation and professional development practices, and andragogy. Chapter III covered the methodology and research design of the study, including information on the population, sample, instrumentation, and data and analysis procedures.

In this chapter is included a summary of the study and a presentation of the data gathered and analyzed in the course of the study. Also included are the purpose and research questions, as well as the methodology, population, and sample. For each round of the Delphi study, data aligned with each research question is presented. The chapter concludes with a summary of findings.

Purpose Statement

The purpose of this Delphi study was to identify the most important elements for SETs (Student Evaluation of Teachers) at the high school level as perceived by a panel of

expert master teachers, administrators, and teacher trainers. In addition, it was the purpose to determine how the results of SETs can best be used by teacher trainers and administrators to inform evaluation and professional development practices for secondary teachers.

Research Questions

The study sought to answer the following research questions:

1. What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

2. How do the panel of master teachers, administrators and teacher trainers rank the importance of the elements of SETs?

3. What do a panel of master teachers, administrators, and teacher trainers identify as strategies for using the data from SETs to inform evaluation and professional development for secondary teachers?

Research Methods and Data Collection Procedures

This study utilized the Delphi method to elicit perceptual data from an expert panel of master teachers, pre- and in-service teacher trainers, and administrators. Electronic questionnaires were used to assess the perceptions of respondents about the content and use of SETs at the secondary level. These questionnaires were administered in three rounds, with the second round divided into two parts to ease processing of the large number of responses in round one. The results of round-one questions were analyzed to inform the creation of the round-two surveys. This process was then applied to the round two responses to create the final set of questions for round three.

65

Population

For this study, the intended populations were administrators and teacher trainers involved in pre- and in-service training of secondary teachers in the state of California. Permission was received from the appropriate authorities from local school districts, teacher training institutions, and teacher training groups to distribute an electronic flyer calling for participation in the study (see Appendix A). These flyers, along with a participant's bill of rights and a request for informed consent, were then distributed through group mailings and listservs. Initially, the flyers elicited responses from only a few qualified participants. The researcher, a California educator with over twenty years of teaching and teacher training experience at the secondary and tertiary levels, then reached out personally via email to experienced administrators and teacher trainers involved in pre- and in-service training of California secondary teachers. Thirty master teachers and pre- and in-service teacher trainers responded and provided informed consent. All thirty respondents were included as expert panelists for the Delphi study and received electronic questionnaires in each of the three rounds of the study. Of the thirty panel members, eleven were master teachers, ten were teacher trainers, and nine were administrators, each according to the criteria established in Table 1. For the first round of the study, twenty-six panelists (86%) completed the survey. Twenty-four panelists (80%) completed round two's surveys, and twenty-six (80%) completed the round three survey.

Sample

Because the participants of a Delphi study should be people who are actively involved in a topic and capable of contributing current and practical knowledge (Hsu & Sanford, 2007), specific criteria were used in selecting panel members (Patton, 2002, p.

238). Participants were solicited from a number of professional and instructional

organization forums, including the California Writing Project (CWP), the California

Association of Teachers of English (CATE), and the Visalia Unified School District's

Beginning Teacher Support and Assessment (BTSA) cohort, as well as through direct

correspondence with directors of teacher training institutions (e.g. the education

department of Fresno Pacific University and local Offices of Education). For purposes of

the study, participants were divided into three groups: master teachers, teacher training

instructors, and administrators. Master teachers and teacher training instructors were

differentiated by where their work lay in the training process (i.e., those involved in the

professional development of current teachers and those involved in the instruction of

future teachers). Panelists for this Delphi study were selected based on their conformity

to separate criteria for each of the three groups, with panel inclusion requiring that a

minimum of three standards be met (see Table 1). The initial set of panelists comprised

thirty members: eleven master teachers, nine teacher training instructors, and ten

administrators. Of the initial thirty participants, twenty-four completed all three rounds

of the survey. Two participants did not send in the second half of round two's survey but

later rejoined the study for the final round of questions.

Demographic Data

The participants of the Delphi study comprise a diverse and highly qualified

group of individuals. Tables 2-6 present the group's demographic data:

Table 2

Primary profession of panelists

Primary Role in Education	Percentage of Participants
Teacher Trainer (pre- or in-service)	66%
Administrator	33%

The group contained a majority of pre- and in-service teachers. At the same time, fully half of the administrators in the group would have also qualified for the study given their teaching experience prior to becoming administrators.

Table 3

Age of panelists

Age of Panelists	Percentage of Participants
30-39	12%
40-49	34%
50-59	34%
60-69	12%
70-79	8%

The largest groups fell in the 40-49 and 50-59 range. Taken together, over half of the study panel comprised mid-career teachers and administrators.

Table 4

Gender of panelists

Gender of Panelists	Percentage of Participants
Female	65%
Male	35%

Almost two-thirds of the twenty-six panelists were female.

Table 5

Education level of panelists

Education Level of Panelists	Percentage of Participants
BA/BS	12%
MA/MS	76%
Ed.D/Ph.D	12%

Over three-fourths of the panelists had an MA or MS. Combined with those

panelists having doctorate, almost ninety percent of the panelists had graduate degrees.

Table 6

Years of work in education

Years of Work in Education of Panelists	Percentage of Participants
5-9	4%
10-19	38%
20-29	23%
30 or more	35%

While all panelists had at least five years of experience in their field (one of the

criteria for inclusion in the study), ninety-five percent of the panelists had at least ten

years of experience in education. When it became necessary to personally invite

panelists due to low response to the electronic calls for participation, the researcher

deliberately reached out to highly-qualified educators and administrators for inclusion.

These included educators and administrators from six different high schools, six districts,

and five teacher training institutions throughout the state of California.

Presentation and Analysis of Data

Data are presented for each research question consecutively, beginning with research question one. Each of the three rounds of the Delphi study is reported consecutively for each research question.

Research Question One

What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

Round One.   In round one, participants were asked to respond to an open-ended question: *If high school students were being surveyed about their teacher's work in their class, and that information might be used for evaluation or professional development purposes, what should we be asking about the teachers?* Instructions accompanying the survey requested that respondents list as many areas to be surveyed of students and, where possible, to include the actual questions to be asked. They were also encouraged to provide justification for particular responses where appropriate (see Appendix C).

The survey was emailed to the thirty participants providing informed consent. Twenty-six panel experts responded to the round one questionnaire. The researcher then reviewed, sorted, and categorized panel members' responses. Similar responses were combined, while multi-part responses were disaggregated. For example, if a respondent mentioned that students should be surveyed on what they spent time doing in a class and then gave examples such as engaging in group work, listening to the teacher talk, or answering questions on a worksheet, each of these choices was added to the list of possible questions to ask in a survey.

From the round-one responses to question one, the researcher generated a list of fifty-one potential items to be considered for inclusion in a SET. This list was narrowed down by the researcher to forty-nine unique items, as outlined in Table 7.

Table 7

Questions potentially to be included in a SET at the secondary level, as reported by a panel of expert teacher trainers and administrators

Potential Question	Frequency of mention
Does your teacher have clear objectives for each day, posted visibly?	11
Does your teacher often have you work with a partner or group during a lesson?	7
Is your teacher available outside of class for extra help?	6
Does your teacher come prepared to class each day?	5
Does your teacher know the subject he/she is teaching well?	5
Does your teacher care about the students in this class?	5
Does your teacher make the material engaging?	5
Can your teacher convey concepts in multiple ways?	5
Do the course materials feel useful and relevant to real life?	5
Does your teacher give you effective feedback on your work in a timely manner?	4
Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?	4
Do you feel safe asking questions, commenting, or asking for help in class?	4
Does your teacher ask you to show that you understand during a lesson?	4
Does your teacher have a good rapport with the students?	4
Is your teacher excited about his/her subject matter?	4
Does your teacher give good instructions?	4
Is your teacher fair and equitable?	4
What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students)	4
Do you feel welcomed and supported by your teacher?	3
Does the teacher ensure that you know what criteria you will be measured against?	3
Does your teacher make good use of class time?	3
Does your teacher use technology in the class? Do students?	3
Do you know how your teacher wants routine classroom actions handled?	3

71

Does your grade in class reflect your learning, or does it reflect other aspects?	3
Do you feel challenged in this class?	2
Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?	2
Does your teacher have a 'can do' attitude towards students' ability and work?	2
Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?	2
Does the homework for this class reinforce the learning done during lessons?	2
How much of class is usually spent in lecture vs. in interactive work?	2
Does your teacher know your individual strengths and weaknesses?	2
Can your teacher think on his/her feet to keep a class moving?	2
Does your teacher change the way he/she teaches based on individual student needs?	2
How flexible is your teacher?	2
Does your teacher require you to write to justify or explain ideas?	1
Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?	1
Does your teacher have high standards for your work?	1
Does your teacher give individual help when necessary?	1
Does the content of the course prepare you for the exams?	1
Do you have a sense of belonging in this class?	1
Do you feel like you accomplish something in class each day?	1
What parts of the class were difficult? Why? (Short answer)	1
How much do you feel you've learned in class this year?	1
Does your teacher move from activity to activity well?	1
When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Short answer)	1
Does your teacher link course content to other subjects/disciplines?	1
What makes a good teacher? (Short answer)	1
What connections have you made in class this year? (Short answer)	1
How did you feel about the subject of this class before you took it? And now? ? (Short answer)	1

Analysis of Round One. All 26 members responding to the first round of the

questionnaire provided multiple examples of what they felt should be included in a SET

at the secondary level. With eleven references, the item most frequently mentioned by

panelists was regarding having posted daily objectives in the classroom. Next in

frequency (seven mentions) came a question about a teacher's use of group work during

the lesson. A teacher's availability for help outside class received six mentions in the

survey. Items ranging from a teacher's preparedness to subject-matter knowledge to the

relevancy of course materials were mentioned five times. As the frequency of mention

decreased, the number of discrete items increased, with fifteen items being mentioned

only one time each.

Emerging Themes of Research Question One. With forty-nine different areas

potentially being covered in a SET, certain themes arose. Various aspects of a teacher's

behaviors in the classroom featured prominently, among them their transitions from

activity to activity, the giving of individual help, the level of engagement established, the

wise use of class time, the giving of timely and effective feedback, and the ability to

provide effective examples and individualized help. Affective factors were also featured,

with students being asked to comment on whether they felt a sense of belonging, how

they felt about the subject matter before and after the course, and if they felt they

accomplished something in class each day. Finally, classroom activities themselves came

into focus, with questions concerning the connection between class work and homework,

the frequency of group work and peer-response activities, and the use of media and

technology to enhance learning. In general, the questions offered by the participants

were of a variety that could be answered on a Likert scale; however, six of the questions

asked for a more extended response, asking students for specific details or a short-answer response.

With this list of potential questions, the researcher then began surveying participants on which and how many of these should be included in a SET, which was the main thrust of research question two.

Research Question Two

How do the panel of master teachers, administrators and teacher trainers rank the importance of the elements of SETs?

Round Two. In the second round of surveys, participants were asked to rank each of the items generated in research question one on a Likert scale, based on how important each was to include in a SET (see Appendix D). The scale points ranged from 1 (Not important) to 6 (Extremely Important). Twenty-six participants responded to this round of the survey, and from the results, the researcher was able to calculate an initial ranking of the possible items to include in a SET.

Table 8

Rankings of possible questions to be included in a SET, as reported by a panel of expert teacher trainers and administrators

Possible question for inclusion in a SET	Survey results on a 1-6 Likert scale
1. Does your teacher give you effective feedback on your work in a timely manner?	5.38
2. Does your teacher come prepared to class each day?	5.38
3. Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?	5.33
4. Do you feel welcomed and supported by your teacher?	5.333
5. Do you feel safe asking questions, commenting, or asking for help in class?	5.33
6. Does your teacher ask you to show that you understand during a lesson?	5.14
7. Does your teacher have a good rapport with the students?	5.10

8. Does your teacher know the subject he/she is teaching well?	5.05
9. Does your teacher require you to write to justify or explain ideas?	5.00
10. Does your teacher care about the students in this class?	5.00
11. Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?	4.95
12. Does your teacher have high standards for your work?	4.86
13. Does your teacher give individual help when necessary?	4.86
14. Does the content of the course prepare you for the exams?	4.86
15. Is your teacher excited about his/her subject matter?	4.81
16. Does your teacher give good instructions?	4.81
17. Do you feel challenged in this class?	4.81
18. Does the teacher ensure that you know what criteria you will be measured against?	4.81
19. Does your teacher make good use of class time?	4.76
20. Does your teacher make the material engaging?	4.67
21. Does your teacher use technology in the class? Do students?	4.67
22. Do you have a sense of belonging in this class?	4.62
23. Is your teacher fair and equitable?	4.62
24. Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?	4.52
25. Does your teacher have a 'can do' attitude towards students' ability and work?	4.52
26. What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students)	4.52
27. Does your teacher often have you work with a partner or group during a lesson?	4.48
28. Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?	4.48
29. Do you feel like you accomplish something in class each day?	4.38
30. What parts of the the class were difficult? Why? (Short answer)	4.33
31. Can your teacher convey concepts in multiple ways?	4.33
32. How much do you feel you've learned in class this year?	4.33
33. Does the homework for this class reinforce the learning done during lessons?	4.33
34. Do you know how your teacher wants routine classroom actions handled?	4.33

35. Does your teacher have clear objectives for each day, posted visibly?	4.29
36. Do the course materials feel useful and relevant to real life?	4.29
37. How much of class is usually spent in lecture vs. in interactive work?	4.29
38. Is your teacher available outside of class for extra help?	4.14
39. Does your grade in class reflect your learning, or does it reflect other aspects?	4.14
40. Does your teacher know your individual strengths and weaknesses?	4.00
41. Does your teacher move from activity to activity well?	3.95
42. When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Short answer)	3.95
43. Does your teacher link course content to other subjects/disciplines?	3.86
44. Can your teacher think on his/her feet to keep a class moving?	3.71
45. Does your teacher change the way he/she teaches based on individual student needs?	3.71
46. What makes a good teacher? (Short answer)	3.67
47. What connections have you made in class this year? (Short answer)	3.38
48. How did you feel about the subject of this class before you took it? And now? (Short answer)	3.14
49. How flexible is your teacher?	2.76

Analysis of Round Two. The participant responses for round two were averaged

for each item, as this has been deemed a robust method for aggregating subjective

judgments (Sommerville, 2008). The results of round two show that there was little

correlation between how often an item was introduced by respondents in round one and

how necessary it was deemed for inclusion in SETs in round two. This is probably due to

panelists recognizing the value in items introduced by other members of the panel. For

example, round one's most often mentioned item, regarding the posting of daily

objectives, ranked only 35th among respondents in round two, thus showing that items

frequently mentioned initially by panelists were not always valued as highly when more

items came into the picture. Conversely, the two most-valued items, regarding effective feedback and preparedness, were mentioned only four and five times, respectively, in the initial survey.

One theme that emerged involved a higher ranking for questions about a teacher's actions and attitudes (e.g. approachability, subject-matter knowledge, ability to give good feedback and instructions, caring for students, etc.) than the demands on students in the classroom (e.g. course challenge, connection between homework and classwork, relevant materials, etc.). Finally, those questions requiring an extended response from students tended to score low on the survey, with only one such question breaking the top twenty-five in the rankings.

