

2014

Faculty Perceptions as a Foundation for Evaluating Use of Student Evaluations of Teaching

Scott Hamilton Baker
University of Vermont

Follow this and additional works at: <https://scholarworks.uvm.edu/graddis>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Higher Education Commons](#)

Recommended Citation

Baker, Scott Hamilton, "Faculty Perceptions as a Foundation for Evaluating Use of Student Evaluations of Teaching" (2014). *Graduate College Dissertations and Theses*. 288.
<https://scholarworks.uvm.edu/graddis/288>

This Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.

FACULTY PERCEPTIONS AS A FOUNDATION FOR EVALUATING USE OF
STUDENT EVALUATIONS OF TEACHING

A Dissertation Presented

by

Scott H. Baker

to

Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Education
Specializing in Education Leadership and Policy Studies

October, 2014

© Copyright by Scott Hamilton Baker
October 2014

ABSTRACT

Amidst ever-growing demands for accountability and increased graduation rates to help justify the rising costs of higher education, few topics in undergraduate education elicit a broader range of responses than student evaluations of teaching (SETs). Despite debates over their efficacy, SETs are increasingly used as formative (pedagogical practices) and summative (employee reviews) assessments of faculty teaching. Proponents contend SETs are a necessary component in measuring the quality of education a student receives, arguing that they further enable educators to reflect upon their own pedagogy and thus informing best practices, and that they are a valid component in summative evaluations of faculty. Skeptics argue that SETs are ineffective as the measurements themselves are invalid and unreliable, students are not qualified evaluators of teaching, and faculty may lower educational standards due to pressure for higher ratings in summative evaluations. This study dives more deeply into this debate by exploring faculty perceptions of SETs.

Through the use of surveys of 27 full- and part-time faculty within one division at a private, four-year teaching-focused college, this study explored faculty perceptions of SETs primarily as an initial step in a larger process seeking to evaluate perceived and potential efficacy of SETs. Both quantitative and qualitative data were collected and analyzed using Patton's (2008) Utilization-Focused Evaluation (UFE) framework for engaging evidence based upon a four-stage process in which evaluation findings are analyzed, interpreted, judged, and recommendations for action are generated, with all steps involving intended users. Overall, the study data suggests that faculty were generally very supportive of SETs for formative assessments, and strongly reported their importance and use for evaluating their own pedagogy. Findings also indicated faculty relied primarily upon the students' written qualitative comments over the quantitative reports generated by externally determined scaled-questions on the SETs. Faculty also reported the importance of SETs as part of their own summative evaluations, yet expressed concern about overreliance upon them and again indicated a desire for a more meaningful process.

The utility of the UFE framework for SETs, has implications beyond the institution studied, nearly every higher education institution is faced with increasing demands for accountability of student learning from multiple stakeholders. Additionally, many institutions are grappling with policies on SETs in summative and formative evaluation and to what extent faculty and administrators do—and perhaps should—utilize SETs in measuring teaching effectiveness is a pertinent question for any institution of higher education to examine. Thus, the study suggests that to what extent faculty reflect upon SETs, and to what extent they utilize feedback, is a salient issue at any institution; and Patton's model has the potential to maximize the utility of SETs for many relevant stakeholders, especially faculty.

DEDICATION

To Rebecca, my rock throughout the tides and companion on every step of this journey, for your unwavering love, support, and sacrifice, and to Noah and Elsa, for your inspiring sense of inquiry about the world around you, and the love and joy you bring me.

ACKNOWLEDGEMENTS

This entire undertaking would not have been possible without the support and encouragement of many. First, I would like to thank my dissertation committee for all of their valuable input, time, and support; Dr. Judith Aiken, Dr. Barri Tinkler, Dr. Jessica Strolin, and especially my Advisor, Dr. Cindy Gerstl-Pepin, who has been incredibly supportive throughout my entire doctoral program and all aspects of the dissertation process. I would also like to thank all of my colleagues, especially the faculty and staff within my institutional division. It is their dedication to students and teaching that inspired this study and continues to inspire my work. Special thanks are due to my colleagues within the administrative team in my division for their continual support and encouragement. Thanks are also due to my University of Vermont doctoral cohort colleagues and classmates; the discussions and experiences enriched my doctoral experience. I'd also like to thank all of my family and friends who've helped me along the way, through both their continual support and forgiveness for my spiritual and physical absences during particularly demanding times throughout my doctoral studies, and especially my parents, Buzz and Terry Baker, for their unyielding encouragement over the years. Lastly, I would like to thank my wife, Rebecca Baker, who has been with me every step of the way and without whom I could not have completed this journey, and to my wonderful children, Noah and Elsa, whose smiles and unique ways of seeing the world continually motivate me.

TABLE OF CONTENTS

| | |
|---|----|
| I. INTRODUCTION | 1 |
| BACKGROUND..... | 1 |
| CONTEXT FOR STUDY..... | 4 |
| OVERVIEW OF RESEARCH QUESTIONS AND SIGNIFICANCE OF STUDY..... | 6 |
| II. LITERATURE REVIEW | 9 |
| MEASURING STUDENT LEARNING IN HIGHER EDUCATION..... | 10 |
| SETS IN FORMATIVE AND SUMMATIVE EVALUATIONS..... | 13 |
| III. METHODOLOGY | 23 |
| UTILIZATION-FOCUSED EVALUATION..... | 23 |
| DATA COLLECTION METHODS..... | 30 |
| PARTICIPATORY ACTION AND ITERATIVE DATA ANALYSIS..... | 36 |
| RESEARCHER'S ROLE..... | 40 |
| LIMITATIONS AND DELIMITATIONS OF STUDY..... | 43 |
| IV. JOURNAL ARTICLE | 46 |
| INTRODUCTION..... | 46 |
| THE HISTORICAL FOUNDATION OF THE ASSESSMENT MOVEMENT..... | 50 |
| CONCEPTUAL FRAMEWORK AND METHODOLOGY..... | 53 |
| FACULTY PERCEPTIONS OF SETS..... | 60 |
| DISCUSSION..... | 75 |
| CONCLUSION..... | 80 |
| ARTICLE REFERENCES..... | 81 |
| REFERENCES | 85 |
| APPENDICES | 93 |
| SURVEY INSTRUMENT..... | 93 |

CHAPTER ONE

INTRODUCTION

BACKGROUND

Perhaps few topics in undergraduate higher education elicit a broader range of responses than measurements of student learning amidst an ever-growing demand for accountability, compounded by concerns of the rising cost of higher education and mounting student debt. Pressure to measure various aspects of student learning through “competency” or “outcomes” assessments has increased considerably, due in part to increasing examinations of how much college students are actually learning. In their landmark study of over 2,000 undergraduates from nearly two dozen institutions (large and small, public and private), Arum & Roska (2010) contend that more than a third of graduates “demonstrated no significant gains in critical thinking, analytical reasoning, and written communications” based off their assessment of data from the widely used “Collegiate Learning Assessment” (CLA) exam¹. Criticisms of their methodology and overreliance on the CLA notwithstanding, their study garnered much more attention to the topic of student learning in higher education.

Accountability for student learning has also been the subject of increased attention from scholars and practitioners both within and outside higher education. The American Council on Education, representing college and universities presidents from over 1,800 member institutions, developed a “National Task Force on Institutional Accreditation” which urged a strengthening and improvement in the quality and public

¹ The CLA is a standardized test developed by the Council for Aid to Education, which they claim measures critical thinking, problem solving, and communication.

accountability of the institutional accreditation process through a report entitled “Assuring Academic Quality in the 21st Century: Self-Regulation in a New Era” (National Task Force on Institutional Accreditation, 2012). Kuh, Kinzie, et. al. (2006) report that several states have begun demanding measurements of student success through required data sharing of persistence and graduation rates. A committee from the State Higher Education Executive Officers (2005), comprised of state governing boards and policy agencies officials, issued a declaration that accountability in higher education is a national imperative. A joint commission of policy makers, higher education administrators and faculty, and business and community leaders from several states formed the “National Commission on Accountability in Higher Education” (2005) and argued that more accountability in higher education is vital to continuous improvement not only of the students’ learning, but also because the knowledge and skills college graduates do—or do not—posses upon entering society and the economy is a critical issue of national interest. As Doubleday (20013) points out, the U.S. federal government is also demanding more accountability in higher education, including proposed legislation tying federal financial aid more directly to measures of student performance, as well as “The White House College Scorecard.” Developed by the U.S. Department of Education’s College Affordability and Transparency Center (2014), the Scorecard allows prospective students and families to readily compare institutions across several measures, primarily tuition costs and graduation rates.

Pressure for accountability will only increase with federal governmental pressure to increase graduation rates. The National Center for Education Statistics (2013) projects post-secondary enrollment projected will increase to 24 million students by 2021.