Round Three. For the final round of the survey, participants were shown the results of the previous round's rankings as presented in Table 8. They were then asked whether each item should remain in its place or be raised or lowered in its position. As seen in Table 9, for only a few of the items was the number of participants choosing to raise or lower an item's position greater than those opting to keep it in its current place. In general, panelists opted to keep items in their current quartile in all but four instances, showing growing consensus regarding the rankings of the potential SET questions.

Table 9

Suggestions for movement of items in the rankings, as reported by a panel of expert teacher trainers and administrators

Possible SET Question	Move up in list	Move down in list	Keep in place	Up vs down	Keep in place vs move
1. Does your teacher give you effective feedback on your work in a timely manner?	6	0	18	+6	+12

2. Does your teacher come prepared to class each day?	6	2	15	+4	+9
3. Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?	11	1	12	+10	+1
4. Do you feel welcomed and supported by your teacher?	7	1	14	+6	+7
5. Do you feel safe asking questions, commenting, or asking for help in class?	4	2	18	+2	+16
6. Does your teacher ask you to show that you understand during a lesson?	13	0	11	+13	-2
7. Does your teacher have a good rapport with the students?	6	6	13	0	+7
8. Does your teacher know the subject he/she is teaching well?	8	7	10	+1	+2
9. Does your teacher require you to write to justify or explain ideas?	7	2	14	+5	+7
10. Does your teacher care about the students in this class?	6	4	14	+2	+8
11. Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?	13	1	12	+12	-1
12. Does your teacher have high standards for your work?	4	3	17	+1	+13
13. Does your teacher give individual help when necessary?	8	1	15	+7	+7
14. Does the content of the course prepare you for the exams?	7	5	12	+2	+5
15. Is your teacher excited about his/her subject matter?	10	3	12	+7	+2
16. Does your teacher give good instructions?	8	2	13	+6	+5
17. Do you feel challenged in this class?	5	5	14	+9	+9

18. Does the teacher ensure that you know what criteria you will be measured against?	14	2	7	+12	-7
19. Does your teacher make good use of class time?	8	3	13	+5	+5
20. Does your teacher make the material engaging?	8	1	15	+7	+7
21. Does your teacher use technology in the class? Do students?	3	7	15	-4	+12
22. Do you have a sense of belonging in this class?	8	4	12	+4	+4
23. Is your teacher fair and equitable?	9	2	13	+7	+4
24. Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?	5	4	14	+1	+9
25. Does your teacher have a 'can do' attitude towards students' ability and work?	8	6	11	+2	+3
26. What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students)	5	5	13	+8	+8
27. Does your teacher often have you work with a partner or group during a lesson?	7	8	9	-1	+1
28. Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?	12	3	10	+9	-2
29. Do you feel like you accomplish something in class each day?	10	3	11	+7	+1
30. What parts of the the class were difficult? Why? (Short answer)	5	7	13	-2	+6
31. Can your teacher convey concepts in multiple ways?	7	2	14	+5	+7
32. How much do you feel you've learned in class this year?	5	8	11	-3	+3

33. Does the homework for this class reinforce the learning done during lessons?	11	3	11	+8	0
34. Do you know how your teacher wants routine classroom actions handled?	7	7	12	0	+5
35. Does your teacher have clear objectives for each day, posted visibly?	9	2	13	+7	+4
36. Do the course materials feel useful and relevant to real life?	8	4	12	+4	+4
37. How much of class is usually spent in lecture vs. in interactive work?	5	5	13	0	+8
38. Is your teacher available outside of class for extra help?	6	2	9	+4	+3
39. Does your grade in class reflect your learning, or does it reflect other aspects?	8	4	12	+4	+4
40. Does your teacher know your individual strengths and weaknesses?	7	4	13	+3	+6
41. Does your teacher move from activity to activity well?	2	7	13	-5	+6
42. When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Short answer)	10	1	13	+9	+3
43. Does your teacher link course content to other subjects/disciplines?	7	4	14	+3	+7
44. Can your teacher think on his/her feet to keep a class moving?	3	8	13	-5	+5
45. Does your teacher change the way he/she teaches based on individual student needs?	8	3	12	+5	+4
46. What makes a good teacher? (Short answer)	8	8	9	0	+1
47. What connections have you made in class this year? (Short answer)	5	8	12	-3	+4

	4	7	12	-3	+5
48. How did you feel about the subject of this class before you took it? And now? (Short answer)	4	7	12	-3	+5
49. How flexible is your teacher?	5	10	11	-6	+1

Analysis of round three. Based on this third round of surveys, the rankings established in round two are largely stable, as usually happens with a Delphi study (Sommerville, 2008). In only four cases out of forty-nine were the number of votes for moving an item in the rankings greater than the number of votes for keeping it in its current place. In all four cases, the respondents showed a preference for moving the items up in the rankings.

From the second-round results, a relative ranking of all items into quartiles was generated. Factoring in the third-round results, four quartiles were established regarding the relative importance of each of the forty-nine items for inclusion in a SET.

Table 10

Final ranking of possible items for inclusion in a SET at the secondary level, divided into quartiles, as reported by a panel of expert teacher trainers and administrators.

Rank	Question	Quartile
1	Does your teacher give you effective feedback on your work in a timely manner?	1
2	Does your teacher come prepared to class each day?	1
3	Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?	1
4	Do you feel welcomed and supported by your teacher?	1
5	Do you feel safe asking questions, commenting, or asking for help in class?	1
6	Does your teacher ask you to show that you understand during a lesson?	1
7	Does your teacher have a good rapport with the students?	1
8	Does your teacher know the subject he/she is teaching well?	1
9	Does your teacher require you to write to justify or explain ideas?	1
10	Does your teacher care about the students in this class?	1

11	Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?	*1*
12	Does the teacher ensure that you know what criteria you will be measured against?	*1*
13	Does your teacher have high standards for your work?	*2*
14	Does your teacher give individual help when necessary?	*2*
15	Does the content of the course prepare you for the exams?	*2*
16	Is your teacher excited about his/her subject matter?	*2*
17	Does your teacher give good instructions?	*2*
18	Do you feel challenged in this class?	*2*
19	Does your teacher make good use of class time?	*2*
20	Does your teacher make the material engaging?	*2*
21	Does your teacher use technology in the class? Do students?	*2*
22	Do you have a sense of belonging in this class?	*2*
23	Is your teacher fair and equitable?	*2*
24	Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?	*2*
25	Does the homework for this class reinforce the learning done during lessons?	*2*
26	Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?	*3*
27	Does your teacher have a 'can do' attitude towards students' ability and work?	*3*
28	What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students)	*3*
29	Does your teacher often have you work with a partner or group during a lesson?	*3*
30	Do you feel like you accomplish something in class each day?	*3*
31	What parts of the the class were difficult? Why? (Short answer)	*3*
32	Can your teacher convey concepts in multiple ways?	*3*
33	How much do you feel you've learned in class this year?	*3*
34	Do you know how your teacher wants routine classroom actions handled?	*3*
35	Does your teacher have clear objectives for each day, posted visibly?	*3*
36	Do the course materials feel useful and relevant to real life?	*3*
37	How much of class is usually spent in lecture vs. in interactive work?	*3*
38	Is your teacher available outside of class for extra help?	*4*
39	Does your grade in class reflect your learning, or does it reflect other aspects?	*4*
40	Does your teacher know your individual strengths and weaknesses?	*4*
41	Does your teacher move from activity to activity well?	*4*

42	When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Short answer)	*4*
43	Does your teacher link course content to other subjects/disciplines?	*4*
44	Can your teacher think on his/her feet to keep a class moving?	*4*
45	Does your teacher change the way he/she teaches based on individual student needs?	*4*
46	What makes a good teacher? (Short answer)	*4*
47	What connections have you made in class this year? (Short answer)	*4*
48	How did you feel about the subject of this class before you took it? And now? (Short answer)	*4*
49	How flexible is your teacher?	*4*

Emerging Themes of Research Question Two. These rounds of surveys took the initial set of possible questions for SETs and ranked them. Looking at the three categories of questions that emerged in the first research question (i.e. teacher behavior, affective factors, content/activities), the rankings reveal a perceived importance for those questions that deal with a teacher's behaviors. Of the twelve questions ranking in the first quartile, nine of them deal with a teacher's actions, competencies, and abilities. Affective factors figured less prominently throughout, with only one or two questions in each quartile dealing with how students feel about various aspects of their classroom, subject, or teacher. Moving down the quartiles, questions concerning classroom content and activities become more prominent. These rankings suggest that the panel feels that more benefit would come from questioning students about their perceptions of their teacher than about their feelings about the class and subject matter or about the activities and content of courses. This is not to say that the other two types of questions should not be used, as they still comprise nearly half of the total questions. Rather, it shows questions about teacher behavior are seen as having importance in the SET process by the

83

panel. If a SET were to have a limited number of questions on which students would respond, the greater proportion of those questions could deal with teacher behaviors.

This raises an interesting point concerning the information that the Delphi group wanted to find out from SETs that might not be available by other means. Two prominent sources of what is useful and expected from teachers can be found in John Hattie's *Visible Learning* (Hattie, 2009) and the California Standards for the Teaching Profession ("CSTPs," 2018). The former provides in Appendix B a ranked list of the relative effect sizes of various initiatives and actions in education based on a meta-analysis of hundreds of studies, while the latter provides a prescriptive list of standards deemed essential to effective teaching practice for California educators. Table 11 shows congruence between the survey items selected by the Delphi panel and the contents of these two lists.

Table 11

A comparison of the forty-nine SET questions selected by a panel of expert teacher trainers and administrators and the items featured in Hattie's list of effective actions and the corresponding CSTP standards and sub-standards.

Delphi Study SET Questions	Corresponding CSTP Standards and Sub-standards	Correspondence with Hattie's Rankings Based on Effect Sizes
1. Does your teacher give you effective feedback on your work in a timely manner?	5.5	10 Feedback
2. Does your teacher come prepared to class each day?		
3. Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?	1.2, 5.4	8 Clarity
4. Do you feel welcomed and supported by your teacher?		11 Teacher-Student Relationships

Question		
5. Do you feel safe asking questions, commenting, or asking for help in class?	1.4, 2.1	11 Teacher-Student Relationships
6. Does your teacher ask you to show that you understand during a lesson?	1.4, 1.5	
7. Does your teacher have a good rapport with the students?		11 Teacher-Student Relationships
8. Does your teacher know the subject he/she is teaching well?	3.1, 3.4, 6.1	125 Teacher Subject-Matter Knowledge
9. Does your teacher require you to write to justify or explain ideas?	1.2	
10. Does your teacher care about the students in this class?		11 Teacher-Student Relationships
11. Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?	1.2	30 Worked Examples
12. Does your teacher have high standards for your work?	1.5, 4.3, 4.4	58 Expectations
13. Does your teacher give individual help when necessary?	1.2	
14. Does the content of the course prepare you for the exams?	5.1	
15. Is your teacher excited about his/her subject matter?		
16. Does your teacher give good instructions?		8 Clarity
17. Do you feel challenged in this class?	1.5, 4.1, 4.4	
18. Does the teacher ensure that you know what criteria you will be measured against?	1.5	
19. Does your teacher make good use of class time?	2.6	70 Time on Task
20. Does your teacher make the material engaging?	1.2	
21. Does your teacher use technology in the class? Do students?	1.2, 2.1, 3.5	71 Computer-Assisted Instruction
22. Do you have a sense of belonging in this class?		11 Teacher-Student Relationships
23. Is your teacher fair and equitable?	1.4, 2.2	

24. Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?	1.2, 3.5	
25. Does your teacher have a 'can do' attitude towards students' ability and work?	1.4	11 Teacher-Student Relationships
26. What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students)	1.2, 3.5	
27. Does your teacher often have you work with a partner or group during a lesson?	1.3, 2.3	24 Cooperative vs. Individualistic Learning
28. Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?	1.2, 1.4, 1.5, 5.3	24 Cooperative vs. Individualistic Learning
29. Do you feel like you accomplish something in class each day?	1.4, 3.5	
30. What parts of the the class were difficult? Why? (Short answer)		
31. Can your teacher convey concepts in multiple ways?	1.2	
32. How much do you feel you've learned in class this year?		
33. Does the homework for this class reinforce the learning done during lessons?	5.1	88 Homework
34. Do you know how your teacher wants routine classroom actions handled?	2.3, 2.5	
35. Does your teacher have clear objectives for each day, posted visibly?	1.5, 4.2	34 Goals
36. Do the course materials feel useful and relevant to real life?	1.1, 1.4	
37. How much of class is usually spent in lecture vs. in interactive work?	1.3	
38. Is your teacher available outside of class for extra help?		11 Teacher-Student Relationships
39. Does your grade in class reflect your learning, or does it reflect other aspects?	5.1, 5.2	
40. Does your teacher know your individual strengths and weaknesses?	3.1, 4.1	11 Teacher-Student Relationships
41. Does your teacher move from activity to activity well?	2.6	

42. When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Short answer)	1.3, 2.3, 4.5, 5.2	
43. Does your teacher link course content to other subjects/disciplines?	1.1, 1.4, 3.3, 4.4	
44. Can your teacher think on his/her feet to keep a class moving?	1.1, 1.2, 3.4, 4.5	
45. Does your teacher change the way he/she teaches based on individual student needs?	1.2, 3.4, 4.1	62 Matching Style of Learning
46. What makes a good teacher? (Short answer)		
47. What connections have you made in class this year? (Short answer)	1.4, 3.3	
48. How did you feel about the subject of this class before you took it? And now? (Short answer)		
49. How flexible is your teacher?	1.2	

This comparison of the three items raises some interesting questions. In the chart it is evident that most of the questions developed by the Delphi panel deal with items contained in the CSTPs. In comparison, a number of questions linking closely with Hattie's data deal with teacher-student relationships, an area that is difficult to assess in a short formal evaluation but, based on Hattie's ranking of the item eleventh in a list of 138 items and assigning it an effect size of .71, has a significant effect on learner success (Hattie, 2009, p. 300). In fact, most of the items in the list of questions that are not covered by a CSTP deal with a teacher's affect and a student's response in the classroom. This suggests that while such items might be difficult to assess in a formal evaluation, they could be useful to elicit from students in the process of on-going teacher reflection and professional development.

87

Research Question Three

What do a panel of master teachers, administrators, and teacher trainers identify

as strategies for using the data from SETs to inform evaluation and professional

development for secondary teachers?

This broad research question was divided into several smaller ones to provide guidance

on the size of a SET, its timing, its audience, and its application. Where applicable, these

questions were also divided to reflect differences in SET use in evaluation and in

professional development.

Size of Survey

Round Two. Having generated a list of potential SET survey questions in round

one, participants were asked in round two to determine how large a survey should be with

the following open-ended question: *Given that we have about 50 possible survey items*

here, we also need to think about how large the SET should be. Thinking about both

*manageability and thoroughness, how many items do you feel should be on this surve*y?

The results were divided into three ranges, as shown in table 12:

Table 12

Number of questions to be included in a SET for secondary students, as reported by a
panel of expert teacher trainers and administrators in round two

Number of questions to be on a SET	Number of respondents
5-9	4
10-19	5
20-30	15

Analysis of round two. The results of this round confirm that the participants

believe that the survey should be limited in size. The majority of twenty-four

respondents favored asking between twenty and thirty questions on a SET.

Round Three. For round three of the survey, participants were then asked to choose from the three ranges resulting from round two.

Table 13

Number of questions to be included in a SET for secondary students, as reported by a panel of expert teacher trainers and administrators in round three

Number of questions to be on a SET	Number of respondents
5-9	6
10-19	9
20-30	11

Analysis of round three. The results of this round saw the participants' views growing more varied. While the twenty-to-thirty-question range still received a majority of the votes, the other two ranges saw an increase in popularity. This could be a reflection of the belief, expressed by one participant, that the purposes of giving a SET would determine how many questions were used, and that the instrument could be designed each time to fit the needs of the given situation. The next factor to be investigated was the timing and frequency of the surveys.

Timing of Surveys for Professional Development Purposes

Round One. The round one survey also included a question concerning the timing and frequency of SETs at the secondary level, with a separate question being asked regarding SETs to inform professional development practices and SETs for evaluation purposes. For the former, an open-ended question was used to elicit a range of answers:

If these surveys were to be used to inform professional development practices (either for individuals or groups), when and how often in the school year should students be surveyed about their teachers?

89

Round Two. The responses were then grouped by the researcher and included in a survey question in round two: *If used for professional development purposes, when should the surveys be given?* The twenty-six respondents were asked to choose from the field of choices, with the results shown for this round in Table 14.

Table 14

Potential timing and frequency of administration of SETs for professional development purposes at the secondary level, as reported by a panel of expert teacher trainers and administrators, as reported by a panel of expert teacher trainers and administrators in round two

Timing and frequency of administration	Percentage of Respondents
Twice a year, at the end of each semester, so adjustments can be made for the second semester and the results can then be viewed at the end of the year.	33%
At 'benchmark' points, such as after the first month of school, around Thanksgiving, February, and again in April.	29%
Quarterly, so that adjustments can be made quicker and more often.	21%
Let the teacher decide.	12%
Near the end of the school year, so that results can inform summer professional development efforts.	5%

Analysis of round two. A clear majority of study participants preferred giving SETs at multiple points during the school year. The most popular choice in this round, to give surveys at the end of each semester, received a third of all votes. Second came giving them at specific benchmark points in the school year. One-fifth of participants favored giving them quarterly in order to allow for adjustments to be made more quickly and more often. Twelve percent wanted the teacher to decide when to give the SETs, which leaves the frequency and timing open. Only five percent preferred a single implementation at the end of the year that would inform summer PD efforts.

Round Three. In order to achieve consensus, in the final round of the survey,

participants were shown the results in Table 14 and again asked to choose from the

options for timing and frequency.

Table 15

*Potential timing and frequency of administration of SETs for professional development
purposes at the secondary level, as reported by a panel of expert teacher trainers and
administrators in round three*

Timing and frequency of administration	Percentage of Respondents
Twice a year, at the end of each semester, so adjustments can be made for the second semester and the results can then be viewed at the end of the year.	62%
At 'benchmark' points, such as after the first month of school, around Thanksgiving, February, and again in April.	24%
Quarterly, so that adjustments can be made quicker and more often.	12%
Let the teacher decide.	4%

Analysis of round three. The group came closer to consensus in this round, with

sixty-two percent of respondents calling for SET use at the end of each semester. Giving

at benchmark points and quarterly received fewer votes, but they still represented over a

third of participants between them. The number of participants choosing to let the

teacher decide decreased. None of the participants opted for a single end-of-year

implementation. The consensus of the group is for multiple implementations of SETs for

professional development purposes throughout the school year.

Timing and Frequency of Surveys for Evaluation Purposes

Round One. As with surveys for professional development uses, the round one

survey also included a question concerning the timing and frequency of SETs at the

secondary level for evaluation purposes. An open-ended question was used to elicit a

range of answers: *If these surveys were to be used in the evaluation process, when and how often in the school year should students be surveyed?*

Round Two. The responses were then grouped by the researcher and included in a survey question in round two: *If used for evaluation purposes, when should the surveys be given?* The twenty-four respondents chose from the field of choices, with the results shown for this round in Table 16.

Table 16

Potential timing and frequency of administration of SETs for evaluation purposes at the secondary level, as reported by a panel of expert teacher trainers and administrators, as reported by a panel of expert teacher trainers and administrators in round two

Timing and frequency of administration	Percentage of Respondents
Student surveys should not be used for evaluation purposes	29%
Twice a year, at the end of each semester	25%
Twice a year, coming mid-fall and prior to the springtime evaluation process	21%
Let the teacher decide	17%
Near the end of the school year (so that results can inform summer professional development efforts)	8%

Analysis of round two. While the results of this survey item were similar to those concerning the use of SETs for professional development purposes (see Table 15), an opinion unique to this item that was voiced in round one was the most popular choice among participants in round two, with seven of the twenty-four respondents suggesting that SETs not be used for evaluation purposes. Nearly half of the respondents preferred twice-a-year implementation, either at the end of the semester or in mid-fall and just prior to the springtime evaluation process. The final quarter of respondents opted for either letting the teacher decide on the timing or limiting SET use to one end-of-year implementation.

Round Three. To come closer to consensus, in the final round of the survey,

twenty-six participants were shown the results in Table 16 and again asked to choose

from the options for timing and frequency.

Table 17

*Potential timing and frequency of administration of SETs for evaluation purposes at the
secondary level, as reported by a panel of expert teacher trainers and administrators in
round three*

Timing and frequency of administration	Percentage of Respondents
Student surveys should not be used for evaluation purposes	58%
Twice a year, at the end of each semester	19%
Twice a year, coming mid-fall and prior to the springtime evaluation process	19%
Let the teacher decide	4%

Analysis of round three. The study group came closer to consensus in round

three, with a majority of respondents (58%) suggesting that SETs not be used for

evaluation purposes. The next most common choices, with nineteen percent each, had

surveys being used twice during the school year. The percentage of participants

preferring to let the teacher decide on the timing and frequency dropped from eight

percent to four percent for the final round of the survey. These results suggest that the

group supports the use of SETs for professional development purposes, but it is less

supportive of using them as part of the evaluation process. The next question to be

addressed by the Delphi group involved how the results of SETs should be disseminated.

Audience for SET Surveys for Professional Development Purposes

Round One. In addition to the content and timing of SETs, participants were

asked to comment on the potential audience for the results of SETs used for professional

development purposes: *If these surveys were to be used to inform professional*

development practices (either for groups or individuals), how should the results be

disseminated (i.e., who should see them, and in what forum)? The responses of this open-

ended prompt were collected and categorized by the researcher into six possible

audiences for SET survey results.

Round Two. The six responses were included as a question in round two, where

participants chose as many items as they deemed appropriate: *If these surveys were to be*

used to inform professional development practices (either for groups or individuals), how

should the results be disseminated (i.e., who should see them, and in what forum)?

(Please mark all that apply). The results of this survey question are shown in Table 18.

Table 18

Potential audiences for the results of SET surveys used for professional development
purposes, as reported by a panel of expert teacher trainers and administrators in round
two

Potential audiences for SET results	Percentage of Respondents
Individuals see their own	88%
Administrators	75%
PLCs, without individual names	54%
All staff, without individual names	50%
Department heads, without individual teacher scores	33%
Department heads, with individual teacher scores	21%

Analysis of round two. The most popular option for the audience for SETs for

professional development purposes were the teachers themselves, with eighty-eight

percent of respondents choosing it. Three-quarters felt that administrators should have

access to SET results. The next three most popular audiences (PLCs, all staff, and

department heads) all asked that anonymity be maintained for individual teachers. In

fact, the least popular choice, to allow department heads to see the results for individual

teachers, was only chosen by twenty-one percent of the respondents.

Round Three. To come closer to consensus, for this round participants were shown the results of round two and asked to again note which audiences they deemed appropriate for SET results.

Table 19

Potential audiences for the results of SET surveys used for professional development purposes, as reported by a panel of expert teacher trainers and administrators in round three

Potential audiences for SET results	Percentage of Respondents
Individuals see their own	92%
Administrators	69%
PLCs, without individual names	50%
All staff, without individual names	19%
Department heads, without individual teacher scores	12%
Department heads, with individual teacher scores	4%

Analysis of round three. Letting teachers see their individual results remained the most popular choice among participants, being chosen by ninety-two percent of the group. Giving access to administrators was again chosen by a majority of respondents, but the percentage of those choosing that option dropped from seventy-five to sixty-nine. Letting PLCs view the results anonymously remained a majority choice, while anonymous viewing by all staff or department heads became less popular as an option. Having department heads view the results for individual teachers was the least popular option, this time garnering only four percent of the group's approval. The results suggest that the group prefers letting individuals and their administrators see named results, but that other groups (all staff, PLC) should only see aggregated data.

Uses for SETs for Professional Development Purposes

Round One. The final aspect of SET use for professional development purposes to be investigated involved their use. In an open-ended question, round-one participants

were asked what should be done with survey results: *How should the results of these surveys be used to improve instructional practices, either for groups or individuals?* The twenty-six responses were grouped into six possible actions to be taken.

Round Two. These six responses were included as a question in round two, where the twenty-four respondents were asked to choose which uses they found appropriate: *How should the results of these surveys be used to improve instructional practices, either for groups or individuals? (Please mark all that apply.)*

Table 20

Potential uses for the results of SET surveys used for professional development purposes, as reported by a panel of expert teacher trainers and administrators in round two

Potential Uses for SET results	Percentage of Respondents
Use the results to differentiate PD initiatives for the needs of the teachers.	80%
Administrators should use the data when planning whole-school PD efforts.	63%
The results should be shared by administrators with individual teachers as part of the evaluation/counseling process.	58%
Administrators and grade levels/bands view the data collaboratively to discuss implications and areas of strength/growth.	58%
The results should be used primarily as a needs assessment for the larger PD efforts of a school/district. They should be part of a larger PD plan.	42%
PD could be conducted by teachers scoring high in particular areas, with possible classroom demonstrations of best practices for visiting teachers.	42%

Analysis of round two. The most popular response had SET results being used to differentiate PD initiatives based on the needs reflected in the data acquired. A majority of participants also supported the use of SETs to inform choices for whole-school PD, in conferences between administrators and individual teachers as part of the counseling/evaluation process, and in collaborations between teacher groups and administrators around areas of strength and growth. Using the results as part of a needs

assessment for a site or at the district level received support from forty-two percent of

participants, as did using the results to select particular teachers with high scores to hold

demonstrations of best practices for visiting teachers.

Round Three. In order to come closer to consensus, in the final round of the

survey, participants were asked to look at the responses shown in Table 20 and again pick

which they thought were the most appropriate uses for SETs for professional

development purposes.

Table 21

Potential uses for the results of SET surveys used for professional development purposes,
as reported by a panel of expert teacher trainers and administrators in round three

Potential Uses for SET results	Percentage of Respondents
Use the results to differentiate PD initiatives for the needs of the teachers.	92%
Administrators should use the data when planning whole-school PD efforts.	69%
The results should be shared by administrators with individual teachers as part of the evaluation/counseling process.	50%
Administrators and grade levels/bands view the data collaboratively to discuss implications and areas of strength/growth.	50%
PD could be conducted by teachers scoring high in particular areas, with possible classroom demonstrations of best practices for visiting teachers.	50%
The results should be used primarily as a needs assessment for the larger PD efforts of a school/district. They should be part of a larger PD plan.	23%

Analysis of round three. The results from round three largely mirrored those of

round two, with the most popular choice again supporting the use of SETs to differentiate

PD initiatives for based on the needs of teachers. Those uses receiving majority approval

in round did so again in round three, with high-performing teachers conducting

demonstration lessons joining the ranks. The only item not receiving majority approval

involved using the results as a needs assessment at the site or district level. This suggests

that the group favored using SET results on a more localized level, with individuals and smaller groups looking at the data in order to plan PD more appropriate to individual needs. The twenty-three percent approval for using results at higher levels suggests that the group felt that data from SETs would be more useful at lower levels.

Weighting of SETs for Evaluation Purposes

Round One. The final aspect of SET use for evaluation purposes to be investigated involved their weighting in a teacher's evaluations. In an open-ended question, round-one participants were asked how heavy an influence SETs should have on a teacher's score: *If these surveys were to be used in the evaluation process, how much weight should they carry in the outcome (i.e., what percentage of a teacher's evaluation score could be based on student survey responses)?* The twenty-six responses were grouped into five possible weightings for the SETs.

Round Two. These five responses were included as a question in round two, where the twenty-four respondents were asked to choose which weighting they found appropriate: *If used for evaluation purposes, how much weight should they carry in a teacher's final evaluation?*

Table 22

Potential weighting for the results of SET surveys used for evaluation purposes, as reported by a panel of expert teacher trainers and administrators in round two

Potential weighting for SET results in a teacher's evaluation	Percentage of Respondents
No weight at all, but it could be a box in the teacher's evaluation	63%
5-10%	21%
20%	13%
30%	4%
50%	0%

98

Analysis of round two. The results of this question seem to confirm what was seen in Table 16, that the majority of the group felt that data from SETs should not be used as a factor in a teacher's evaluation score. The remaining participants set the weighing for SETs in an evaluation at no higher than thirty percent of a teacher's overall score, with the most popular weight being from five to ten percent, which was chosen by twenty-one percent of respondents. No respondents chose the option of giving SETs a weight of fifty percent of a teacher's score.

Round Three. In order to come closer to consensus, in the final round of the survey, participants were asked to look at the responses shown in Table 22 and again pick what they thought was the most appropriate weighting for SETs in a teacher's evaluation.

Table 23

Potential weighting for the results of SET surveys used for evaluation purposes, as reported by a panel of expert teacher trainers and administrators in round three

Potential weighting for SET results in a teacher's evaluation	Percentage of Respondents
No weight at all, but it could be a box in the teacher's evaluation	81%
5-10%	15%
30%	4%

Analysis of round three. The results of this round confirmed those of round two. Eighty-one percent of respondents opted to give SET results no weighting in a teacher's evaluation, with the remaining members choosing to give them either a five-to-ten percent weight or a thirty-percent weight. These findings, coupled with those earlier regarding the uses of SETs, suggest that the group sees the greatest benefits of SET use coming from their ability to inform the PD process rather than in their use in evaluations.

Emerging Themes of Research Question Three. The major themes emerging from the survey questions surrounding Research Question Three show a preference for SET use in PD rather than for evaluative purposes, for local rather than larger-scale

application, and for limited dissemination of individual teachers' results. When given the option to limit the use and weighting of SETs in teacher evaluations, a majority of participants consistently chose it. This is perhaps best demonstrated in eighty-one percent of participants preferring to give no weight to SET results on a teacher's evaluation. In both dissemination and use of results, the participants often chose options that kept individual teachers' results known only to the teachers and/or their administrators. The one instance where participants opted to have results known more widely concerned having teachers identified as successful in a given item giving demonstration lessons to others. Beyond that, the panel preferred a biannual implementation of SETs for PD purposes, with the results being shared anonymously with PLCs, departments, and staff in order to help them differentiate and inform PD events for individuals and small groups.

Additional Comments

Delphi panel members were also asked to reply to two more questions regarding the advantages and disadvantages of using SETs to inform PD and evaluation processes. Their replies were combined by the researcher and then sent out in Round Two, where participants chose those that they felt were appropriate. Although these responses were not used as a factor in answering the three Research Questions, they do raise interesting points about the perceptions of SETs. The results regarding the advantages of using SETs at the secondary level are seen in Table 24.

Table 24

Potential advantages of using SETs for PD or evaluative purposes at the secondary level, as reported by a panel of expert teacher trainers and administrators in round two.