Supported by President Obama, the U.S. Department of Education issued a report entitled “Meeting the Nation's 2020 Goal: State Targets for Increasing the Number and Percentage of College Graduates with Degrees”, which issued a national charge to increase the national proportion of college graduates by 50% (from 40% to 60%) by 2020. Yet along with a desire to increase graduation rates is an even larger increase in the average cost of obtaining a college degree, which has increased exponentially over the last thirty years. The National Center for Public Policy and Higher Education (2008) reported between 1982-2007, the average cost of college tuition increased 439%; nearly 200% more than the rise in average health care costs, and outpacing average family incomes by nearly 300%. With increased concerns for accountability of learning accompanied by a national increase in enrollment and perpetually rising costs, demands for measurement of student learning in higher education are likely to intensify.

One increasingly common method of attempting to measure undergraduate student learning is student evaluations of teaching (SETs), used both as formative (pedagogical practices) and summative (employee reviews) assessments of faculty teaching. SETs have been in use for quite sometime and are common throughout academe (Seldin, 1993; Astin & Antonio, 2012). While institutions of higher learning often utilize other evaluations such as administrator evaluations, peer evaluations, and self-evaluations to measure teaching effectiveness (Smith, 2007), student evaluations have perhaps become the most commonly used (Richardson, 2005; Berk, 2013). Proponents of SETs have long contended that they are a necessary component in measuring the quality of education a student receives (Marsh & Roche, 1993), that they further enable educators to reflect upon their own pedagogy (Centra, 1993), and that they

are a valid component in summative evaluations of faculty (Cashin, 1988). SETs critics such as Schmelkin, Spencer, and Gellman (1997), and Simpson and Sigauw (2000), have long argued that SETs are ineffective measures of faculty teaching rife with issues. These criticisms include that the measurements themselves are invalid and unreliable, students are not qualified evaluators of teaching, students may use them as tools for revenge or are little more than a popularity contests, and may even lower educational standards if faculty feel pressure for higher ratings in summative evaluations by their institutions. Thus, at the crux of many debates concerning the efficacy of SETs is to what extent, if at all, they should be used in formative or summative faculty evaluations (Rifkin, 1995), or both (Donahue, 2000). Common assumptions and conclusions primarily indicate faculty underutilization of SETs due to suspicion and mistrust of their efficacy, as well as institutions' overreliance upon them for formative and summative evaluations; "one gets the impression from the anecdotal literature that there is widespread resistance on the part of faculty to student ratings" (Schmelkin, Spencer, & Gellman, 1997, p. 576). To explore this issue, this study sought to examine faculty perceptions of student evaluations.

CONTEXT FOR THE STUDY

Personally, I was interested in faculty perceptions of student responses to evaluations of teaching effectiveness for several reasons. As an educator, SETs play a role in my own pedagogical practices as well as my own summative evaluations. As a faculty-administrator of over 30 full- and part-time faculty and over 500 undergraduate students at a small, private college in the northeastern United States, I sensed that while some faculty were supportive of measuring student-perceptions of their own learning, there was a general malaise among the majority towards the validity and reliability of

such evaluations, as well as a general concern of how (and indeed, even why) the administration utilized them in determining faculty's summative evaluations. As my institution currently utilizes such evaluations for both formative and summative evaluations of faculty, and my responsibilities include evaluation of faculty, this is also a very salient policy issue for me. First, I sensed a bilateral "overgeneralization" of assumptions and conclusions related to faculty teaching evaluations both among, and between, faculty and administration in higher education. My suspicion was that there is a general dissatisfaction among faculty towards such evaluations, and a general mistrust of how administration utilizes them in determining faculty's overall evaluations. Yet this sense, which is hypocritical to my statements above, is also understood by me to be insufficient to draw any realistic conclusions within my own organization of higher education, let alone academe on a national or global level; hence my desire to conduct research at my own institution to obtain a greater empirical sense of this topic. Secondly, I was curious as to what extent faculty reflect upon, and utilize, SETs to inform their own pedagogy. Lastly, I was interested in faculty perceptions of SETs when used for summative evaluations of teaching effectiveness.

At a broader level, I believe such considerations are relevant far beyond one single institution of higher learning for several reasons. First, every institution is faced with increasing demands for accountability of student learning from government agencies, accreditation bodies, market and peer institution pressures, and students themselves (and their families). Secondly, many institutions are likely grappling with their own (or its accrediting bodies') policies on SETs in summative and formative evaluation. Third, the extent to which faculty reflect upon student responses to teaching

effectiveness, and how they utilize this feedback, is a salient issue at any institution. Lastly, to what extent faculty and administrators do—and perhaps should—utilize SETs in measuring teaching effectiveness is a pertinent question for any institution of higher education. Demands for accountability of student learning in higher education are only going to intensify, and SETs appear to be mainstay component of formative and summative evaluations of teaching effectiveness and student learning.

With the increasing use and reliance on SETs for increased accountability, SETs and how they are utilized (or underutilized) will become even more of a salient issue. Moreover, SETs have been, and will likely continue to be, a contentious and high-stakes issue for faculty and administrators not only as they relate to accountability of student learning and teaching effectiveness, but also because results and interpretations are quite often tied to promotion, tenure, salary and merit-pay decisions (McCarthy, 2012). Therefore, a better understanding of undergraduate faculty perceptions and self-reported use of SETs may add clarity to this important issue and perhaps aid in the enhancement or development of more inclusive, stakeholder-focused formative and summative evaluation models of teaching and learning within undergraduate higher education.

OVERVIEW OF RESEARCH QUESTIONS AND SIGNIFICANCE OF STUDY

The purpose of this study was to better understand how faculty perceive the value of student evaluations of their teaching and the impact of such data on their pedagogy. Through the analysis of survey response data from 27 full- and part-time faculty at a private, four-year teaching-focused college, this study explored answers to the following research questions: *What are faculty perceptions of the efficacy of student evaluations of teaching (SETs) in measuring and informing teaching effectiveness and practices? To*

what extent do faculty utilize student evaluations formatively in their pedagogy? What are faculty perceptions of how SETs are utilized by their institutions? Both quantitative and qualitative data were collected and analyzed within the context of a literature review and through the lens of a establishing a foundation for a formative evaluation of SET use based on Patton's (2008) framework for Utilization-focused Evaluation (UFE).

This study is relevant beyond one single institution of higher learning for several reasons. Institutions of higher education are faced with increasing demands for accountability of student learning from government agencies, accreditation bodies, market and peer institution pressures, and students themselves (and their families). Secondly, many institutions are likely grappling with their own (or its accrediting bodies') policies on SETs in summative and formative evaluation. Thirdly, to what extent faculty reflect upon student responses to teaching effectiveness, and to what extent they utilize this feedback, is a salient issue at any institution. Lastly, to what extent faculty and administrators do—and perhaps should—utilize SETs in measuring teaching effectiveness is a pertinent question for any institution of higher education. Therefore, a better understanding of faculty perceptions, and self-reported use of SETs, will expand the literature and ideally aid in enhancing or developing more inclusive student evaluation of teaching models.

The rest of this dissertation will be organized as follows. Chapter Two, Literature Review provides an overview of relevant literature to set the context for understanding the evolution of SETs as an increasingly common component of accountability of student learning in higher education. A deeper understanding of practices and perceptions thereof from the literature will augment the context and debates surrounding SETs and the

significance of this study. Next, in Chapter Three, Methodology I present a conceptual framework and establish the relevance and applicability of Utilization-Focused Evaluation (UFE) methodology, which is an action-oriented process centered on the principle that those who are being evaluated should have a primary role in the overall evaluation process. The UFE methodology will be used to further illustrate the importance of including faculty perceptions of SETs through an original survey, which inherently increases faculty as stakeholders in the SET process within the research site. Data collection methods, plans for analysis, and use of the data and findings within a “next-steps” framework will also be outlined as are reflections on the limitations of the study, as well as my role as researcher. Lastly, in Chapter Four, Journal Article, this study and its findings have been developed into a journal article for publication.

CHAPTER TWO

LITERATURE REVIEW

The importance of assessing student learning in higher education, while certainly not a new topic of conversation or debate, as illustrated above, has garnered increased attention within the context of escalating demands for accountability of student learning and teaching effectiveness from multiple stakeholders amidst increasing concerns about rising costs. An examination of the literature provides a historical foundation of the assessment movement in higher education and more context for the increasing practice and debates surrounding the role and value of student evaluations of teaching (SETs) in assessing student learning. This literature review is divided into two sections; the first provides historical context for measuring student learning within overall trends of assessment and accountability, and the second examines the debates surrounding increased use of SETs in formative and summative evaluations. As will be highlighted, SETs have become a de facto tool for measurement with strong proponents of its benefits yet also with continued concerns for its shortcomings (Astin, 1977; Cashin, 1988; Marsh & Roche, 1993; Kite, 2012). Although the reliability and validity of these measurements has long been a matter of debate, these deliberations have generally evolved from whether or not SETs should be used to how they should be used more effectively (Marsh, 1982; Aleamoni, 1987; Marsh & Roche, 1993; McKeachie, 1997; Abrami, 2001; Gravestock & Gregor-Greenleaf, 2008; Kite, 2012). A richer appreciation for the different arguments surrounding SET use for both formative (pedagogy) and summative (employee review) reasons provide more context for the importance of this study, as well

as the increased inclusion of faculty perceptions and self-reported use of them, especially within the Utilization-Focused Evaluation framework.