Potential advantages to using SETs at the secondary level	Frequency of Response
Surveys provide a perspective that cannot be seen from observations and walk-throughs.	76%
Students are shown that their voices count.	69%
Professional development practices can be improved if teaching is examined as a two-way street: the instructor's knowledge meets the learner's needs.	65%
Positive data can give teachers clarity and confidence.	58%
There is accountability and perspective to the population actually being served by the teacher.	46%
Professional development choices will be based on student needs, not on the strengths of the teachers or the current trends at the district level.	38%
Students spend the most time with teachers, so their insights about their practice can be the most informed.	31%
Survey data tell an administrator if parent or student complaints are warranted and provides evidence for suggested teaching improvements.	19%

Analysis of Round Two. The results show that the study group's perceptions regarding the value of SETs at the secondary level match what research has shown: that students' perceptions are valid and reliable (Costin et al., 1971; Schmelkin et al., 1997), that SETs can validate student voices (Fielding, 2004; Jackson, 2004; Williams et al., 2012), and that can teachers benefit from reflecting on the results of student input (Fisher et al., 1995; Ferguson, 2010; Mertler, 1999).

The panel's list of disadvantages of using SETs at the secondary level was also interesting in that it showcased the negative perceptions of the group regarding SET use, despite the advantages expressed in Table 18. The process for eliciting and evaluating the potential disadvantages of SET use was the same as for the advantages. The group's perceptions are shown in Table 25.

Table 25

Potential disadvantages of using SETs for PD or evaluative purposes at the secondary level, as reported by a panel of expert teacher trainers and administrators in round two.

Potential disadvantages to using SETs at the secondary level	Frequency of Response
Surveys can be subjective, and the results can vary from day to day.	69%
Surveys can become a popularity contest, not a read reflection of teaching.	58%
Students can give higher marks in those classes they chose (electives, areas of interest) and lower marks in classes they're forced to take.	58%
It is nearly impossible to craft a multiple-choice survey that really encapsulates teacher performance.	46%
Students can give higher marks to teachers who give easier grades.	46%
Needs vary by class, so what works in one class may not be needed in another.	38%
Students can be nasty, and teachers don't like reading bad things about themselves.	23%
There is potential for abuse from those in power.	19%

Analysis of Round Two. Contrary to the previous section on the advantages of SET use, many of the perceived disadvantages expressed by the panel in Round Two run counter to the research. SET results are stable over time (MET Project, 2012). Students do not treat them like a popularity contest (Costin et al., 1971), nor do they rate their instructors based on grades received (Scheurich et al., 1983). There is, however, a small correlation between student ratings and whether a course was required or an elective choice (Costin et al., 1971). Studies have shown that it is possible to craft an instrument that accurately reflects a teacher's practice and provides actionable information (Kane & Staiger, 2012). Given the information in the last two tables, it appears that the panel recognizes what advantages SET use can bring to the PD and evaluation environments, but they are unaware of or unconvinced by the research refuting their misgivings about SET use at the secondary level.

Summary

Contained in chapter IV are the purpose of the study, the three research questions, the methodology, the population and sample, and the presentation of data aligned to each of the three research questions. Also included was additional research on the perceived advantages and disadvantages of SET use.

In round one of the Delphi study, participants were asked to identify possible questions for inclusion in a SET for evaluation or PD purposes at the secondary level. Twenty-seven of the thirty panel members responded, identifying fifty-one potential questions for use in a SET.

The unique responses to this question were collected and combined into forty-nine potential questions by the researcher, and these became the basis for round two of the study, where participants were asked to rank each on a Likert scale according to their importance for inclusion in a SET. They were also asked to weigh in on the appropriate length for such a survey. Twenty-four of the twenty-seven panel members responded to this round of the survey, and the results were used to provide a preliminary ranking for the forty-nine SET question items and possible ranges in the number of items to be included in such an instrument.

A second set of questions was sent out to participants in round two, concerning the administration, audience, and use of SETs. Participants were asked to identify how often and when SETs should be implemented, how their results should be disseminated, how much weight the responses should have in evaluations, and how the results should be used to inform PD practices. The advantages and disadvantages of SET use were also

elicited from participants. Twenty-six of the twenty-seven participants responded to this round of the survey.

In round three, panel members were provided with the initial rankings of the potential SET questions and asked how each should be moved in the rankings. They were also provided with the initial results of the second round-two survey and asked to weigh in on all questions asked. Twenty-six of the twenty-seven panel members responded. The researcher reviewed the responses, analyzing the data and presenting the emerging themes through narrative and tables corresponding to each of the research questions. Analysis of the discussion of the advantages and disadvantages of SET use at the secondary level was also provided.

Chapter V presents conclusions, implications, and recommendations for future research.

CHAPTER V: FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The study examined the perceptions of master teachers, teacher trainers, and administrators regarding the use of SETs (Student Evaluation of Teachers) at the secondary level. This study also sought to identify the most important elements for inclusion in the construction of SETs. In addition, it intended to determine how the results of SETs could best be used by teacher trainers and administrators to inform evaluation and professional development practices for secondary teachers.

Chapter I of this study provided background about current attitudes and procedures towards the use of SETs, evaluations, and professional development (PD) and an introduction to the research study. Chapter II presented a review of literature about andragogy, the history and attitudes towards SETs, current evaluation and PD practices in the US, and student voice.   Chapter III explained the research design and methodology of the study, including population, sample, instrumentation, data collection, and analysis procedures. Chapter IV provided a brief description of the research design, population, sample, and data collection and analysis procedures. Data was presented aligned to each research question, grouped by rounds of the Delphi study. Chapter IV concluded with a summary of findings.

Chapter V presents an overview of the study, which includes the purpose, research questions, and methodology. A summary of major findings is presented, followed by conclusions, recommendations for further research, and concluding remarks and reflections.

Purpose Statement

The purpose of this Delphi study was to identify the most important elements for SETs (Student Evaluation of Teachers) at the high school level as perceived by a panel of expert master teachers, administrators, and teacher trainers. In addition, it was the purpose to determine how the results of SETs could best be used by teacher trainers and administrators to inform evaluation and professional development practices for secondary teachers.

Research Questions

The following questions were investigated to address the purpose of the study:

1. What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

2. How do the panel of master teachers, administrators and teacher trainers rank the elements of SET?

3. What do a panel of master teachers, administrators, and teacher trainers identify as strategies for using the data from SETs to inform evaluation and professional development for secondary teachers?

Methodology

This study utilized the Delphi method to elicit perceptual data from an expert panel of master teachers, pre- and in-service teacher trainers, and administrators. Electronic questionnaires were used to assess the perceptions of respondents about the content and use of SETs at the secondary level. These questionnaires were administered

in three rounds, with the second round divided into two parts to ease processing of the large number of responses in round one.

Twenty-six of the panel members (87%) responded to the electronic questionnaire for the first round of the Delphi study. Results of study participants' responses were analyzed by the researcher and became the basis for the two parts of the second round of the study. Twenty-four of the thirty panel members (80%) responded to the electronic survey for both parts of the second round. As with round one, these responses became the basis for the third round of the survey. In the third round of the survey, the two panelists not completing the second half of the round two survey came back to complete the final survey, bring the number of respondents back to twenty-six (80% of the initial panel members).

Major Findings

The findings related to each of the three research questions are presented here, along with associated emerging themes. These results are divided by research question, with the findings and emerging themes presented sequentially by survey round.

Research Question One

What do a panel of master teachers, administrators, and teacher trainers identify as important elements of Student Evaluation of Teachers (SETs) at the high school level for secondary teachers?

Round One. Thirty panel members were asked by electronic questionnaire to answer the question *If high school students were being surveyed about their teacher's work in their class, and that information might be used for evaluation or professional development purposes, what should we be asking about the teachers?* Respondents were

also asked to include any justification they felt appropriate for their answers and, if they wanted, to include the actual questions they would want on the survey. From the forty-nine questions culled from the panelists' responses, three main areas for survey questions emerged.

Research Question One Findings. A majority of the questions dealt with the teacher's attitudes and behaviors, eliciting answers on topics such as approachability, frequency and effectiveness of feedback, content knowledge, and preparedness. These were followed in frequency by questions about the activities and content of the course, such as the amount of group work and peer feedback, the relevance of the material covered to real-world situations, and the use of multiple media and technology in lessons. There was marked congruence (see Table 11) between the survey questions generated by the panel and John Hattie's *Visible Learning* (Hattie, 2009) and the California Standards for the Teaching Profession ("CSTPs," 2018). Finally, a few questions concerned affective factors in the classroom, with questions asking about how students felt about the material before and after instruction, how welcomed students felt by the teacher, and how much they felt they learned. With a list of potential questions generated, it was now time to have the study group determine the ranking of the potential test items.

Research Question Two

How do the panel of master teachers, administrators and teacher trainers rank the importance of the elements of SETs?

Round Two. For the second round of the survey, panelists were asked to rank each of the forty-nine questions on a six-point Likert scale with the following prompt: *Below you'll see the range of answers generated in the first round of the study. For each,*

please rate how important you feel this item would be to include in a Student Evaluation of Teachers (SET) at the secondary level. The twenty-four respondents ranked each item, and the mean scores ranged from 5.38 ("Does your teacher give you effective feedback on your work in a timely manner?") to 2.76 ("Is your teacher flexible?"). From these scores, the researcher was able to rank the forty-nine questions and place them into quartiles.

The panelists were also asked a question concerning the size of a potential survey: *Given that we have about 50 possible survey items here, we also need to think about how large the SET should be. Thinking about both manageability and thoroughness, how many items do you feel should be on this survey?* The open-ended question yielded a range of answers from five up to fifty.

Round Three. The panelists were presented with the rankings of potential SET questions and asked to state whether each item should remain in its current place or be moved up or down in the quartiles. In all but four cases, the majority opinion favored keeping an item in its current ranking. Three items were moved up in the quartiles, and one was moved down. This resulted in a list of forty-nine ranked questions for possible inclusion in a SET.

The study group was also shown the member's preferences for the size of the survey, including how many voted for each range. A shift occurred from the first round, with the most panelists voting for a survey containing between twenty and thirty questions, and no one opting for a survey containing more.

Research Question Two Findings. An emerging trend in the rankings had panelists tending to give preference to questions concerning a teacher's attitudes and

behaviors, with the lower quartiles concentrating on the activities and content of a course. Given that the Delphi group preferred a survey in the twenty-to-thirty-question range, items in the higher quartiles would presumably be given priority from among the forty-nine choices. This suggests that the panel wanted the surveys to concentrate more on what teachers were doing in the classroom than on what they were teaching. With the content of SETs established, the remaining questions concerned the timing, frequency, audience, weighting, and uses of such surveys.

Research Question Three

What do a panel of master teachers, administrators, and teacher trainers identify as strategies for using the data from SETs to inform evaluation and professional development for secondary teachers?

Round One. In the first round of the Delphi study, respondents were asked a number of open-ended questions regarding the administration and use of SETs. Regarding the use of SETs for professional development purposes, questions were asked concerning the possible timing and frequency of administration, as well as how they might be used to inform PD practices. An additional question elicited responses on the potential audience(s) for SET survey results. For SET use in evaluations, timing and frequency were also investigated, as well as the potential weighting of SET survey data in a teacher's final evaluation score. A final set of questions asked panelists to provide their perceptions of the potential advantages and disadvantages of SET use at the secondary level. The results of each of these questions were analyzed by the researcher and formed the basis for the second round of surveys on research question three.

Round Two. Survey questions in this round were based in both form and content on the responses given by respondents in round one. Panelists were also, where applicable, provided with the anonymous feedback justifying responses from the earlier round.

Round Three. The final round of survey questions continued in the model of round two, with participants responding to questions containing the data and feedback from the previous round. In most areas, round two results were confirmed, with the panelists' views coming closer to consensus.

Research Question Three Findings. For SET use in PD, a number of trends emerged. The panel preferred multiple administration of surveys, with the dates falling at the ends of semesters or at key benchmark points in the curriculum. It also opted to make the audience for survey results individuals and their administrators having access to disaggregated results, with PLCs and departments possibly seeing the results for the given group. Panelists generally selected a more local dissemination and use of survey results, especially in differentiating and planning the content of PD practices for individuals and groups.

For SET use in teacher evaluations, the group continued to support multiple administrations throughout the year. They also preferred to give SET results little, if any, weight on a teacher's formal evaluation.

Additional Survey Results

Panelists' perceptions of SET use were also elicited over the course of the study. Participants were initially asked open-ended questions about the advantages and disadvantages of using SETs for evaluation and PD purposes. Their responses were then

conflated into a list for round two, from which panelists chose those items they felt were most pertinent.

From the data, it is evident that the most popular responses concerning the advantages of using SETs to inform PD and evaluation dealt with what students' perceptions could add to the process. Panelists recognized that students can add perspectives not seen using current teacher observation practices. Affective factors also came up, with panelists acknowledging how eliciting student opinions can be motivating for students as their voices are being heard, and also for teachers as the positive aspects of their work are confirmed. Also noted was the capacity for improving PD practices because SETs would function as a needs assessment, allowing schools to focus on areas requiring improvement rather than relying on current educational trends or uninformed choices. These responses were expected, as they confirmed much of what was said in the research about the positive effects of eliciting student opinions (Ferguson, 2012; Jezequel, 2008).

The panelists' responses regarding the disadvantages of using SETs to inform PD and evaluation, however, raised questions about the continuing negative perceptions of students' ability to be impartial. Contrary to the findings in the literature regarding the reliability and validity of student opinion (Colorado Legacy Foundation, 2013; Elbow & Boice, 1992; Mertler, 1999), the three responses most often chosen by the panel involved the subjectivity and variability of student responses, and students' propensity to treat the process as a popularity contest or to award higher rankings based on their grades or whether the class was required or an elective. This finding explains the trend in the study data for the group's preferences that SET results remain largely anonymous and that

112

survey results be noted but not weighted in evaluations. The conclusion from this is that as long as teachers and administrators harbor doubt about their students' abilities to provide reliable and unbiased data on teachers' practices, they will be reluctant to give full credence to survey results.

Unexpected Findings

As noted above, the panel expressed a reluctance to give SETs much if any weight in a teacher's formal evaluation. This finding, along with the disadvantages of SET use brought up in the final survey questions, suggests that despite a century of use at the tertiary level and much research to the contrary, SETs are still perceived as being potentially unreliable as a credible source of information on a teacher's practice at the secondary level. For any institution wanting to implement SET use at this level, serious consideration must be given to developing processes to alleviate staff concerns about issues like bias, variations due to course content and grading, and potential misuse or unwarranted dissemination of survey data by administrators.

Conclusions

This study was designed to gain insight into what the content of secondary student evaluations of their teachers should be. It also sought to find out how a panel of educational experts would rank that content. Next, it attempted to discover how SETs could best be used to inform professional development and evaluation practices. Finally, it sought to elicit the perceptions of teachers and administrators regarding SET use. Based on the findings and literature review, several conclusions can be drawn regarding the design and use of SETs at the secondary level. Successful SET use in informing

113

professional development and evaluation practices is dependent on prioritization and focus in the following areas:

1. Based on the research and study results, current evaluation practices do not provide the substance and specificity needed for teachers to raise classroom achievement. Where effective evaluation practices must result in information on current difficulties and viable paths for improvement (Darling-Hammond et al., 2012), in some studies two-thirds of teachers undergoing evaluation received no specific feedback on how to improve their classroom performance (Weisberg et al., 2009). As long as there is little connection between evaluations and professional development initiatives, the current system of teacher development will continue to show minimal improvement in teaching practices. Studies have shown, however, that students can provide valuable and valid input regarding a teacher's classroom performance and practices (Burniske & Meibaum, 2012; Ferguson, 2012). Incorporating student evaluations of teachers into the evaluation process will provide all stakeholders with useful data for improving individual and site- and district-wide classroom performance.