MEASURING STUDENT LEARNING IN HIGHER EDUCATION

Assessment as a measure of student learning in higher education is a concept that, while by no means new, has gained increased attention over the last twenty-five years. As Astin & Antonio (2012) illustrate, Harris' 1970 study argued that course grades were poor indicators of student learning. Ewell (2002) argues that the "intellectual roots of assessment as scholarship" existed well before the concept of assessment became a "recognizable movement" starting in the mid-1980s (p. 3). Martell and Calderon (2005) highlight the increased concern over the "quality of learning at higher education institutions" being largely driven in the early 1980s by discussions focusing on "the failure of higher education institutions as centers of learning" (pp.1-2). Walvoord (2004) also attributes much of the higher education "assessment movement" as being driven by the concerns of extrinsic groups—politicians, employers, learning communities, movements for problem-based learning, etc.—"who were disappointed with the quality of college graduates and the rising costs of higher education"(p. 5).

These "external stakeholders" as Ewell (2002) refers to them, have grown to include accreditation agencies, market forces, and the media, all of which have become more "performance-conscious" and "data-hungry" than ever, making assessment for institutions of higher education "an unavoidable condition of doing business: institutions can no more abandon assessment than they can do without a development office" (p. 22). Assessment will continue to be a critical component of the academy, and likely, with only increased scrutiny and expectations as the costs of higher education continue to rise

exponentially. Despite the national attention on rising health care costs in the United States, which have risen 251% since 1982, the average cost of higher education has risen 439% in the same time frame, compared to the Consumer Price Index rising a little over 100%; (The College Board, 2009, p. 8). Undoubtedly, increased cost to students pursuing higher education degrees is adding to demands for more accountability of student learning.

Martell & Calderon (2005) dichotomize the measurement of learning with assessment as either “indirect,” through, for example, student surveys of their own perceived learning, or “direct,” demonstrations of the skills and knowledge learned. The authors further frame the motivations for—and perhaps expectations of—assessment within two categories; “performance based” versus “value-added,” with the former requiring institutions to demonstrate that students are meeting faculty expectations and the latter demonstrating that learning has improved over the course of their college career. This is an important distinction, because as they argue, value-added assessment “seeks to demonstrate that students learn because of the instruction they receive” (Martell & Calderon, 2005, p. 8).

Ewell (2002) provides an excellent synopsis of trends since the 1960s in assessment in higher education, delineating between four different yet complimentary areas; student learning, retention and student behavior, evaluation and “scientific management,” and mastery learning (pp. 3-8). The concept of assessing “value-added” learning has its roots in series of works by Astin (1977), who utilized longitudinal studies involving over 200,000 students and 300 institutions to examine the overall benefits of higher education for students. Among his many findings, Astin argues that student

satisfaction—and thus, likely retention—is impacted not only by their relationships with faculty and level of contact with them, but also in students’ perceptions of their own learning and development. Drawing upon the work of Astin (1977) and others related to these perceptions, Kuh (1993) summarizes several innate and external factors that influence students’ perceptions of their own learning. In arguing that the benefits of attending college, which were under public scrutiny in the 1970s, were indeed still valid, Pace (1979) drew upon five decades of research on student achievement. This public scrutiny was in part due to increasing attrition rates, which led to an additional foci of higher education assessment practitioners; retention (Ewell, 2002). Increased federal attention, and expenditures, on higher education and student learning led to the development of quantitative and qualitative measures of student learning as an output for cost-benefit studies and investigations of social return on investment (Ewell, 2002).

Lastly, in response to such measures of student learning, a movement within higher education assessment was created to develop means for measuring mastery of knowledge and competencies by students, rather than what protagonists saw as merely testing content knowledge to justify expenditures (Ewell, 2002, p. 6). All of these trends, regardless of their primary motivations, led to what Ewell (2002) called the “birth of the assessment movement in higher education” (p. 7) in 1985, with one seminal report being “Involvement in Learning” which, among other recommendations to enhance student learning, argued that institutions of higher learning could themselves learn from student responses on the institutions’ own performance (Study Group on Conditions of Excellence in American Higher Education, 1984). These student responses are normally

captured through surveys developed internally or by one of several organizations that provide them for institutions.

The instrument currently used for student responses to teaching effectiveness at my institution was developed out of this argument that colleges and universities should listen to students' perceptions of the institutions' effectiveness in helping them learn. Designed by the IDEA Center, a non-profit organization providing assessment and feedback systems for higher education, their primary instrument has evolved from its first employment in 1975 at Kansas State University, to a nationally recognized and widely used instrument entitled "Student Ratings of Instruction." IDEA claims more than 365 institutions use their services through more than 5 million surveys across 250,000 classes each year (IDEA Center). The IDEA Center describes its Student Ratings of Instruction as "soliciting students' feedback on their own learning progress, effort, and motivation, as well as their perceptions of the instructor's use of 20 instructional strategies and teaching methods. In addition, the system surveys instructors regarding their overall goals and highlights for them in the analysis and report" (IDEA Center, 1999). Thus, the student responses to teaching evaluations survey examined in this research project would fall within both Martell & Calderon's (2005) "indirect" and "value-added" categories of assessment. However, the use of SETs as one component of assessment in measuring student learning has also led to much debate in terms of their validity and reliability, but more recently, about their utility as a tool in formative and summative evaluations.

SETs IN FORMATIVE AND SUMMATIVE EVALUATION

The concept of utilizing student evaluations of teaching effectiveness (SETs) is by no means new to education at any level, and the concept of rating learners' perceptions of

instruction utilizing scales was introduced 90 years ago by Freyd (1923) with his research and argument for developing valid and reliable rating scales for teachers. The use of student ratings is well researched in the literature, as Berk (2013) notes “There is more research and experience in higher education with student ratings than with all of the other measures of teaching effectiveness combined” (p. 16). Benton and Cashin (2012) have found more than 2000 articles on the topic, with the vast majority focusing on debates over the reliability and validity of SETs in measuring teaching effectiveness. While some attention has been placed on SETs’ utility for formative and summative evaluations, much of the research and literature on SETs has indeed focused largely on their validity and reliability. This focus has likely been concurrent with increased demands for accountability of student learning—and therefore teaching effectiveness—and the corresponding increase of use by institutions over the last two decades.

As Gravestock and Gregor-Greenleaf (2008) summarize, collective research questioning reliability and validity of SETs has claimed that perhaps more than two dozen variables may influence how students rate a course and an instructor, across four main categories illustrated below:

| <i>Categories</i> | <i>Administrative</i> | <i>Course Characteristics</i> | <i>Instructor</i> | <i>Student</i> |
|---------------------------|---|---|---|--|
| <i>Possible Variables</i> | Timing of evaluations Instructions to students Anonymity Presence of Instructor Purpose | Class size Time of day Elective vs. Required Course Workload/Difficulty Course Level Discipline/Field of study | Age Research productivity Race Personality/Popularity Expressiveness Rank and Experience Gender | Age Gender Year of Study GPA Personality Gender Motivation Attendance Grades |

Additionally, Berk (2013) notes the common contention that students themselves are not fully qualified to rate instructors and this needs to be taken into consideration concerning the use of SETs (p. 16).

Yet as Gravestock & Gregor-Greenleaf (2008) note, the literature generally supports the validity and reliability of SETs and that, in fact, “It should be noted that any effect on overall ratings from any of these particular variables, even when statistically significant, is almost always very small—often changing the ratings by less than one-tenth of 1%,” (p. 44) with validity almost always maintained by limiting scores to no more than one decimal place or as part of a broad category. Greenwald (1997) analyzed much of the research on the validity of SETs over a twenty-year period from 1975-1995, and indicated the majority of findings within this set of literature argued for their validity. Well-constructed and tested SETs appear to consistently measure specific components of teaching practice (Centra, 1993; Marsh, 1984; Aleamoni, 1999). Ory and Ryan (2001) stated increased reliability could be augmented through the common practice of multi-section testing. McKeachie (1997) also reported SETs to be a valid method for indicating teaching effectiveness. Theall & Franklin (2001) argued that attempts to discredit SETs’ validity and reliability was akin to a “witch hunt” and instead, focus should be on properly utilizing their findings. Gravestock & Gregor-Greenleaf (2008) concur with numerous studies that SETs are valid and reliable measures of teaching effectiveness, and moreover, that none of the variables above should by themselves be considered as having any significant biasing effect.

Although contentions against the validity and reliability of SETs as a measurement of teaching effectiveness have been outweighed by those in support of

their validity and reliability, concerns that students' expected grades within a class are positively correlated to their ratings of instructors on SETs and may even lead to grade inflation in pursuit of higher evaluations have been posited for years (Wilson, 1998; Simpson & Siguaw, 2000). However, Aleamoni's study (1999) found that the majority of studies arguing expected grades influenced SET ratings had no significant relationship. Moreover, Heckert, et. al. (2006) conducted a study in which they found that higher evaluations on SETs were given to more demanding instructors in which students' expectations of the difficulty of the course were met. Benton & Cashin (2012) summarize numerous studies over the last 50 years that the authors claim have perpetuated misconceptions about about SETs which are not supported by the research and make utilization of SETs to improve practice more difficult.

Another issue is related to comparisons of faculty performance with their peers. Abrami (2001) recommend institutions need to decide whether or not norm-based or criterion-based referencing should be utilized and to frame the results in the corresponding manner if they are to have more formative benefit for faculty. The utilization of norm-based referencing for faculty can be beneficial to faculty so they have a better understanding of where their own SET results are situated within their own department and how they compare to previously taught sections of the same course (Gravestock & Gregor-Greenleaf, 2008). Cashin (1990) argues that without this comparative context SETs lose some utility for formative evaluations. It is common for institutions of higher education to provide departmental/divisional and/or institutional means on each faculty SETs, normally with graphical information and detailed manuals (or access to them) describing the evaluation tool and guides for interpreting results

(Gravestock & Gregor-Greenleaf, 2008). Department chairs, deans, and institutional researchers charged with some aspect of faculty evaluation are normally given access to this norm-based referencing. What is not clear, however, is the extent to which merely providing faculty with norm-referenced results is the most effective method of enabling faculty to use their SETs for reflective and formative purposes themselves. In an analysis of over 18,000 students' evaluations of teaching in language courses, Zabaleta (2007) determined that SETs should not be used to compare instructors and should only be used for summative purposes very cautiously and only within the context of increasing teaching effectiveness. However, some argue caution if SETs are used in norm-based referencing for summative evaluations by administrators. The primary concern of these arguments is rooted in the level and number of variables that affect any given classes' SETs, such as differences in course outcomes and goals, teaching styles, content, etc., so that comparing one faculty members' scores directly with others invalidates the SETs and should be avoided (McKeachie, 1997; Algozine, 2004; Zabaleta, 2007). Rather, these authors advocate for looking at the distribution of the student ratings for a particular faculty member and reliably categorizing them across only a few levels of overall student perceptions, e.g. where the faculty rated as excellent overall (McKeachie, 1997; Algozine, 2004; Zabaleta, 2007).

Concerns of overreliance on SETs for summative decisions by administrators have been voiced for years (Wilson, 1998), and have been a continuing concern for faculty and researchers (Schmelkin, Spencer, & Gellman, 1997; Zabaleta, 2007; Berk, 2013). Astin & Antonio (2012) also caution against institutions over-use of SETs for both formative and summative assessments, due to compromising the potential for faculty

development in the personal review process by using measurements of teaching performance. They argue the value of SETs as formative assessments and benefits to faculty pedagogy can be detracted by their simultaneous use as summative measurements. They also contend that student evaluations are perhaps not the best method for summative evaluations, and that improvements in measuring teaching efficacy are most likely to be gained by focusing on the educational development of the students. They suggest using other data demonstrating student progress as well as products of students' academic work as more direct evidence of faculty teaching effectiveness ("faculty outcomes"), and state that if an institution uses student assessments of teaching, they should be primarily for the benefit of the instructor for formative reflection.

In their meta-analysis of research on SETs, Penny and Coe (2004) concluded that SETs have considerable utility formatively, especially with proper consultation of feedback including active involvement of instructors in the learning process and adequate time for reflection and dialogue. Kember et al. (2008) also emphasized the importance of instructor reflection on SET feedback, and categorized such self-reflection as ranging from "nonreflection" whereby results are viewed but ignored or dismissed, to "understanding" where results are registered by instructors but not consciously related to their experience in the class, to "reflection" when instructors do relate the results to their experience. The last, and most effective stage of reflection they argue, is "critical reflection" when the instructor not only relate the results to their experiences, but take "transformative" action to incorporate necessary changes or augment successful practices. Changes to pedagogy are often seen as overwhelming or difficult, yet

McGowan and Graham (2009) argue that “meaningful change” does not require complete overhauls of teaching practices, but rather, are achievable changes related to creating more “active” learning, increasing student-teacher interactions, establishing expectations and maintaining high standards, being properly prepared for class and examining grading practices and feedback given to students. Such suggestions indicate that perhaps a more inclusive process of assisting instructors in better appreciating the possible formative merits of SETs might be conducive to higher levels of acceptance, reflection, and utilization in assessing their own teaching practices.

Astin and Antonio (2012) also claim that one SET instrument should not be used for both formative and summative evaluations, but rather, that two sets of ratings should be used, one for faculty consumption for formative purposes and one for institution consumption with students being told which rating is being used for which purpose. The authors also suggest that institutions hire teaching consultants to discuss students’ perceptions of faculty performance directly with students. Lastly, they argue that traditional motivations for assessing faculty are too often driven by purposes of hiring, tenuring, and promoting rather than to enhance the development of the faculty as researchers and teachers. In other words, the potential value of assessing faculty formatively is all too often lost in the institutions utilization of them as summative measurements. Beran, Violato, & Kline (2007) stress the importance of multiple sources of information in evaluating teaching effectiveness, and Benton & Cashin (2012) are very clear in their argument that “writers on faculty evaluation are almost universal in recommending the use of multiple sources of data” (p. 1) and that SETs alone do not

provide “sufficient information to make a valid judgment about an instructor’s overall teaching effectiveness” (pp. 1-2).

While some studies on SETs have focused on their utility as formative and summative evaluations, even fewer have included faculty perceptions of SETs. In one of the few studies actually drawing upon faculty perceptions of SETs and involving a survey of 52 faculty at various institutions in the United States, (Simpson & Sigauw, 2000) found that “Overall, faculty appeared to believe that SETs encourage instructors to lower educational standards, serve as a tool for student revenge, encourage overreliance on ratings in performance evaluations, and are rife with measurement issues” (p. 209). The study also indicated that given these perceptions, faculty responses to SETs, could be categorized as either ignoring them, consciously or unconsciously adopting lower course and grading standards to appease students, or actually implementing teaching practices aimed at positively improving ratings rather than for increases in student learning. This study certainly indicated the perceived—and therefore perhaps realized in practice and in reflection—concerns that faculty have with SETs despite the literature ascertaining their reliability and validity and the increasing institutional practice of utilizing SETs.

In sum, the implications for a deeper understanding of perceptions and practices concerning SETs are profound. Demands for accountability of student learning in higher education are intensifying from multiple stakeholders, from increasing state and federal attention, employer and graduate school expectations, and concerns over the rising costs of higher education for students’ and their families. These demands for measurements of student learning are placing more pressure on institutions and their faculty, who are frequently utilizing SETs more as one such measurement. Additionally, SETs are also

being utilized more for formative and summative evaluations of teaching effectiveness. Yet, as the literature indicates, many debates concerning the utility of SETs—particularly from the faculty perspective—still exist amidst their increased use and weight placed upon their findings. Institutions are using SETs more and more as an answer (or “the” answer) for demands of accountability of student learning and teaching effectiveness, yet seemingly without a deeper understanding of faculty’s self-reported perceptions and practices concerning SETs. This could be leading to increased concerns or resentment towards SETs from faculty, especially when they are factored into summative evaluations of teaching, which might in turn be overshadowing any potential formative utility that may be perceived.

Therefore, the literature suggests it would be beneficial to the overall discussion if more research were conducted on faculty perceptions of SETs, focusing specifically on their self-reported use of them formatively and their opinions of their use for summative measurements. Use of SETs is likely only to increase, so including faculty perspectives coupled with an applied framework such as UFE has the potential to broaden and enrich the discussion surrounding SETs. Additional feedback from faculty concerning their perspective and self-reported use of SETs through UFE could also lead to very beneficial conversations at the institutional level. How faculty perceive SETs’ results, whether or not they report using them to inform their teaching practices, and how they feel about their use in summative evaluations, has the potential to lead to conversations, reflections, and perhaps changes, in use and impact for each institution. These conversations and related research could then collectively enrich the larger discussion and practice

surrounding SETs within the larger contexts of attempting to measure student learning and teaching effectiveness.

CHAPTER THREE

METHODOLOGY

UTILIZATION-FOCUSED EVALUATION

As mentioned, the purpose of this study was to better understand how faculty members perceive the value of student evaluations of their teaching and the impact of such data on their pedagogy as a foundation for initiating an inclusive evaluation of SET utility. Again, my primary research questions are:

- *What are faculty perceptions of the efficacy of student evaluations of teaching (SETs) in measuring and informing teaching effectiveness and practices?*
- *To what extent do faculty utilize student evaluations formatively in their pedagogy?*
- *What are faculty perceptions of how SETs are utilized by their institutions?*

My first research question was primarily exploratory, as I wanted to learn more about faculty perceptions of the efficacy of SETs. The second question, concerning faculty perceptions and formative use of SETs, is as Nardi (2005) describes, “evaluative research” designed and implemented to “evaluate specific outcomes and to provide explanations for why and how a particular result occurred” (p. 11). The third question is perhaps, in Nardi’s (2005) lexicon, more conducive to “explanatory research” because as he describes, I was interested in being “able to explain or perhaps predict” faculty and administrators’ opinions about, and behaviors as a result of, the use of student responses to teaching effectiveness evaluations (p. 10). This research question was designed in a manner conducive for discovery of faculty perceptions of SETs utilization for summative

evaluations while ideally keeping the primary research and methods focused on faculty perceptions of SETs for formative purposes. Given my interest in not only understanding faculty perceptions of SETs within my division, but also in institutional improvement, this study was conducted as a foundation for initiating a formative evaluation of SETs based upon Patton's (2008) framework for "Utilization-Focused Evaluation" ("UFE").