2. Based on the research and study results, the current lack of focus and coherence in PD practices at the secondary level results in ineffective PD experiences for secondary teachers. While effective professional development initiatives link teachers' evaluation data and developmental needs to training initiatives (Darling-Hammond et al., 1983; Fogerty & Pete, 2009), too often professional development decisions are arbitrary in

114

nature, with little connection to actual teacher needs (Kelleher, 2003). In

many cases, the connection between evaluation data and the focus of

professional development is tenuous (Stecher et al., 2012; Webster-

Wright, 2009), leading to professional development initiatives that are

unfocused and of low quality (Desimone, 2011; Royce, 2010). SET use is

needed to inform and improve these efforts by providing both valuable

and actionable information for targeted professional development to

decision makers and material for self-reflection for teachers undergoing

the process.

3. Teachers need feedback that focuses their reflection on the effects of their

actions and affect in the classroom. Because each lesson and class are

different, teachers need more than just a list of best practices to implement

universally (Hattie, 2009); rather, they need ongoing feedback on their

choices in the classroom (TFEE, 2012). Studies have shown that frequent

and targeted feedback for teachers leads to increased student achievement

as they continually question habitual patterns of activity and thinking

(Webster-Wright, 2009). Because teaching is a multifaceted endeavor,

any survey attempting to capture a teacher's practice will need to be

equally multifaceted. In order to capture this complexity, the study found

that SETs used to inform PD and evaluation processes should focus on

three main areas, in order of importance: what the teacher does in the

classroom, how the students feel about themselves and their learning in the

class, and what activities and content are being used in the teaching process.

4. Teachers must receive feedback on their classroom management strategies and actions in order to improve their teaching practices. While teachers would benefit more from feedback regarding how their behaviors affect students and the classroom atmosphere (McMillan & Schumacher, 2010). current evaluative practices focus mainly on the activities and content being used in the classroom (Webb, 1995) Therefore, when deciding on which items to use in a SET, preference should be given to questions dealing with a teacher's actions and affect in the classroom. This was confirmed in the study, where of the twelve questions ranking in the first quartile, nine of them deal with a teacher's actions, competencies, and abilities.

5. For professional development initiatives and teacher reflection to be effective, teachers and administrators require concise, timely, and actionable information on classroom practices. Unfortunately, current evaluation practices provide very little actionable feedback to teachers and administrators (TFEE, 2012; Weisberg et al., 2009). Student evaluations can provide teachers and administrators with timely and meaningful data on the aspects of teachers' classroom practices that should be targeted for improvement and development (Hanover Research, 2013; Youngs, 2013). Therefore, based on study results, useful and actionable data for teachers and administrators can be effectively obtained through the use of student

evaluations of teacher that are implemented multiple times during the school year, either at the ends of semesters or at strategic times (e.g., at benchmark points, close to major breaks, or at the ends of teaching units). These SETs should contain between ten and thirty questions. Individual teachers and their administrators should review the disaggregated results, while PLCs, departments, and whole staffs should look at aggregated and anonymous data to determine where PD efforts should be concentrated. These results will allow for increased differentiation to meet individual teachers' needs. The analysis of data should also lead to individual teachers being asked to conduct PD efforts because of their demonstrated success in certain areas.

6. When SETs are weighted in the formal evaluation process, they focus teachers' attention on compliance and lose their power to cultivate authentic reflection on how to improve practice. In fact, faculty resistance to student evaluations tends to focus on their formal inclusion in evaluations (Schmelkin et al., 1997). Student evaluations can be used more effectively in an unofficial, unweighted manner, with the resulting data being used to promote individual reflection (Elbow & Boice, 1992). Therefore, based on study results, if SETs are used in the formal evaluation process, data should be used to inform the reflective process but not be weighted in a teacher's formal evaluation. As with SET use for PD purposes, multiple implementations should be conducted, either at the end of each semester or coinciding with the evaluation cycle.

7. While there is currently widespread resistance among secondary school teachers to the implementation of student evaluations, this can be countered. If the processes are explained well and understood by teachers, they are more likely to be respected and accepted, especially if they are seen as a mechanism for schoolwide improvement (Goe et al., 2008). Any secondary school or district wanting to implement SETs for evaluation or PD purposes must address and counter negative views towards student evaluations as they introduce the process to their staff.

Recommendations for Action

If California educators are to become more informed and reflective practitioners of their craft, it is vital that they be given effective feedback on their attitudes and actions in the classroom. Current professional development and evaluation practices are ignoring a vital source of information on what teachers do in their classrooms each day: the students in those classrooms. As the observers and recipients of teachers' efforts for hundreds of hours each year, students are best situated to provide valid and reliable input regarding what their teachers do well and where they need to be concentrating their professional development efforts. To that end, a number of recommendations are being made:

1. Secondary students should be completing SETs in all of their classes, multiple times each year. These SETs should include Likert-scale and open-ended questions about the teacher's affect and actions, the classroom atmosphere, and the content and activities of the course.

2. When SETs are first rolled out at a school, it is recommended that all stakeholders be involved in the process. Administrators must anticipate the staff's potential objections to their use and provide training that highlights the reliability and validity of student views. Students will require training in completing the surveys, and particular attention will need to be paid to ensuring that they understand what each survey item is assessing. Ownership from the all stakeholders can be ensured by letting each group have a voice in which of the potential survey items are included in the final instrument. This also allows for the foci of the surveys to change over time as the staff continue to hone their craft.

3. The data obtained from these surveys should be used by individual teachers in reflecting upon their practices, either in isolation or in conference with their master teacher and/or administrator. Individual teachers will use the data to continuously reflect upon and improve all aspects of their craft.

4. When SETs are to be used in the formal evaluation process, they should hold little if any weight in a teacher's final score. That being said, the results should still be included in the evaluation, and the evaluator should conference with the teacher regarding the input provided by the students and the implications for further professional development. Though the data would not hold weight in the formal evaluation, they would still have a significant impact on teacher development.

5. The data obtained from these surveys should be used by PLCs and departments in order to highlight trends and inform collective professional development efforts. The success of professional development efforts will be monitored through analysis of SET and student achievement data.

6. The data obtained from these surveys should be used by schools to determine the foci of professional development efforts. When particular areas for improvement are identified, local teachers with demonstrated skill in the particular areas, as shown by SET results, should be encouraged to spearhead PD initiatives in those areas.

7. The data obtained from these surveys should be used by districts to report Dashboard data as it relates to the LCAP, thus making aggregated student survey data available and transparent to ensure public accountability.

The process of how a SET might be implemented in a California secondary school is currently being explored in the researcher's school in Visalia, California. While still in its nascent stages, it could provide a model for other institutions wishing to follow suit. Here is a brief explanation of what is being attempted.

The idea for using a SET at the researcher's secondary school was presented first to the school's Committee for School Improvement, a voluntary weekly before-school meeting hosted by the principal and attended by members of the staff and administration wishing to discuss possible actions to be taken to improve their school environment. This venue was chosen for introducing the idea of using a SET on campus because those attending these volunteer meetings were among the most involved and influential adults on campus. The researcher presented the preliminary findings of the Delphi study,

including the ranked list of potential questions that were generated. The committee expressed their desire to implement a site-generated SET in all classes. As a precautionary measure, the local teacher union was then consulted, and legality of the use of a SET for professional development purposes was confirmed.

As student input was also desired, students from five student homerooms, one homeroom from each grade level and one multi-grade homeroom, were given the forty-nine questions and asked to choose the twenty they felt most strongly about wanting to use in offering feedback to their teachers. As was the case with the Delphi group, the students gave emphasis to questions that allowed them to comment on their teachers' affect and actions in the classroom.

At the next staff development meeting, the principal introduced the idea of surveying students to all the teachers, being careful to couch it in terms of professional development and not evaluation, and extolling the benefits of he himself having undergone a 360-degree peer evaluation recently. The survey items were not discussed at that time.

Moving forward, the plan is to introduce the forty-nine questions at the next monthly staff development meeting and have the teaching staff choose which items they would like to see on a site-specific SET for professional development purposes. These results will be compared with the student-generated results, and the Committee for School Improvement (CSI) will then prepare a SET for end-of-year implementation in all classes. The results of the SET will be analyzed by CSI members over the summer, and the topics for the school's fall semester professional development initiatives will be informed by this analysis. Aggregated results will also be passed on to department heads

121

for distribution among PLCs as they start their work in the new school year. The CSI will then convene to analysis the process and the results as they prepare to repeat a SET implementation in the spring semester.

Recommendations for Future Research

Although SETs have been used in colleges and universities for over a century, they are still a relatively new factor at the secondary school level and below. As such, there are still several areas demanding further research:

1. In a state like California, where many districts contain strong collective bargaining units for teachers, how much of an effect will these units have on teachers being receptive to implementing SETs?

2. While the validity and reliability of secondary students' opinions is well documented in the literature, far less is known about younger students. Can students in elementary and middle schools also provide effective feedback on their teachers?

3. If students can provide effective feedback on secondary school teachers, could not these same teachers provide effective feedback on their administrators? What are the potential advantages and disadvantages of having secondary school teachers fill out surveys on their administrators?

4. What is the correlation between SET use for professional development purposes and student achievement data?

5. Once areas for improvement are identified through the use of SETs, how can they best be addressed? Should professional development efforts be led by local experts from within or by hired experts from without? If from

within, what positive effects would this have on teacher self-actualization?

6. What are the concrete benefits in student motivation to having them take part in the creation of the SET instrument? In other words, what benefits would accrue if students not only completed the surveys, but also helped choose which questions were asked?

7. Should a school use a single instrument when implementing SETs, or would each department/PLC/group benefit from choosing the specific items to be included in their SET?

Concluding Remarks and Reflections

James A. Belasco once said, "Evaluate what you want - because what gets measured, gets produced" ("Belasco quotes," 2016). This sentence reminds me of how ineffective the evaluations I have undergone over the past twenty years of teaching have been for me. Every time I go through the two- or three-year cycle of having someone come into my classroom to watch me teach for an hour and then fill out a prescribed form about the experience, the result is always the same: I am given a clean bill of health and told that I should keep on doing what I am doing. While it is always gratifying to hear that I am doing my job well, it is also frustrating because I am never given anything useful to help me improve my practice. From Belasco's point of view, the evaluations being done in my class only serve to perpetuate the status quo. What my administrators and I should really be doing is getting useful input from the people who are in a far better position to help me improve as a teacher, the students. Until collecting their voices is part of the process, we will never be getting the full picture of what is going on in my

classroom. And until that full picture is seen, we will never have a clear focus for my professional development efforts.

Conducting this study has been a life-changing process for me. Seeing the diversity of opinion on educational topics through these years of research, I have come to realize just how divided we educators are about what really works in the classroom. When standards and processes change with every trend or administration, it is easy to see why teachers view progress in education like they view the weather in Texas: if you dislike what is happening at the moment, just wait five minutes for it to change. At the same time, I am heartened by the potential I see for transforming classroom practice by trying something as simple and obvious as incorporating student voices into the process. Still, it is a little daunting to consider doing so.

As a reflective practitioner of my trade, I know it is not always easy to hear criticism about something I spend so much time and effort trying to improve. Part of me is still afraid of what I will hear from my students if I ever put the forty-nine questions develop by the Delphi panel in front of them. Before starting this study, I was completely unaware of just how valid the perspectives of my students can be. I know it will still feel risky to open myself up to the honest opinions of the two hundred teenagers I interact with every day, but I also know that doing so is necessary and useful and will be a solid step in improving the educational practices for myself, my department, and my school.

REFERENCES

A new way to rate LA Unified's teachers. (2012). Retrieved from

    http://articles.latimes.com/2012/dec/05/opinion/la-ed-teacher-evaluation-

    20121205

Algozzine, B., Beattie, J., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Sponner,

    F. (2004). Student evaluation of college teaching: A practice in search of

    principles. Retrieved from http://www.jstor.org/stable/27559201

Archer, J., Kerr, K. A., & Pianta, R. C. (2014). Why measure effective teaching? In T.

    Kane, K. A. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems* (pp.

    1-8). Retrieved from http://www.general-ebooks.com/book/185695980-

    designing-teacher-evaluation-systems

Association for Supervision and Curriculum Development. (2005). Effective

    professional development. Retrieved from

    http://webserver3.ascd.org/ossd/planning.html

Ball, D., & Cohen, D. K. (1999). Developing practice, developing practicioners:

    Toward a practice-based theory of professional education. In G. Sykes & L.

    Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of*

    policy and practice (pp. 3-32). Retrieved from http://www-

    personal.umich.edu/~dball/chapters/BallCohenDevelopingPractice.pdf

Battey, D., & Franke, M. L. (2008, Summer 2008). Transforming identities:

    Understanding teachers across professional development and classroom practice.

    Teacher Education Quarterly, *35*(3), 127-149. Retrieved from

    http://www.jstor.org/stable/23478985

Beyers, C. (2010, 07 Aug 2010). The hermeneutics of student evaluations. *College Teaching*, *56*, 102-106. http://dx.doi.org/10.3200/CTCH.56.2.102-106

Bias in survey sampling. (2015). Retrieved from http://stattrek.com/survey-research/survey-bias.aspx

Bias in survey sampling. (2015). Retrieved from http://stattrek.com/survey-research/survey-bias.aspx

Boulton-Lewis, G. M., Wilss, L., & Mutch, S. (1996, July). Teachers as adult learners: Their knowledge of their own learning and implications for teaching. *High Education*, *32*(1), 89-106. Retrieved from http://www.jstor.org/stable/3447898

Brown-Easton, L. (2008, June). From professional development to professional learning. *The Phi Delta Kappan*, *89*, 755-759. Retrieved from http://www.jstor.org/stable/40792272

Burniske, J., & Meibaum, D. (2012). The use of student perceptual data as a measure of teaching effectiveness. Retrieved from http://txcc.sedl.org/resources/briefs/number_8/index.php

California Department of Education. (2010). *2009–10 Highly Qualified Teacher Data* [Excel file]. Retrieved from http://www.cde.ca.gov/nclb/sr/tq/schlstfrpt.asp

California Department of Education. (2016). Local control funding formula: Information about the funding and accountability provisions of the Local Control Funding Formula. Retrieved from http://www.cde.ca.gov/fg/aa/lc/lcffoverview.asp

California standards for the teaching profession (CSTP) 2009. (2018). Retrieved from https://www.scribd.com/document/345792715/cstps-at-a-glance-2009

126

California Teachers Association. (2011). Teacher development & evaluation principles. Retrieved from http://www.cta.org/Issues-and-Action/Teacher-Quality/Teacher-Evaluation-Principles.aspx

Center for Research on Learning and Teaching. (2014). Guidelines for evaluating teaching. Retrieved from http://www.crlt.umich.edu/tstrategies/guidelines

Chalwa, V., & Thurkal, P. (2011). Effects of student feedback on teaching competence of student teachers: A microteaching experiement. *Contemporary Educational Technology*, 2(1).

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. Retrieved from http://obs.rc.fas.harvard.edu/chetty/value_added.html

Christopherson, K., Elstad, E., & Turmo, A. (2010). Is teacher accountability possible? The case of Norweigian high school science. *Scandinavian Journal of Educational Research*, *54*(5). http://dx.doi.org/10.1080/00313831.2010.508906

Colorado Legacy Foundation. (2013). Positioning students as experts on instructions: An analysis of open-ended responses from the student perception survey. Retrieved from http://colegacy.org/news/wp-content/uploads/2013/09/SPS-Technical-Report-FINAL_final.pdf

Colorado Legacy Foundation. (2013). Positioning students as experts on instructions: An analysis of open-ended responses from the student perception survey. Retrieved from http://colegacy.org/news/wp-content/uploads/2013/09/SPS-Technical-Report-FINAL_final.pdf

Commission on Effective Teachers and Teaching. (2012). *Transforming teaching: Connecting professional responsibility with student learning*. Retrieved from http://www.nea.org/assets/docs/Transformingteaching2012.pdf

Cook-Sather, A. (2006, Winter). Sound, presence, and power: "Student voice" in educational research and reform. *Curruculum Inquiry*, *36*, 359-390. Retrieved from http://www.jstor.org/stable/4124743

Costin, F., Greenough, W. T., & Menges, R. J. (1971, December 1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, *41*, 511-535. Retrieved from http://www.jstor.org/stable/1169890

Crow, T. (2011, December 2011). The view from the seats: Student input provides a clearer picture of what works in schools. *Journal of Sustainable Development*, *32*(6), 24-30. Retrieved from http://learningforward.org/docs/jsd-december-2011/december2011jsdlearningguide.pdf?sfvrsn=2

Darling-Hammond, L., Amrein-Beardsley, A., Hartel, E., & Rothstein, J. (2012, March 1). Evaluating teacher evaluation. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2012/03/01/kappan_hammond.html.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983, Autumn 1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, *53*(3), 285-328. Retrieved from http://www.jstor.org/stable/1170367

Desimone, L. M. (2011, March). A primer on effective professional development. *The Phi Delta Kappan*, *92*(6), 68-71. Retrieved from http://www.jstor.org/stable/25822820

Dillon, S. (Dec 10, 2010, December 10). What works in the classroom: Ask the

    students. *New York Times*, p. A15. Retrieved from

    http://www.nytimes.com/2010/12/11/education/11education.html

Dozier, C. (2012). Religious private high school students' perceptions of effective

    teaching. *Journal of Instructional Pedagogies*, *9*(1).

Dresel, M., & Rindermann, H. (2011, November 2011). Counseling university

    instructors based on student evaluations of their teaching effectiveness: A

    multilevel test of its effectiveness under consideration of bias and unfairness

    variables. *Reserach in Higher Education*, *52*, 717-737. Retrieved from

    http://www.jstor.org/stable/41483813

Education code section 44660-44665. (2005). Retrieved from

    http://law.onecle.com/california/education/44664.html

Elbow, P., & Boice, R. (1992). Making better use of student evaluations of teachers.

    Profession, 42-48. Retrieved from http://www.jstor.org/stable/25595488

Fenwick, T. (2006). Work, learning, and education in the knowledge economy: A

    working-class perspective. *Curriculum Inquiry*, *36*, 454-466. Retrieved from

    https://www.academia.edu/12456640/Work_Learning_and_Education_in_the_Kn

    owledge_Economy_A_Working-Class_Perspective

Ferguson, R. F. (2010). Student perceptions of teaching effectiveness. Retrieved from

    http://www.gse.harvard.edu/ncte/news/Using_Student_Perceptions_Ferguson.pdf

Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta

    Kappan*, *94*(3), 24-28. Retrieved from

    http://jeb.sagepub.com/content/39/5/394.full.pdf+html

Fielding, M. (2004, November 2004). 'New wave' student voice and the renewal of civic society. *London Review of Education*, *2*. http://dx.doi.org/DOI: 10.1080/1474846042000302834

Fingertip facts on education in California. (2016). Retrieved from http://www.cde.ca.gov/ds/sd/cb/ceffingertipfacts.asp

Fisher, D., Fraser, B., & Cresswell, J. (1995). Using the "Questionnaire on Teacher Interaction" in the professional development of teachers. *Australian Journal of Teacher Education*, *20*(1), 7-19. Retrieved from http://dx.doi.org/10.14221/ajte.1995v20n1.2

Fogerty, R., & Pete, B. (2009, Dec., 2009 - Jan., 2010). Professional learning 101: A syllabus of seven protocols. *The Phi Delta Kappan*, *91*, 32-24. Retrieved from http://www.jstor.org/stable/25594677

Forrest, III, P., & Peterson, T. O. (2006, Mar. 2006). It's called andragogy. *Academy of Management Learning & Education*, *5*(1), 113-122. Retrieved from http://www.jstor.org/stable/40212539

Fulmer, G. (2013). Measuring model-based high school science instruction: Development and application of a survey. *Journal of Science, Education, and Technology*, *23*(), 37-46. http://dx.doi.org/DOI 10.1007/s10956-012-9374-z

Gay, L. R., & Airasian, P. (1996). *Educational research*. Columbus, OH: Merrill Prentice Hall.

Gentile, M., & Pisanu, F. (2014, June). Considering the student voice. *Recercazione*, *6*(1), 18-25. Retrieved from https://www.academia.edu/7669133/Considering_the_student_voice

Glickman, C., Gordon, S., & Ross-Gordon, J. (2010). *Supervision and instructional leadership: A developmental approach* (8 ed.). Boston, MA: Pearson.

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf

Grooters, G. S. (2008). Secondary student and teacher perceptions of building relationships for personalizing education. Retrieved from http://search.proquest.com/docview/304536472?accountid=

Guskey, T. R., & Yoon, K. S. (2009, March 2009). What works in professional development? *Phi Delta Kappa*, *90*, 495-500. Retrieved from http://www.jstor.org/stable/20446159

Hallowell, M. (2009, April). *Techniques to minimize bias when using the Delphi technique to quantify construction safety and health risks*. Paper presented at the University of Colorado at Boulder, Boulder, Co. Retrieved from https://www.researchgate.net/publication/269124457_Techniques_to_Minimize_Bias_When_Using_the_Delphi_Method_to_Quantify_Construction_Safety_and_Health_Risks

Hanover Research. (2013). Student perception surveys and teacher assessments. Retrieved from http://dese.mo.gov/sites/default/files/Hanover-Research-Student-Surveys.pdf

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

131

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.

Helding, K., & Frasier, B. (2013). Effectiveness of national board certified (NBC) teachers in terms of classroom environment, attitudes, and achievement among secondary science students. *Learning Environment Research*, *16*(), 1-21. http://dx.doi.org/DOI 10.1007/s10984-012-9104-8

Hibler, W. D., & Snyder, J. A. (2015, Spring 2015). Observations on teacher evaluations. *Teaching Matters*, *12*(1), 33-47. Retrieved from http://www.jstor.org/stable/10.1086/680693

Hill, H. C. (2009, March 2009). Fixing teacher professional development. *Phi Delta Kappan*, *90*, 470-476. Retrieved from http://www.jstor.org/stable/20446155

Hirsh, S. (2007, April 1). NSDC standards and tools help strengthen professional development. *SEDL Letter*, *19*(1). Retrieved from http://www.sedl.org/pubs/sedl-letter/v19n01/nsdc-standards-tools.html

Hsu, C., & Sanford, B. A. (2007, August 2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, *10*(10). Retrieved from http://pareonline.net/getvn.asp?v=12&n=10

Isaac, S., & Michael, W. B. (1981). *Handbook in research and evaluation*. San Diego, CA: EdITS Publishers.

Jackson, D. (2004). Why pupil voice? Retrieved from http://culturelanguage.wikispaces.com/file/view/nexus-se-pnsln-why-pupil-voice.pdf

James A. Belasco quotes. (2016). Retrieved from

      http://thinkexist.com/quotes/james_a._belasco/

Jensen, B., Sonnemann, J., Roberts-Hall, K., & Hunter, A. (2016). Beyond PD: Teacher

      professional learning in high-performing systems. Retrieved from

      http://www.ncee.org/beyondpd/

Jezequel, J. L. (2008). *The impact of student evaluation of teachers on teacher*

      practices in a secondary school (Doctoral dissertation, Northcentral University).

      Retrieved from http://search.proquest.com/docview/304823465?accountid=10051

Johnson, B. (2012). Should teachers evaluate their teachers? Retrieved from

      http://www.edutopia.org/blog/student-evaluation-teachers-ben-johnson

Kane, T. J., & Cantrell, S. (2010). *Learning about teaching: Initial findings from the*

      Measures of Effective Teaching project. Retrieved from GatesFoundation.org:

      http://www.teachingquality.org/sites/default/files/Preliminary_Finding-

      Policy_Brief_.pdf

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining

      high-quality observations with studetn surveys and achievement gains. Retrieved

      from http://eric.ed.gov/?id=ED540961

Keane, E., & MacLabhrainn, I. (2005). Obtaining student feedback on teaching and

      course quality. Retrieved from

      www.nuigalway.ie/celt/documents/evaluation_ofteaching.pdf

Kelleher, J. (2003, June). A model for assessment-driven professional development.

      The Phi Delta Kappan, *84*, 751-756. Retrieved from

      http://www.jstor.org/stable/20440476

Kelly, M. (2015). State versus national standards. Retrieved from

http://712educators.about.com/od/curriculumandlessonplans/a/standards.htm

Kember, D., & Wong, A. (2000). Implications for evaluation from a study of students'

perceptions of good and poor teaching. Retrieved from

http://www.jstor.org/stable/3447952

Knowles, M., Holton, E., & Swanon, R. (2005). *The adult learner* (6 ed.). Burlington,

MA: Elsevier.

Lacireno-Paquet, N., Bocala, C., & Bailey, J. (2016). *Relationship between school*

professional climate and teachers' satisfaction with the evaluation process.

Retrieved from

http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2016133.pdf

Lawson, A., Leach, M., & Burrows, S. (2012). The implications for learners, teachers,

and institutions of using student satisfaction as a measure of success: a review of

the literature. *Education Journal*, *138*.

Lester, J. H. (2003, Fall 2003). Planning effective secondary professional development

programs. *American Secondary Education*, *32*(1), 49-61. Retrieved from

http://www.jstor.org/stable/41064504

Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher

effectiveness. Retrieved from

http://www.tqsource.org/publications/practicalGuide.pdf

Local control funding formula overview. (2016). Retrieved from

http://www.cde.ca.gov/fg/aa/lc/lcffoverview.asp

Massachusetts Department of Elementary & Secondary Education. (2013). Education

    laws and regulations: Final regulations on evaluation of educators. Retrieved

    from http://www.doe.mass.edu/lawsregs/603cmr35.html?section=07

Mayer, M., & Phillips, V. L. (Eds.). (2012). *Primary sources 2012: America's teachers*

    on the teaching profession. Retrieved from

    http://www.scholastic.com/primarysources/pdfs/Gates2012_full.pdf

McKeachie, W., & McKeachie, W. (1957, Winter 1957). Student Ratings of Faculty.

    Improving College and University Teaching, *5*, 4-8. Retrieved from

    http://www.jstor.org/stable/27561792

McMillan, J. H., & Schumacher, S. (2010). *Research in education: Evidence-based*

    inquiry (7 ed.). Upper Saddle River, NJ: Pearson Education.

Membership trends and projections. (2013). Retrieved from

    http://annualreport.acsa.org/2012/membership/membership-trends/

Mertler, C. A. (1999, Spring 1999). Teacher perception of studetns as stakeholders in

    teacher evaluation. *American Secondary Education*, *27*(3), 17-30. Retrieved

    from http://www.jstor.org/stable/41064317

MET Project. (2012). Gathering feedback for teaching: Combining high-quality

    observations with student surveys and achievement gains. Retrieved from

    http://metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Milanowski, A. (2004). The relationship between teacher performance evaluation scores

    and student achievement evidence from Cincinatti. *Peabody Journal of*

    Education, *79*(4), 33-53. Retrieved from http://www.jstor.org/stable/1493307

New Teacher Project. (2013). Teacher evaluation 2.0. Retrieved from

      http://tntp.org/assets/documents/Teacher-Evaluation-Oct10F.pdf

Newton, E. S. (1977, Feb.). Andragogy: Understanding the adult as learner. *Journal of*

      Reading, *20*, 361-363. Retrieved from http://www.jstor.org/stable/40032981

Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002, Suimmer 2002). A

      cost framework for professional development. *Journal of Educational Finance*,

      28(1), 51-74. Retrieved from http://www.jstor.org/stable/40704157

O'Shea, M. D. (2006). *Student perceptions of teacher support: Effect on student*

      achievement (Doctoral dissertation). Retrieved from

      http://search.proquest.com/docview/305353638?accountid=10051

Pallas, A. M. (2011, December 2010/January 2011). Measuring What Matters. *The Phi*

      Delta Kappan, *92*(4), 68-71. Retrieved from

      http://www.jstor.org/stable/27922491

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods* (3 ed.). Thousand

      Oaks, CA: Sage.

Pedagogy. (2016). In *Oxford English dictionary*. Retrieved from

      http://www.oed.com/view/Entry/139520?redirectedFrom=pedagogy#eid

Pedagogy. (2016). Retrieved from

      http://etymonline.com/index.php?allowed_in_frame=0&search=pedagogy

Phillips, V., & Hughes, R. L. (2012, December 4, 2012). Teacher collaboration: The

      essential common-core ingredient. *Education Week*, *32*(13), 32-35. Retrieved

      from www.edweek.org/ew/articles/2012/12/05/13hughes.h32.html?print=1

Pioneers In our field: Jean Piaget - champion of children's ideas. (2016). Retrieved from

http://www.scholastic.com/teachers/article/pioneers-our-field-jean-piaget-

champion-childrens-ideas

Quaglia, R. J., & Corso, M. J. (2014). *Student voice: The instrument of change*.

Retrieved from

https://books.google.com/books?id=0V5WBAAAQBAJ&printsec=frontcover&d

q=student+voice+quaglia&hl=en&sa=X&ved=0ahUKEwiEloyN6J_MAhUUwW

MKHS1cACEQ6AEIIzAB#v=onepage&q=student%20voice%20quaglia&f=false

Rada, H., & Knowles, M. (1980, Fall). An interview with Malcolm Knowles. *Journal

of Developmental & Remedial Education*, *4*(1), 2-4. Retrieved from

http://www.jstor.org/stable/42774527

Ramsdell, R. (2011). *Enhancing teacher evaluation and feedback systems with Tripod

student surveys* [PowerPoint slides]. Retrieved from Arizona Department of

Education: http://www.azed.gov/highly-qualified-

professionals/files/2012/01/tripod_ramsdell_2011-11-14_arizona_publish.pdf

Rayens, M. K., & Hahn, E. J. (2000). Buliding consensus using the policy Delphi

method. *Policy, Politics, & Nursing Practice*, *1*, 308-315. Retrieved from

http://www.mc.uky.edu/tobaccopolicy/ResearchProduct/BuildingConsensus.pdf

Roberts, C. M. (2010). *The dissertation journey: A practical and comprehensive guide

to planning, writing, and defending your dissertation* (2 ed.). Thousand Oaks,

CA: Corwin.

Rockoff, J. E., & Speroni, C. (2010, May 2010). Subjectve and objective evaluations of

    teacher effectiveness. *American Economic Review*, *100*, 261-266. Retrieved

    from http://www.jstor.org/stable/27805001

Rodin, M., & Rodin, B. (1972, September 29). Student Evaluation of Teachers.