Patton describes UFE as follows:

Evaluation done for and with specific intended primary users, for specific, intended uses. *Utilization-focused evaluation* begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything that is done, from beginning to end, will affect use. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore the focus in utilization-focused evaluation is on intended use by intended users. (p. 37)

For this mixed-methods, action-research oriented study, the UFE methodology was chosen for its inclusiveness of the stakeholders most affected by the evaluation throughout the entire process (p. 78) thus ensuring them a role in the process of decision making and not solely the decision (p. 91), and primarily, for its "commitment to use" emphasizing "knowledge for action—finding out things that can be used." (p. 51). This methodological framework is the most conducive for inquiring about faculty's perceptions and use of SETs, to what degree they do or do not utilize them to inform their own pedagogical practices, and to what extent faculty feel SETs should be used by administrators as a component of measuring their effectiveness in teaching. It is my intent to share the results of my study with the faculty and administrators at my site to begin a larger conversation about SETs and their perceived utility by those most deeply affected by them; the faculty.

Patton (2008) argues that UFE allows participants to “engage” in the overall evaluative processes and its outcomes rather than simply evaluations being “used” (p. 158). He outlines several categories of uses that could be applied within my case to possibly enhance—or develop a new—model of engaging SETs more fully, depending upon faculty feedback. He argues that engaging evaluation means infusing them into the organizational culture through training, and enhancing shared understandings of the evaluation itself in part through building processes that allow for the data to be tied to specific outcomes (e.g. teaching effectiveness, student’s self-report progress on learning through SETs). Patton also contends that engaging evaluations through UFE requires the ability of participants to monitor their own progress, and that they themselves learn more about the evaluation and its impact (e.g. SETs as formative and summative tools). Another directly applicable aspect of UFE is Patton’s call for evaluations to increase engagement through empowering those involved through reflective practice and self-evaluation, which depending upon faculty responses, could be a foundation for inquiring as to how SETs may or may not be utilized even more as a tool for faculty in reflecting upon their own teaching practices as well as one indicator of students’ (self-reported) progress towards learning and course outcomes.

The findings from a faculty survey within my own division provided an appropriate scope for inquiry into my research questions and preliminary steps for initiating Patton’s (2008) UFE framework. The UFE framework’s concept of engaging evaluations applies directly with my overall inquiry, the design of my research questions, and my intention behind the development, implementation, and analysis of my own survey instrument. UFE methodology argues for evaluations to focus on the utility of the

information for action and decision-making, and I plan to use this study as the foundational groundwork in attempting to develop a process (program) that explicitly connects my data collection and analysis with enhancing achievement through feedback and use of the data to inform decisions and increase understanding of SETs and faculty's perceptions of them. Patton (2008) also describes a primary intention for his UFE model as building capacity of those involved, which directly applies to faculty and students. This could be incredibly powerful in developing more inclusive models of SETs and their formative use by faculty. Thus, it is my hope that through this study and Patton's UFE framework I will be able to help the faculty and the institution learn more about faculty perceptions of SETs at my own institution for two primary purposes.

Primarily, I wish to use Patton's (2008) UFE framework and the data I collect to discern if any recommendations might be made in light of "improvement-oriented, formative evaluations" that indicated "strengths and weaknesses with the expectation that both will be found and each can be used to inform an ongoing cycle of reflection and innovation" (p. 116) within my own division at my institution. Patton's process is a viable framework for perhaps assisting faculty in utilizing SETs more formatively. If faculty self-reported utilizing SETs, and perceiving their merit, in helping to reflect upon and perhaps improve their own teaching practices, then perhaps there is a need for process improvement based on Patton's UFE that could be developed within my division and institution. If faculty responses reported little trust or use in SETs for formative purposes, then Patton's framework would also be viable for envisioning a process with faculty input that might ultimately help change the perception of SETs utility in reflective teaching practices.

One fundamental purpose of evaluations as Patton (2008) defines them can be for “formative improvement and learning,” which allows administrators, staff, and participants who are regularly involved in the program to all be the primary users (p. 139). This model would potentially make UFE through this lens of formative improvement learning a very viable framework for interpreting and incorporating the results of faculty perceptions of SETs into a more inclusive model through which all stakeholders as defined by Patton become more involved through making adjustments, acting upon feedback, and enhancing implementation and outcomes (2008). As Patton describes, the primary purpose of using evaluations in this regard is to learn and improve programs (in my research, teaching and overall student learning). Incorporating aspects of Patton’s UFE through this “learning” lens asks stakeholders to examine what is and isn’t working as well, discover strengths and weaknesses, and to inquire more deeply into participant (student) reactions—to discern if certain patterns emerge among different groups (e.g. classes or majors)—in an attempt to increase outcomes (e.g. teaching effectiveness and student learning). Participant feedback, reflective practice and appreciative inquiry are common approaches in this particular practice of UFE, and Patton describes the key factors in using UFE for formative learning and improvement as “creating a learning climate, openness to feedback and change” and “trust” that lead to use of the findings in “actionable” steps.

This aspect of UFE might be very viable in (re)evaluating SETs as perhaps being a viable and impactful formative tool for faculty within my division. If faculty seemed to be receptive to SETs for formative purposes then they might be receptive to this model of UFE for furthering the utility of SETs in their teaching practices. If faculty reported

apprehension or skepticism of SETs for formative purposes, then this model of UFE might help change that perception by incorporating faculty perceptions into an initial inquiry as to why they do not deem (or use) SETs formatively. I would also incorporate previous research highlighted above indicating effective strategies for overcoming misconceptions of SETs perceived utility by faculty into a UFE framework within my case.

It is also important to include faculty perceptions of SETs as “summative, judgment-oriented evaluations” that help determine the “overall merit, worth, significance, or value” of SETs (Patton, 2008, p. 113) by including the faculty as primary stakeholders in a process that profoundly impacts them; SETs as a tool used in summative evaluations within my division. Faculty responses to the survey questions related to their perceptions of SETs as summative evaluations are crucial in this regard, and could lead to different courses of possible action through Patton’s framework. If their responses indicated their belief that SETs are used primarily for accountability, then not only could Patton’s framework for accountability evaluations be instructive for next steps my division and institution might consider, but perhaps a process (and culture) change designed to help faculty envision and perhaps utilize SETs more for “formative improvement and learning” as Patton describes it above. If faculty saw some merit in SETs but believe they are overused as summative measurements, then perhaps this information can help administrators discern how they’d like to use SETs; would the feedback from my research provide incentive to make changes as to how they are used by the institution? Does the institution wish to use SETs for summative purposes, or might there be a receptive opportunity to assist faculty in utilizing them more formatively (and

if so, how might the institution decouple the summative use from the formative use SETs provide? The hope is that this study will help inform both the faculty and the institution's perspectives, and utilization, of SETs through Patton's concept of an "accountability system" through which he argues "the extent to which those held accountable actually have the capacity to achieve those things over which they are held accountable" is critical (p. 121).

Thus, Patton's UFE framework (2008) corresponded very well with my research questions in gauging faculty perceptions of SETs as evaluations by including them more as stakeholders in the overall process; what are their thoughts on SETs as formative and summative measurements of teaching? How might their perceptions as to the utility or current practice of incorporating information from SETs correspond with Patton's framework for evaluations to be used by the primary intended users? How might the faculty's own perceptions lead to perhaps a more "improvement-oriented, formative" and summative processes through Patton's UFE framework? These questions, initially examined at the level of my division, and through Patton's framework for stakeholder involvement in teaching evaluation processes, are crucial for collaborative efforts of continuously improving student learning as one component of increasingly required accountability in higher education. These questions and this process are relevant not only on the individual institutional level, but also as part of the ever-increasing larger conversations on student learning in higher education at state and the national level. As the multi-stakeholder National Commission on Accountability in Higher Education (2005) declared, "Real improvement in higher education will come when accountability in higher education is a democratic process through which shared goals are explicitly

established, progress is measured, and work to improve performance is motivated and guided” (p. 7).

In sum, the primary focus of the study was how SETs, analyzed through a UFE framework (Patton, 2008), might impact perceptions and practice of faculty, and also the extent that such information informs summative evaluations. This study was designed to not only be an act of inquiry for myself as a researcher, but also to inform and initiate action through the UFE framework in further evaluating SETs. There are several profound potential outcomes for my division, and perhaps subsequently, my institution. Incorporating faculty perceptions of SETs as formative and summative tools could begin a larger conversation within my division and potentially at my institution about faculty perception of SETs’ efficacy in their utilization of SETs. Patton’s UFE is also conducive for better engaging teaching faculty as primary stakeholders to participate more inclusively in the SET evaluation process itself, through a combination of “designing instruments...overseeing implementation, and interpreting findings” (p. 77).