    Science, *177*, 1164-1166. Retrieved from

    http://www.jstor.org.libproxy.chapman.edu/stable/1734252

Rothstein, J. (2011). Review of learning about teaching. Retrieved from

    http://nepc.colorado.edu/thinktank/review-learning-about-teaching

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and

    analysis. *International Journal of Forecasting*, *15*, 353-375. Retrieved from

    http://www.forecastingprinciples.com/files/delphi%20technique%20Rowe%20Wr

    ight.pdf

Royce, C. A. (2010, November). A revolutionary model of professional development.

    Science Scope, *34*(3), 6-9. Retrieved from http://www.jstor.org/stable/43182921

Sawchuk, S. (2010, November 10). Full cost of professional development hidden.

    Education Week, *30*(11), 14-16. Retrieved from

    http://www.edweek.org/ew/articles/2010/11/10/11pd_costs.h30.html?tkn=OUOF

    ymp5AnGfOCZ40kLdKORH6id4Bu3kZhmh&print=1

Scheurich, V., Graham, B., & Drolette, M. (1983). Expected grades versus specific

    evaluations of the teacher as predictors of students' overall evaluation of a

    teacher. *Research in Higher Education*, *19*, 159-173. Retrieved from

    http://www.jstor.org/stable/40195559

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997, October 1997). Faculty

perspectives on course and teacher evaluation. *Research in Higher Education*, *38*,

575-592. Retrieved from http://www.jstor.org/stable/40196249

School Services of California, Inc. (2016). *Local control and accountability plan

development workshop* [Lecture notes]. Retrieved from

https://www.sscal.com/workshops.cfm?action=display_workshop&workshop_ID

=715

Shadreck, M., & Isaac, M. (2012). Science teacher quality and effectiveness: Gweru

Uran Junior Secondary School students' points of view. *Asian Social Science*,

8(8). http://dx.doi.org/DOI: 10.5539/ass.v8n8p160

Shulman, L. S. (1986, February 1986). Those who understand: Knowledge growth in

teaching. *Educational Researcher*, *15*(2), 4-14. Retrieved from

http://www.jstor.org/stable/1175860

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate

research. *Journal of Information Technology Education*, *6*(), 1-21. Retrieved

from http://www.jite.org/documents/Vol6/JITEv6p001-021Skulmoski212.pdf

Sommerville, J. A. (2008). *Effective use of the Delphi process in research:Its

characteristics, strengths and limitations* (Unpublished doctoral dissertation).

Oregon State University, Corvallis, Or. Retrieved from

https://eric.ed.gov/?id=EJ897797

State schools superintendent Tom Torlakson announces healthy kids survey for 2014-05

school year. (2016). Retrieved from

http://www.cde.ca.gov/nr/ne/yr16/yr16rel50.asp

Stecher, B., Garet, M., Holtzman, D., & Hamilton, L. (2012, November). Implementing

measures of teacher effectiveness. *The Phi Delta Kappan*, *94*(3), 39-43.

Retrieved from http://www.jstor.org/stable/41763674

Student feedback forms. (2016). Retrieved from http://www.temple.edu/ira/assessment-

and-evaluation/student-feedback-forms/index.html

Task Force on Educator Excellence. (2012). *Greatness by design*. Retrieved from

http://www.cde.ca.gov/eo/in/documents/greatnessfinal.pdf

The Bill and Melinda Gates Foundation. (2012). Gathering feedback for teaching:

Combining high-quality observations with student surveys and achievement

gains. Policy and practice summary. Retrieved from

http://metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Thiessen, D. (2006, Winter). Student knowledge, engagement, and voice in educational

reform. *Curriculum Inquiry*, *36*, 345-358. Retrieved from

http://www.jstor.org/stable/4124742

Thorne, G. L. (1980, Mar-Apr 1980). Student ratings of instructors: From scores to

administrative decisions. *The Journal of Higher Education*, *51*, 207-214.

Retrieved from http://www.jstor.org/stable/1981375

Torff, B., & Sessions, D. (2009, Summer). Principals' perceptions of the causes of

teacher ineffectiveness in different secondary subjects. *Teacher Education

Quarterly*, *36*(3), 127-148. Retrieved from http://www.jstor.org/stable/23479193

Towe, P. B. (2012). *An investigation of the role of a teacher evaluation system and its

influence on teacher practice and professional growth in four urban high schools*

(Doctoral dissertation, Seton Hall University). Retrieved from

http://search.proquest.com/docview/1033658578?accountid=10051

Walker, T. (2014, November 2, 2014). NEA Survey: Nearly half of teachers consider

leaving profession due to standardized testing. *NEA Today*. Retrieved from

http://neatoday.org/2014/11/02/nea-survey-nearly-half-of-teachers-consider-

leaving-profession-due-to-standardized-testing-2/

Webb, K. M. (1995, December 1995). Not even close: Teacher evaluation and teachers'

personal practical knowledge. *The Journal of Educational Thought*, *29*, 205-226.

Retrieved from http://www.jstor.org/stable/23767725

Webster-Wright, A. (2009, June 2009). Reframing professional development through

understanding authentic professional learning. *Review of Educational Research*,

79, 702-739. Retrieved from http://www.jstor.org/stable/40469054

Weingand, D. E. (1996, Winter). Continuing education: A reminder about andragogy.

Journal of Education for Library and Information Sciences, *37*(1), 79-80.

Retrieved from http://www.jstor.org/stable/40324288

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect.

Retrieved from tntp.org/publications/view/the-widget-effect-failure-to-act-on-

differences-in-teacher-effectiveness

What are the Common Core Standards? (2012). Retrieved from

http://www.cde.ca.gov/re/cc/tl/whatareccss.asp.

White, R. (1976, Spring 1976). Some added support justifying administrative use of

student evaluations of teachers. *Journal of Economic Education*, *7*, 120-124.

Retrieved from http://www.jstor.org/stable/1182777

Williams, P., Sullivan, S., & Kohn, L. (2012). Out of the mouths of babes: What do

secondary students believe about outstanding teachers? *American Secondary*

Educatoin, *40*. Retrieved from http://www.questia.com/library/p62110/american-

secondary-education/i3319326/vol-41-no-2-spring

Worrell, M. N., & Dey, R. (2008). Student voice: Philosopher's stone or Pandora's

box? Retrieved from https://studentshare.net/education-essay/622521-student-

voice-pandora--box-or-philosopher-s-stone/

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shipley, K. L. (2007). *Reviewing*

the evidence on how teacher professional development affects student

achievement. Retrieved from http://ies.ed.gov/ncee/edlabs

Youngs, P. (2013). Using teacher evaluation reform and professional development to

support common core assessments. Retrieved from

http://www.americanprogress.org/issues/education/report/http://www.americanpr

ogress.org/education/report/2013/02/05/51410/using-teacher-evaluation-reform-

and-professional-development-to-support-common-core-assessments/

Yousuf, M. I. (2007, May 2007). Using experts' opinions through the Delphi technique.

Practical Assessment, Research & Evaluation, *12*(4). Retrieved from

http://pareonline.net/getvn.asp?v=12&n=4

Zemke, R., & Zemke, S. (1984, March 9). 30 things we know for sure about adult

learning. *Innovation Abstracts*, *6*(8). Retrieved from

http://eric.ed.gov/?id=ED248920

Zemyov, S. I. (1998). Andragogy: Origins, developments and trends. *International Review of Education*, *44*(1), 103-108. Retrieved from http://www.jstor.org/stable/3445079

Zimmerman, J. A., & Jackson-May, J. (2003, Spring 2003). Providing effective professional development: What's holding us back? . *American Secondary Education*, *31*(2), 37-48. Retrieved from http://www.jstor.org/stable/41064485

APPENDICES

Letter of Invitation to Research Subjects

_____:

I am a doctoral student in the field of Organizational Leadership in the School of Education at Brandman University. I am conducting a study into the use of Student Evaluations of Teachers (SETs) at the secondary level to inform evaluation and professional development practices. In particular, I am seeking assemble an expert group of teacher trainers and administrators to investigate how SETs could be formulated and used to provide more effective and targeted professional development for California high school teachers.

I am asking for your assistance in the study by requesting that you respond to a series of three electronic questionnaires as part of a Delphi study. The questionnaires will be administered in three rounds. Each round will take approximately 15-20 minutes to complete. Rounds will be administered in 7-10 day increments, beginning on Monday, (date to be determined). You will have the opportunity to respond to each round at your convenience within the time period designated for each round.

If you agree to participate in the electronic questionnaire, be assured that it will be completely confidential. Your name will not be attached to your electronic survey response. All information will remain in electronic files accessible only to the researchers. No employer will have access to the electronic questionnaire information. You will be free to withdraw from the study at any time. Further, you may be assured that the researcher is not affiliated with your employing agency.

Please review the attached _Informed Consent_ and _Research Participant's Bill of Rights_. If you agree to participate, please respond to this email indicating that you have read the attachements and agree to participate. (You do not need to print and sign the forms; your email response will suffice as your informed consent.) When I receive your response, I will send the first questionnaire.

I am available by phone at (559) 920-2381 to answer any questions you may have. Your participation would be greatly valued.

Sincerely,

Lawrence Jarocki

Informed Consent Form

CONSENT TO PARTICIPATE IN RESEARCH

The Use of Student Feedback in Teacher Development—A Delphi Study

BRANDMAN UNIVERSITY
16355 LAGUNA CANYON ROAD
IRVINE, CA 92618

Responsible Investigator: Lawrence Jarocki

Purpose of Study: The purpose of this Delphi study is to identify the most important elements for SETs (Student Evaluation of Teachers) at the high school level as perceived by a panel of expert master teachers, administrators, and teacher trainers. In addition, it is the purpose to determine how the results of SETs can best be used by teacher trainers and administrators to inform evaluation and professional development practices for secondary teachers.

Procedures: In participating in this study, I agree to respond to a series of three electronic survey questionnaires administered in 7-10 day increments over a period of no more than 30 days as part of a Delphi Study. Each survey will take approximately 15-20 minutes to complete.

   a. Round one of the electronic questionnaire will require participants to type responses to three open-ended questions.
   b. Round two of the electronic questionnaire will require participants to rate the level of importance of items related to responses to round-one questions on a predetermined Likert scale.
   c. Round three of the electronic questionnaire will require participants to rate the level of importance of items related to responses to round-one questions on a predetermined Likert scale and type responses to open-ended questions related to ratings generated during round two.

I understand that:
   a. There are minimal foreseeable risks involved in this research study. The identity of all participants will be anonymous throughout the duration of the study, though email addresses of participants will be required for electronic survey participation.
   b. The possible benefits of this study to the field of education include contributing to the growing body of research related to the use of SETs to inform evaluation and professional development practices at the secondary

level and potentially informing the development of SETs for public school application.

    c.   Any questions I have concerning my participation in this study will be answered by Lawrence Jarocki, M.A. at (559) 429-9862 or jaro2601@mail.brandman.edu.

    d.   I understand that I may refuse to participate or may withdraw from this study at any time without any negative consequences. Also, the Investigator may stop the study at any time.

    e.   I also understand that no information that identifies me will be released without my separate consent and that identifiable information will be protected to the limits allowed by law. If the study design or the use of the data is to be changed, I will be informed and my consent reobtained. I understand that if I have any questions, comments, or concerns about the study or the informed consent process, I may write or call the Office of the Executive Vice Chancellor of Academic Affairs, Brandman University, at 16355 Laguna Canyon Road, Irvine, CA 92618, (949) 341-7641.

Acknowledgement: I acknowledge that I have received a copy of this form and the "Research Participant's Bill of Rights."

Consent: I have read the above and understand it and hereby consent to the procedure(s) set forth.

_____

Printed Name of Participant

_____   _____

Signature of Participant   **Date**

_____   _____

Signature of Principal Investigator   **Date**

APPENDIX C

Delphi Study Round One Questionnaire

Sent to participants electronically via Google Forms:

https://docs.google.com/forms/d/e/1FAIpQLSduQALsinJHAL_WIeTelYVz0Pyx

E3Puc3XwDyIC0NN6VPk48g/viewform?usp=sf_link

Welcome to this Delphi Study on the use of Student Evaluations of Teachers (SETs) to inform professional development and/or evaluation at the secondary level. We'll be involved in a few rounds of discussion on what we might put in such surveys and how they could be used to improve how we evaluate and engage in professional development with secondary school teachers. For this first round, we'll start with some open-ended questions. As you answer them, feel free to expand on your responses, providing insight and justification for your opinions. We'll also have a few more closed-ended questions to help establish some of the logistics on the implementation and use of SETs at the secondary level. If you have any questions about the questions, feel free to contact me.

After I've received the group's responses, I'll be analyzing them and then forming the second-round survey, where we'll work towards consensus on the content and use of SETs. Most Delphi Studies end after the third round.

Thank you for your participation in this study. I hope you look forward to engaging with colleagues on this potentially rich source of information on how to improve instruction in California high school classrooms.

Question 1: If high school students were being surveyed about their teacher's work in their class, and that information might be used for evaluation or professional development purposes, what should we be asking about the teachers? (Feel free to include as many aspects of instruction as you see fit. Where possible, please provide justification as to why that aspect should be considered. Also, if you'd like to include the actual questions you think should be included, please do so.)

```
```

Question 2: If these surveys were to be used to inform professional development practices (either for individuals or groups), when and how often in the school year should students be surveyed about their teachers?

```
```

Question 3: If these surveys were to be used to inform professional development practices (either for groups or individuals), how should the results be disseminated (i.e., who should see them, and in what forum)?

```
```

Question 4: How should the results of these surveys be used to improve instructional practices, either for groups or individuals?

```
```

Question 5: If these surveys were to be used in the evaluation process, when and how often in the school year should students be surveyed?

```
┌──────────────────────────────────────────────────────────────┐
│                                                              │
│                                                              │
└──────────────────────────────────────────────────────────────┘
```

Question 6: If these surveys were to be used in the evaluation process, how much weight should they carry in the outcome (i.e., what percentage of a teacher's evaluation score could be based on student survey responses)?

```
┌──────────────────────────────────────────────────────────────┐
│                                                              │
│                                                              │
└──────────────────────────────────────────────────────────────┘
```

Question 7: What advantages do you see in the use of student surveys to inform evaluation or professional development practices?

```
┌──────────────────────────────────────────────────────────────┐
│                                                              │
│                                                              │
└──────────────────────────────────────────────────────────────┘
```

Question 8: What disadvantages do you see in the use of student surveys to inform evaluation or professional development practices?

```
┌──────────────────────────────────────────────────────────────┐
│                                                              │
│                                                              │
└──────────────────────────────────────────────────────────────┘
```

APPENDIX D

Delphi Study Round Two, Part One Questionnaire

Sent to participants electronically via Google Forms:

https://docs.google.com/forms/d/e/1FAIpQLScfrAL6_uUBXv5i86ALZESLXAV

Air2Mg4Jc-2WAGRvGfp23Nw/viewform?usp=sf_link

Welcome back to the study. This is the first part of the second of three surveys. In this one, we'll be looking at the answers you gave in the first round and attempting to come to consensus on some of the issues. If all goes well, we'll have one more round, and then I'll write up the results and send them out. Thank you for taking part in this.