Additionally, the study seeks to add to the larger conversation in the literature as an evolving case study by including faculty perceptions of SETs concerning their formative and summative utility, and applying Patton’s UFE framework to develop a more inclusive, reflective, and ideally, beneficial process of incorporating SETs into pedagogy and administrative practice.

DATA COLLECTION METHODS: SURVEY DESIGN

Nardi (2005) argued that each method for collecting data has its own advantages and disadvantages, but he described surveys as being “ideal for asking about opinions and attitudes” and “better for sensitive and personal topics,” which fits precisely within my

research questions (pp. 18-19). The purpose of the survey then is to help lay the foundation for both “descriptive research” in order to obtain as much information about the characteristics and issues related to this topic (Nardi, 2005, pp. 9-10). A survey is also well suited for providing the complimentary merits merging quantitative and qualitative research. Sechrest and Sidani (1995) described an inquiry whereby a questionnaire and anecdotal evidence (e.g. “conversations”) provided what they refer to as “formulaic” (quantitative) and “clinical” (qualitative) approaches, respectively (pp. 82-84). Patton (2008) contended that, while both qualitative and quantitative data are beneficial to UFE, “in many cases, both qualitative and quantitative methods should be used together” and “that there are no logical reasons why both kinds of data cannot be used together” (p. 438). Moreover, he argued for “methodological pluralism” by which evaluators would be “wise” to not use qualitative or quantitative approaches and analyses as a singular approach (Patton, p. 466) and to ideally use both as “each can contribute in important ways” (Patton, p. 434).

Therefore my survey was designed to capture both quantitative and qualitative data through scaled questions for descriptive statistical analysis as well as open-ended questions allowing respondents to clarify and describe their own experiences and opinions concerning SETs. While it was informative to see what patterns emerged from the quantitative data, the qualitative data provided deeper and invaluable insight into faculty perceptions of SETs. Patton (2008) summarized this approach by stating that “qualitative data capture personal meaning and portray the diversity of ways people express themselves; quantitative data facilitate comparisons because all program participants respond to the same questions on standardized scales within predetermined

response categories” (p. 435). He further described the merits of a mixed approach by stating that “qualitative inquiry involves small ‘purposeful samples’ of information-rich cases” that provides “rich” data that is “particularly useful” (Patton, p. 459) within the UFE framework, such as that of my own division at my institution (Patton, p. 458). He also posited that while generalizability of both qualitative and quantitative findings within a UFE framework can be considered since program evaluations (e.g. perceptions and reported use of SETs) are certainly context specific, he argues that “extrapolation” as Cronbach and others in the Stanford Evaluation Consortium (1980) define it is a more appropriate expectation as they are “modest speculations on the likely applicability of findings to other situations under similar, but not identical, conditions” (Patton, p. 459). Thus, a mixed-methods approach through a survey instrument was an integral part of my own attempt to use my research in exploring, describing, and ideally, evaluating the construct and efficacy of SETs by first sampling faculty perceptions of them as a foundation for future incorporation of their responses and own interpretation of the data into a framework for further reflection and discussion about SET use for formative and summative purposes based on Patton’s UFE framework.

In utilizing Dillman, et al.’s (2009) suggestions for developing survey questions, I felt it prudent to compare my research on assessing faculty perceptions and attitudes towards instructional surveys conducted to measure their teaching effectiveness with the plethora of issues the authors advise considering before drafting questions (Dillman, Smyth, & Christian, 2009). My questions below have attempted to heed the authors’ cautionary/guiding questions below:

- 1) What survey mode(s) will be used to ask the questions?

- 2) Is this question being repeated from another survey, and/or will answers be compared to previously collected data?
- 3) Will respondents be willing and motivated to answer correctly?
- 4) What type of information is the question asking for?

I selected an Internet survey mode for several major reasons; confidentiality, anonymity, and convenience for respondents, and attempting to garner as much qualitative and quantitative data within my case as possible. In order to (best attempt to) truly capture faculty's perceptions and attitudes, respondents' answers needed to be anonymous and great care was taken in my solicitation to convey the importance I placed on anonymity and confidentiality (see survey instrument and messages below). As Dillman, Smyth, & Christian (2009) argued, "of considerable concern for some survey respondents is how the information will be used and who will have access to it, particularly if they are disclosing information that is personal or sensitive" (p. 28). Given the potentially sensitive nature of faculty self-reporting their perceptions and use of SETs, and the fact that SETs are utilized by administration for formative and summative purposes, I felt a strong need for the faculty to trust the claimed intention and use of my survey and tried to convey that in my survey invitations and instructions. Lastly, I obtained site access through permission from the Provost and Office of Academic Affairs at my institution, and I ensured that my survey instrument was approved by both UVM's and the site-institution's IRBs. This approval was conveyed to prospective respondents.

Dillman, Smyth, & Christian (2009) highlight the advantage of reducing the "perceived costs of responding" to surveys, and that "one of the biggest costs of responding to survey requests is the time it takes to complete the survey" (p. 26). In an

effort to reduce these “costs” I refined the survey to a shorter instrument of 15 questions through numerous drafts, modifications, and feedback (the initial survey draft was over 30 questions). Moreover, further revisions indicated several redundancies in the nature of some questions, and the final version (below) was paired down to 15 questions (see Appendix A for survey instrument). The majority of questions were developed as closed-ended with ordinal scales to measure gradations of “opinions, attitudes, behaviors, and attributes” (Dillman, Smyth, & Christian, p. 135-136), with response categories limited to no more than five to allow a realistic “continuum” in the respondents minds without overburdening them and thus potentially negatively impacting their response or participation in each question (Dillman, Smyth, & Christian, p. 137), and the scales were also constructed with an intent to provide category balance (Dillman, Smyth, & Christian, p. 141). Additionally, several of the closed-ended ordinal scale questions allowed respondents an open-ended text box and thus opportunity to clarify or make additional comments should they wish to do so (as opposed to requiring such a response or the instrument containing more open-ended questions). Only two questions were specifically designed as open-ended, as these were “expensive” for respondents given the time investment to answer each one (Dillman, Smyth, & Christian, p. 113). Thus, I attempted to maximize opportunity for responses to the survey questions by providing closed-ended scalar questions yet with the voluntary option to expand upon answer choices.

The entire population of full- and part-time faculty within my division were surveyed. As the survey population of all full- and part-time faculty is relatively small (n=30), this survey attempted to avoid sampling error by collecting data from all of the faculty within my division. Steps were taken (below) to obtain as high a response rate as

possible to try and gain a confidence level of 90% and thus increase the validity of the findings. Additionally, I wanted to initiate my inquiry into faculty perceptions of SETs and their utility first within my own division in order to more tangibly apply some of Patton's framework into a larger discussion within my division, possibly leading to some actionable steps within his framework. A smaller population and sample size mean inferential statistical analysis cannot be conducted, however, descriptive analysis was possible. Moreover, since the primary focus was on faculty perceptions of SETs and their self-reported use of them, the qualitative data received in addition to the quantitative was of particular interest to this study.

As to eliciting high response rates, I employed the suggestions of Dillman, Smyth, & Christian (2009, pp. 23-24) for "increasing the benefits of participation." Information was provided about the survey (through prenotice letters, announcements in person and via institutional email), and the purpose of this survey was explained as designed to determine their anonymous perceptions of SETs in general as well as the IDEA Student Ratings of Instruction. As mentioned, the survey length was also reduced in part to help elicit higher response rates. Each survey question was refined repeatedly in light of Dillman, Smyth, & Christian's (2009) guidelines for question word choice, response wording and scale, and question ordering. Questions that could potentially make respondents in a smaller population feel categorized were not made mandatory to further ensure their trust in the instrument as well as elicit higher response rates. This survey was, ultimately, a chance for faculty to demonstrate how they felt about such a survey, why it is administered, and to what, and whose, benefit. Moreover, since the IDEA Student Ratings of Instruction is currently used by administration as a significant portion

of faculty teaching evaluations, my instrument was a medium through which they could express their opinions about this practice. Given the high response rate and rich qualitative comments received (discussed below), it appears that faculty did indeed feel motivated and comfortable in providing their valuable perceptions of SETs utility.

PARTICIPATORY ACTION & ITERATIVE DATA ANALYSIS

Of paramount importance, however, was my making explicit the intent of my survey to be a positive step towards participatory action in assessing student learning by including the faculty's voice concerning SETs, and that the survey and its data will be used to help generate future conversations about SETs and how they can—and perhaps should be—utilized through Patton's UFE framework. Angelo (2002) argued that “the main attraction and promise” of “a Scholarship of Assessment” among faculty is “its focus on improving student learning, a focus that many faculty share” (pp. 197-198). Moreover, since a culture of assessment is slowly taking root at my institution, such a survey was timely and will ideally initiate further discussion so that “something important will ultimately happen as a result of this survey,” namely, that faculty perceptions of SETs and their practice will be heard and further discussed (Dillman, Smyth, & Christian, 2009, p. 28). Another strategy for increasing action through inquiries such as this research project is what Angelo (2002) discussed as the early engagement of “opinion leaders” who can become not only participants in initial analyses of my survey data, but ideally, will become champions of further discussion and action with their colleagues (p. 197). This would be an integral part of Patton's (2008) “intended use by intended user” tenet of his UFE framework.