Question One: What should be included in the survey? Below you'll see the range of answers generated in the first round of the study. For each, please rate using the scale below how important you feel each item would be to include in a Student Evaluation of Teachers (SET) at the secondary level. Participants have come up with quite a few possibilities, and we might want to whittle down the list to the most important aspects and attributes of teaching. Also, if you feel strongly about any of the items, please note the question number and make a comment in the optional field at the end. These will be included in the next round of the survey as we look at the rankings. (Note: The wording will probably change on some of these; eliciting information about the given area is the important consideration.) *Descriptors added for the Likert scale.*

1	2	3	4	5	6
○	○	○	○	○	○
Not Important	Slightly Important	Somewhat Important	Quite Important	Very Important	Extremely Important

1. Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work?

2. Does your teacher often have you work with a partner or group during a lesson?

3. Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means?

4. Does your teacher require you to write to explain or justify ideas?

5. Does your teacher ask you to show that you understand during the lesson?

6. Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson?

7. Does your teacher have clear objectives for each day, posted visibly?

8. Does your teacher have high standards for your work?

9. Does your teacher give individual help when necessary?

10. Does your teacher give you effective feedback on your work in a timely manner?

11. Is your teacher excited about his/her subject matter?

12. Does your teacher come prepared to class each day?

13. Does your teacher know the subject he/she is teaching well?

14. Does your teacher make the material engaging?

15. Does your teacher have a 'can do' attitude towards students' ability and work?

16. Can your teacher think on his/her feet to keep a class moving?

17. Does the content of the course prepare you for the exams?

18. Is your teacher available outside of class for extra help?

19. How much of class is usually spent in lecture vs. in interactive work?

20. What parts of the class were difficult? Why?

21. Does your teacher have a good rapport with the students?

22. Does your teacher give good instructions?

23. Can your teacher convey concepts in multiple ways?

24. How much do you feel you've learned in class this year?

25. What connections have you made in class this year? (Written response here)

26. Do you feel welcomed and supported by your teacher?

27. What is one of the ways your teacher teaches the lesson that is effective or 'works' for you?

28. Do you feel like you accomplish something in class each day?

29. Does your teacher make good use of class time?

30. Do you have a sense of belonging in this class?

31. How did you feel about the subject of this class before you took it? And now?

32. What makes a good teacher? (Would require a written response)

33. How flexible is your teacher?

34. Does your teacher change the way he/she teaches based on individual student needs?

35. Do the course materials feel useful and relevant to real life?

36. Does your teacher link course content to other subjects/disciplines?

37. Do you feel safe asking questions, commenting, or asking for help in class?

38. Does your teacher move from activity to activity well?

39. Does the teacher use technology in the class? Do students?

40. Is your teacher fair and equitable?

41. Does your grade in class reflect your learning, or does it reflect other aspects (e.g. homework completion)?

42. Do you feel challenged in this class?

43. Does the homework for this class reinforce the learning done during lessons?

44. Does your teacher care about the students in this class?

45. Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates?

46. Does your teacher know your individual strengths and weaknesses?

47. Do you know how your teacher wants routine classroom actions handled? (e.g. raise hands to answer questions, how to request to use the restroom, what to do if you are absent)

48. Does the teacher ensure that you know what criteria you will be measured against?

49. When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Would require written response)

(Note: Each of these were followed with the Likert-scale response form seen following question one.)

Question Two: When you are working on independent or small group work, how does the teacher monitor your understanding and progress? (Would require written response)

Question Three: Do you have any comments that you'd like to make regarding any of the survey items? Please include the item number(s) in your entry.

Delphi Study Round Two, Part Two Questionnaire

Sent to participants electronically via Google Forms:

https://docs.google.com/forms/d/e/1FAIpQLSfW7y93_8v06s22qQfkWTQj8IVJC

x5y5tWUN3coAD0sg81Neg/viewform?usp=sf_link

Welcome back to round two. In this part of the second round, we'll be looking at

the implementation and use of surveys. I've pulled these responses from your input in the

first round of the surveys. I've also attached a document containing quotes from you on

each of the questions below. Feel free to explore that as you go through the choices

below. Once the results from both parts of round two's surveys are compiled, I'll be

sending out a final survey for round three. Thank you for your insightful participation in

my study.

Question One: If used for professional development purposes, when should the

surveys be given?

1. Twice a year, at the end of each semester (so adjustments can be made for the
   second semester and the results can then be viewed at the end of the year)

2. Quarterly (so that adjustments can be made quicker and more often)

3. Near the end of the school year (so that results can inform summer
   professional development efforts)

4. At 'benchmark' points, such as after the first month of school, around
   Thanksgiving, February, and again in April

5. Let the teacher decide

Question Two: If used for evaluation purposes, when should the surveys be given?

1. Twice a year, at the end of each semester

2. Twice a year, coming mid-fall and prior to the springtime evaluation process

3. Near the end of the school year (so that results can inform summer professional development efforts)

4. At 'benchmark' points, such as after the first month of school, around Thanksgiving, February, and again in April

5. Student surveys should not be used for evaluation purposes

6. Let the teacher decide

```



```

Question Three: If used for evaluation purposes, how much weight should they carry in a teacher's final evaluation?

1. No weight at all, but it could be a box in the teacher's evaluation

2. 5-10%

3. 20%

4. 30%

5. 50%

```



```

Question Four: If these surveys were to be used to inform professional development practices (either for groups or individuals), how should the results be disseminated (i.e., who should see them, and in what forum)? (Please mark all that apply)

1. Administrators

2. Individual Teachers see their own

3. Department Heads, with individual teacher scores

4. Department Heads, without individual teacher scores

5. PLCs (without individual names)

6. All staff (without individual names)

Question Five: How should the results of these surveys be used to improve instructional practices, either for groups or individuals? (Please mark all that apply)

1. Administrators and grade levels/bands view the data collaboratively to discuss implications and areas of strength/growth.

2. The results should be used primarily as a needs assessment for the larger PD efforts of a school/district. They should be part of a larger PD plan.

3. Administrators should use the data when planning whole-school PD efforts.

4. The results should be shared by administrators with individual teachers as part of the evaluation/counseling process.

5. Use the results to differentiate PD initiatives for the needs of the teachers.

6. PD could be conducted by teachers scoring high in particular areas, with possible classroom demonstrations of best practices for visiting teachers

> [ ]

Question Six: What are the main advantages that you see in the use of student surveys at the secondary level? (Please mark all that apply)

1. Students spend the most time with teachers, so their insights about their practice can be the most informed.

2. There is accountability and perspective to the population actually being served by the teacher.

3. Professional development practices can be improved if teaching is examined as a two way street: the instructor's knowledge meets the learner's needs.

4. Survey data tell an administrator if parent or student complaints are warranted and provides evidence for suggested teaching improvements.

5. Students are shown that their voices count.

6. Students are shown that their voices count.

7. Surveys provide a perspective that cannot be seen from observations and walk-throughs.

8. Professional development choices will be based on student needs, not on the strengths of the teachers or the current trends at the district level.

> [ ]

Question Seven: What are the main disadvantages that you see in the use of student surveys at the secondary level? (Please mark all that apply)

1. Needs vary by class, so what works in one class may not be needed in another.

2. Students can be nasty, and no one likes reading bad things about themself.

3. Surveys can become a popularity contest, not a read reflection of teaching.

4. Students can give higher marks to teachers who give easier grades.

5. Surveys can be subjective, and the results can vary from day to day.

6. There is potential for abuse from those in power.

7. Students can give higher marks in those classes they chose (electives, areas of interest) and lower marks in classes they're forced to take.

8. It is nearly impossible to craft a multiple-choice survey that really encapsulates teacher performance.

APPENDIX F

Delphi Study Round Three Questionnaire

Sent to participants electronically via Google Forms:

https://docs.google.com/forms/d/e/1FAIpQLSe27CTAKv12u6htLhz15YMcx2Qcr

F0ByHzHgLsAPIFMFnZTNw/viewform?usp=sf_link

Thanks for sticking with this this far. It's time for the final round of questions. Over the previous three surveys, we've been moving toward consensus on the whats, hows, whens, and whys of using student surveys for evaluation and professional development at the secondary level. This final survey will attempt to come to some tentative answers, which I will then write up as my doctoral dissertation. While the results aren't meant to be definitive, I hope that they will be useful to educational systems and personnel wanting to investigate the use of student surveys to inform their practices. Again, thank you for your patience, support, and expertise in this process.

Question One: From the first two surveys, I've collected opinions about what should be included in the surveys. Each of the forty-nine suggestions have been ranked using a 1-6 Likert scale. What follows is the list in descending order. Given that the length of such a survey is still up for consideration, where a question falls in this order can influence whether it ends up on a final document. If you feel that something is more or less important than other items near it, please mark the question accordingly. If you feel that its placement is roughly correct, you don't need to answer for that item.

1. Does your teacher give you effective feedback on your work in a timely manner? (Group score: 5.380952381/6 on a 1-6 Likert scale)

   ○ Raise in importance

   ○ Lower in importance

   ○ Keep in current position

   ○ Add option or **ADD "OTHER"**

2. Does your teacher come prepared to class each day? (5.380952381/6)

3. Does your teacher clarify things that are confusing or provide additional support before moving on in the lesson? (5.333333333/6)

4. Do you feel welcomed and supported by your teacher? (5.333333333/6)

5. Do you feel safe asking questions, commenting, or asking for help in class? (5.333333333/6)

6. Does your teacher ask you to show that you understand during a lesson? (5.142857143/6)

7. Does your teacher has a good rapport with the students? (5.095238095/6)

8. Does your teacher know the subject he/she is teaching well? (5.047619048/6)

9. Does your teacher require you to write to justify or explain ideas? (5/6)

10. Does your teacher care about the students in this class? (5/6)

11. Does your teacher give you concrete examples or demonstrations of the skills you need to apply before you are asked to do independent work? (4.952380952/6)

12. Does your teacher have high standards for your work? (4.857142857/6)

13. Does your teacher give individual help when necessary? (4.857142857/6)

14. Does the content of the course prepare you for the exams? (4.857142857/6)

15. Is your teacher excited about his/her subject matter? (4.80952381/6)

16. Does your teacher give good instructions? (4.80952381/6)

17. Do you feel challenged in this class? (4.80952381/6)

18. Does the teacher ensure that you know what criteria you will be measured against? (4.80952381/6)

19. Does your teacher make good use of class time? (4.761904762/6)

20. Does your teacher Does your teacher use technology in the class? Do students? (4.666666667/6)

21. Does the teacher make the material engaging? (4.666666667/6)

22. Do you have a sense of belonging in this class? (4.619047619/6)

23. Is your teacher fair and equitable? (4.619047619/6)

24. Does your teacher engage you in the ideas or content you are learning about with visuals, media, art, music or other means? (4.523809524/6)

25. Does your teacher have a 'can do' attitude towards students' ability and work? (4.523809524/6)

26. What is one of the ways your teacher teaches the lesson that is effective or 'works' for you? (Short answer from students) (4.523809524/6)

27. Does your teacher often have you work with a partner or group during a lesson? (4.476190476/6)

28. Are students in this class asked to listen to, comment on, and question the contribution of their teammates and classmates? (4.476190476/6)

29. Do you feel like you accomplish something in class each day?

(4.380952381/6)

30. What parts of the the class were difficult? Why? (Short answer)

(4.333333333/6)

31. Can your teacher convey concepts in multiple ways? (4.333333333/6)

32. How much do you feel you've learned in class this year? (4.333333333/6)

33. Does the homework for this class reinforce the learning done during lessons?

? (4.333333333/6)

34. Do you know how your teacher wants routine classroom actions handled?

(4.333333333/6)

35. Does your teacher have clear objectives for each day, posted visibly?

(4.285714286/6)

36. Do the course materials feel useful and relevant to real life? (4.285714286/6)

37. How much of class is usually spent in lecture vs. in interactive work?

(4.238095238/6)

38. Is your teacher available outside of class for extra help? (4.142857143/6)

39. Does your grade in class reflect your learning, or does it reflect other aspects?

(4.142857143/6)

40. Does your teacher know your individual strengths and weaknesses? (4/6)

41. Does your teacher move from activity to activity well? (3.952380952/6)

42. When you are working on independent or small group work, how does the

teacher monitor your understanding and progress? (Short answer)

(3.952380952/6)

43. Does your teacher link course content to other subjects/disciplines?

(3.857142857/6)

44. Can your teacher think on his/her feet to keep a class moving?

(3.714285714/6)

45. Does your teacher change the way he/she teaches based on individual student needs? (3.714285714/6)

46. What makes a good teacher? (Short answer) (3.666666667/6)

47. What connections have you made in class this year? (Short answer)

(3.380952381/6)

48. How did you feel about the subject of this class before you took it? And now? (Short answer) (3.142857143/6)

49. How flexible is your teacher? (2.761904762/6)

(Note: Each of these were followed with the response form seen following question one.)

For the next few questions, we're going to revisit the timing and use of these surveys. Below you'll find the original statements and the percentage of respondents choosing each. Please mark them again as you see fit.

Question Two: Given that we have about 50 possible survey items here, we need to consider the length the student survey should be. Thinking about both manageability and thoroughness, how many items do you feel should be on this survey?

1. 20-30 questions (70% of respondents)

2. 10-19 questions (15% of respondents)

3. 5-9 questions (15% of respondents)

Question Three: If used for professional development purposes, when should the surveys be given? (Mark one)

1. Twice a year, at the end of each semester, so adjustments can be made for the second semester and the results can then be viewed at the end of the year. (33%)

2. At 'benchmark' points, such as after the first month of school, around Thanksgiving, February, and again in April. (29%)

3. Quarterly, so that adjustments can be made quicker and more often. (21%)

4. Let the teacher decide. (12%)

5. Near the end of the school year, so that results can inform summer professional development efforts. (5%)

Question Four: If used for evaluation purposes, when should the surveys be given? (Mark one)

1. Student surveys should not be used for evaluation purposes. (29%)

2. Twice a year, at the end of each semester. (25%)

3. Twice a year, coming mid-fall and prior to the springtime evaluation process. (21%)

4. Let the teacher decide. (17%)

5. Near the end of school, so that results can inform the summer professional development efforts. (8%)

Question Five: If used for evaluation purposes, how much weight should SETs carry in a teacher's final evaluation?

1. No weight at all, but it could be a box in a teacher's evaluation (62%)

2. 5-10% (21%)

3. 20% (12%)

4. 30% (5%)

Question Six: If these surveys were to be used to inform professional development practices (either for groups or individuals), how should the results be disseminated (i.e., who should see them, and in what forum)? (Please mark all that apply)

1. Individual teachers see their own. (88%)

2. Administrators. (75%)

3. PLCs, without individual names. (54%)

4. All staff, without individual names. (50%)

5. Department heads, without individual teacher scores. (33%)

6. Department heads, with individual teacher scores. (21%)

Question Seven: How should the results of these surveys be used to improve instructional practices, either for groups or individuals? (Please mark all that apply)

1. Use the results to differentiate PD initiatives for the needs of the teachers. (80%)

2. Administrators should use the data when planning whole-school PD efforts. (63%)

3. The results should be shared by administrators with individual teachers as part of the evaluation/counseling process. (58%)

4. Administrators and grade levels/bands view the data collaboratively to discuss

implications and areas of strength/growth. (58%)

5. The results should be used primarily as a needs assessment for the larger PD

efforts of a school/district. They should be part of a larger PD plan. (42%)

6. PD could be conducted by teachers scoring high in particular areas, with

possible classroom demonstrations of best practices for visiting teachers.

(42%)