The data collected was not compared to previously selected data, as no such instrument or process currently exists at my institution that I am aware of. I followed data collection and protection protocols of both UVM and the site institution's IRBs (again, both of which granted approval for my study) to ensure that proper steps were taken to protect the rights and well-being of the faculty who took this survey, and of course, of how the data is protected, and disseminated. I approached data analysis with a goal of utilizing it as part of an iterative process whereby the data and results of the survey will be used as a foundation for further inquiry through an ensuing UFE process evaluating faculty perceptions of SET utility and practice in general, as well as providing a feedback mechanism for conversations between faculty and administration as to the perceived and actualized use of the IDEA Student Ratings of Instruction specifically. This again was the precise intent behind my selection of Patton's UFE framework.

In analyzing the data, I drew primarily upon Patton's (2008) UFE framework which as aforementioned he posits should be based on mixed-methods of collection and a synthesis of both qualitative and quantitative analysis that moves beyond the methodological paradigms of one or the other (p. 466). As Glesne (2006) describes, data analysis at its essence involves the synthesis and interpretation of data in order to make sense of what might be learned through descriptions, explanations, and positing of hypotheses (p. 147). Stake (1995) argues that data analysis is an iterative and reflective process including the use of field notes by the researcher to use for their own interpretations and reflections of the data collection and analysis. Miles & Huberman (1994) define qualitative data analysis as containing three "concurrent flows of activity" including data reduction, display, and conclusion drawing and verification (p. 10). The

authors argue that data reduction such as interpretation and paraphrasing is an analytic choice at every juncture, and should be conducted within the context of the data being displayed in an “immediately accessible, compact form” so that ultimately conclusions can be drawn by the analyst and, ideally, verified either through further reflection by the analyst or through “intersubjective consensus” among colleagues (Miles & Huberman, pp.10-12).

Miles & Huberman’s “interactive model of data analysis” (1994, p.12) contains remarkable parallels to Patton’s UFE framework for data analysis and use. Patton (2008) developed a four-stage framework for analyzing and engaging findings, the latter being a seminal goal of any utilization-focused evaluation (p.478). Similar to Miles & Huberman’s (1994) data reduction, display, and conclusion drawing mentioned above, Patton’s (2008) four-stage model first calls for a description and analysis of both qualitative and quantitative data into a format revealing patterns specifically so that intended users can readily understand and interpret the results; for Patton, “analysis” is the organization of data into understandable formats that best reveal patterns (pp. 478-481), similar to Miles & Huberman’s (1994) first two steps of data reduction and display. Patton’s (2008) second step involves “interpreting” the findings by “going beyond the data” to add “context, determine meaning, and tease out substantive significance” (p. 478). The first iteration will thus involve my own interpretation of the quantitative and qualitative data, in “deciding what the findings mean” (p. 486). Throughout this process I will take field notes, or “analytic notes,” which Glesne (2006) notes is a part of the data analysis process itself and can be instrumental in understanding patterns and themes (p. 59). Key to UFE is the inclusion of additional reflection by capturing and reviewing field

notes, which become a crucial part of the analysis by helping to guide the analysis itself and in discerning emerging patterns from the data (Patton, 2008). The third step is that of “judgment,” in which “values” are applied to the data and interpretation in determining merit upon certain findings versus others (Patton, 2008, p. 500). Lastly, Patton’s fourth step in his framework for analyzing and engaging findings is “recommendations” which “add action” (p. 478).

The study followed Patton’s (2008) articulation of UFE as a four-stage model of engaging evidence (analysis, interpretation, judgment, and recommendations), which should be conducted by the primary evaluator (myself in this study) concurrently with the primary stakeholders (faculty within my division). However, for this research study, my analytic approach contained a two-step process given my overall goals for this research study to be a pilot project that translates into a larger conversation about current perceptions and use of SETs. The first step was my own analytical iteration of Patton’s four-stage framework so that I may not only develop a journal article meeting the requirements of the dissertation, but also so that I become more proficient in Patton’s UFE framework before the second-step of engaging my colleagues in his four-stage process. As Patton (2008) describes, I, as the primary evaluator, can analyze the data through his four-stage process prior to sharing my findings and thus leading my colleagues through their own iteration of his process (p. 500). Moreover, I wished to avoid the scenario Patton describes in which poorly constructed (or received) recommendations can detract from all previous three steps and discredit the entire evaluation (p. 501) by going through his process first before attempting to do so with my

colleagues. I believe this iterative process of analysis will better enable possibilities for truly engaging the evidence in a call for action as Patton's UFE framework intends.

Pursuant to Miles & Huberman's (1994) "verification" and primarily Patton's (2008) UFE framework, it is indeed critical that, as a part of further creating a scholarship of assessment at my institution, the results, and ensuing collectively-developed action plans from my study be continually shared with colleagues at my institution in addition to my attempts to add to the current literature through an initial (Chapter below) and subsequent journal article(s). Patton (2008) contends that findings contribute to theory through the stakeholders' (respondents') descriptions and exploration of the findings, and should be analyzed not only drawing upon descriptive and inferential statistics and content analysis, but ultimately, through lessons learned and extrapolations of the findings developing shared and useful findings for stakeholders (p. 467). It is my hope that this instrument and my own initial as well as collective analysis through Patton's UFE framework become an embedded part of a continuing conversation between faculty and administration—and perhaps students themselves—about students' perceptions of teaching effectiveness in not only my own institution, but ideally at others as well.

RESEARCHER'S ROLE

While I believe my qualifications for pursuing this research are undoubtedly in progress, my experiences as an educator and administrator in higher education, combined with my experiences and content knowledge/skills developing through the Educational Leadership and Policy Studies doctoral program at the University of Vermont, make me qualified to pursue the implementation of this instrument and aforementioned research. I do, however, view my role as participatory and reflective; I will need to continually learn

just as much from the UFE process itself as I learn from the outputs of the process. As a faculty-administrator within my division, one of my responsibilities is to work with faculty concerning all matters of assessment; given that my institution administers mandatory student surveys of faculty teaching effectiveness once per semester, I am professionally as well as pedagogically-motivated to explore the construct of faculty perceptions of these evaluations. I personally believe SETs in general have potential pedagogical benefits for faculty (and thus students), but as aforementioned, I sensed that perhaps skepticism towards SETs how they were used by the institution might be overshadowing perceived value among faculty. It is therefore my hope to discover more about the faculty perceptions of these evaluations; how they view them, how they do (or do not) use them, and how they perceive the administrations' use of them. It should be noted that in my current administrative capacity, I review part-time faculty IDEA ratings and provide feedback, and do not review full-time faculty ratings.

In my personal opinion, I believe my division and institution need to have deeper—and more inclusive—discussions concerning SETs and the IDEA instrument. Faculty, staff, and administrators at my institution care deeply about student learning and teaching effectiveness, and I believe would be receptive to such discussions. However, I suspect a disconnect between institutional emphasis on SETs and the IDEA instrument and faculty perceptions of their utility. On one hand, administrators in each division within my institution as well as the senior administration keep reiterating the importance of the IDEA instrument and reports, as well as the need for faculty to correctly align course objectives within this survey with those of their class so as to avoid any major

disconnects in student responses or learning outcomes.² On the other hand, I suspect faculty feel these surveys are too heavily utilized by the administration in performance and compensation decisions, and feel to some extent that the surveys themselves do not accurately reflect the learning occurring in their classrooms. In my role as an educator, a faculty member, and an administrator, I have personal and professional interests and obligations to explore the accuracy and extent of faculty perceptions of SETs. I employed my dual-role in every aspect of this study design, from the research questions to the survey instrument questions themselves; I tried to see this study from both an administrative and a faculty perspective, as I am both.

Throughout the initial analysis in this study and all subsequent analyses, and through following Patton's four-stage process for "engaging evidence," I will conscientiously attempt to consider my own biases towards SETs, and attempt to limit their influence on my data analysis, interpretation, judgment and recommendations. Creswell (2007) reminds us that researcher bias is a critical component of validation, as is member checking through garnering respondents' opinions of my interpretations, perhaps through a focus group (p. 208). Member checking is also an essential part of Patton's UFE framework for engaging evidence as discussed above, and thus I utilized an impartial peer debriefer who as Creswell (2007) contended played an invaluable role as a "devil's advocate" (p. 208) in checking not only my interpretations but also in my biases (they did not have access to the raw data). In full disclosure of bias, I tend to agree with the "protagonists" mentioned in the literature review section of this study who argue that

² The IDEA survey has 12 predetermined course objectives, of which faculty are encouraged to choose which of these objectives are "Essential," "Important," or of "Minor/No Importance" to each specific course being assessed.

SETs and garnering students' perceptions of their own learning is an important part of measuring student learning and teaching effectiveness, and can be very beneficial formatively in perpetuating or modifying pedagogy in the classroom (I personally utilize SETs in my own reflections of my teaching practices). However, I also agree with some of the skeptics' and protagonists' cautions that SETs not be used as solitary measurements of student learning nor teaching effectiveness, and should not be the sole basis for summative evaluations.

LIMITATIONS & DELIMITATIONS OF STUDY

Some of the primary limitations of this study were related both to the site institution and its use in this study as the single case. The research site is one division within a private four-year, career-oriented college the northeastern United States. Moreover, the division and institution are focused primarily on a practitioner-scholar model placing much higher expectations for faculty on teaching than research, with a normal teaching load of eight courses per academic year. Another limitation, which despite aforementioned efforts, is the very nature of respondents' self-reporting. Despite all efforts in transmitting anonymity and confidentiality to all potential respondents, it is possible that part-time faculty felt more vulnerable to IDEA ratings and their summative use, although the site institution does not offer tenure. Nonetheless, the respondents' status as part-time or full-time might have affected their participation and responses. A primary delimitation of this study was limiting the survey to 30 full- and part-time undergraduate faculty within my current division at my institution. Other delimitations are the focus on my division within my institution as a single case study, and restricting

my survey questions and invited population to faculty who are involved primarily in undergraduate teaching.

Yet these limitations and delimitations also make this institution and this study's survey population very relevant to further inquiry concerning SETs precisely because of the heavy emphasis on, and culture of, teaching. Sechrest and Sidani (1995) describe the intent of inquiry as an attempt to reduce uncertainty through the obtainment of relevant information; "Information is contained in any message that *reduces uncertainty* about an issue of interest (Sechrest & Sidani, 1995). In "Why We Do Research," Nardi (2005) described the understandable behavior of using reflections of our experiences to "make sense of reality" (p. 4). He also states "we are generally good at assessing the climate of opinions about controversial topics among our peers and those whom we encounter regularly..."(Nardi, 2005, p. 4). The "reflections" captured within this study are integral to not only my gaining a deeper understanding of faculty perceptions towards SETs, but moreover, are critical foundational steps towards the development of an evaluation process based on the UFE framework.

While I would generally agree with Nardi within his context of our tendency to use such "limited sampling" to draw larger conclusions about the larger world around us, I would politely disagree within the construct of student responses to faculty evaluations of teaching effectiveness; certainly one of the more "controversial topics" in academia (and other echelons of education). Despite the use of my own reflections, and my feeling that I have a good general sense of my peers' opinions, I am also aware that such a controversial—and personal—topic should not be assumed to be understood simply because I encounter my colleagues regularly. Despite Nardi's (2005) valid concern of

overgeneralization when we assign patterns to groups and infer conclusions based on few observations, this study still provided invaluable information for me, my division, and perhaps my institution and its practices involving SETs. This study and its findings may also provide some added value to the research and practice related to SETs at other institutions through, as Patton (2008) describes above, through extrapolations from this case that others may find analogous or beneficial to their contexts. I chose to begin my inquiry to my division specifically so that the chances of developing an evaluation process of SETs might be more likely, as starting on a divisional scale is more manageable and advisable. I also did not wish to elevate the entire institution's expectations for examining SETs and their impact without first piloting such an inquiry within my own division, and participating in a divisional-level UFE evaluation process prior to an institution-wide evaluation.

CHAPTER FOUR

JOURNAL ARTICLE

Including Faculty Perceptions in the Use of Student Evaluations of Teaching

Abstract

One increasingly common method of measurement is student evaluations of teaching (SETs) used as formative (pedagogical practices) and summative (employee reviews) assessments of faculty teaching. Through the use of surveys of 27 full- and part-time faculty within one division at a private, four-year teaching-focused college, this study explored faculty perceptions as an initial step in a larger process seeking to evaluate perceived and potential efficacy of SETs. Both quantitative and qualitative data were collected and analyzed as a foundation for using Patton's (2008) Utilization-Focused Evaluation (UFE) framework for engaging evidence based upon a four-stage process in which evaluation findings are analyzed, interpreted, judged, and recommendations for action are generated, with all steps involving intended users. Overall, the study data suggests that the faculty surveyed were generally supportive of using SETs for formative assessments, and strongly reported their importance and use for evaluating their own pedagogy. They also indicated that they relied primarily upon the students' written qualitative comments rather than the quantitative reports generated by externally determined scaled-questions on the SETs, and indicated a desire for a more robust evaluation process that would provide more institutional support in interpreting and incorporating all feedback. Faculty also reported the importance of SETs as part of their own summative evaluations, yet expressed concern about overreliance upon them and again indicated a desire for a more meaningful process. Thus, the study suggests that Patton's model has the potential to maximize the utility of SETs for many relevant stakeholders, especially faculty and their evaluators.

Introduction

Perhaps few topics in undergraduate higher education elicit a broader range of responses than measurements of student learning amidst an ever-growing demand for accountability, compounded by concerns of the rising cost of higher education and mounting student debt. Pressure to measure various aspects of student learning through

“competency” or “outcomes” assessments has increased considerably, due in part to increasing examinations of how much college students are actually learning.

Accountability for student learning has been the subject of increased attention from stakeholders both within and outside higher education. In their landmark study of over 2,000 undergraduates from nearly two dozen different types of institutions, Arum & Roska (2010) contend that more than a third of graduates demonstrated no significant proficiency in critical thinking, analytical reasoning, and written communications based off their assessment of data from the widely used “Collegiate Learning Assessment” (CLA) exam. Kuh, Kinzie, et. al. (2006) report that several states have begun demanding measurements of student success including required data sharing of graduation rates. A joint commission of policy makers, higher education administrators and faculty, and business and community leaders formed the “National Commission on Accountability in Higher Education” (2005) and argued that more accountability in higher education is vital to continuous improvement not only of the students’ learning, but also because the knowledge and skills college graduates do—or do not—posses upon entering society and the economy is a critical issue of state and national interest.

As Doubleday (2013) illustrates, the U.S. federal government is also demanding more accountability in higher education, including proposed legislation tying federal financial aid more directly to measures of student performance, as well as “The White House College Scorecard.” Developed by the U.S. Department of Education’s College Affordability and Transparency Center (2014), the Scorecard allows prospective students and families to readily compare institutions across several measures, primarily tuition costs and graduation rates. Supported by President Obama, the U.S. Department of

Education (2011) issued a report entitled “Meeting the Nation's 2020 Goal: State Targets for Increasing the Number and Percentage of College Graduates with Degrees”, which issued a national charge to increase the national proportion of college graduates by 50% (from 40% to 60%) by 2020. Yet along with a desire to increase graduation rates is an even larger increase in the average cost of obtaining a college degree, which has increased exponentially over the last thirty years. The National Center for Public Policy and Higher Education (2008) reported between 1982-2007, the average cost of college tuition increased 439%; nearly 200% more than the rise in average health care costs, and outpacing average family incomes by nearly 300%. With increased concerns for accountability of learning accompanied by a national increase in enrollment and perpetually rising costs, demands for measurement of student learning in higher (Cronbach, 1980) education are likely to intensify.

Student Evaluations of Teaching: Measuring Learning and Teaching Effectiveness

One increasingly common method of attempting to measure undergraduate student learning is student evaluations of teaching (SETs), used both as formative (pedagogical practices) and summative (employee reviews) assessments of faculty teaching. SETs have been in use for quite sometime and are common throughout academe (Seldin, 1993; Astin & Antonio, 2012). While institutions of higher learning utilize other evaluations such as administrator evaluations, peer evaluations, and self-evaluations to measure teaching effectiveness (Smith, 2007), student evaluations have perhaps become the most commonly used (Richardson, 2005; Berk, 2013). Proponents of SETs have long contended that they are a necessary component in measuring the quality of education a student receives (Marsh & Roche, 1993), that they further enable

educators to reflect upon their own pedagogy (Centra, 1993), and that they are a valid component in summative evaluations of faculty (Cashin, 1988).

SETs critics such as Schmelkin, Spencer, & Gellman (1997), and Simpson & Siguaw (2000), have long argued that SETs are ineffective measures of faculty teaching rife with issues; the measurements themselves are invalid and unreliable, students are not qualified evaluators of teaching, and may even lower educational standards if faculty feel pressure for higher ratings in summative evaluations by their institutions. Common assumptions and conclusions primarily indicate faculty underutilization of SETs due to suspicion and mistrust of their efficacy, as well as institutions' overreliance upon them for formative and summative evaluations; "one gets the impression from the anecdotal literature that there is widespread resistance on the part of faculty to student ratings" (Schmelkin, Spencer, & Gellman, 1997, p. 576). With the increasing use and reliance on SETs for increased accountability, SETs and how they are utilized (or underutilized) will become even more of a salient issue. Moreover, SETs have been, and will likely continue to be, a contentious and high-stakes issue for faculty and administrators not only as they relate to accountability of student learning and teaching effectiveness, but also because results and interpretations are quite often tied to promotion, tenure, salary and merit-pay decisions (McCarthy, 2012). This paper explores faculty perceptions of the value of student evaluations of their teaching and the impact of such data on their pedagogy as an initial part of an action-research evaluation project within my institution.

The next section places SETs within the historical evolution of the assessment movement and their use in faculty evaluations. Then the conceptual and methodological

conceptualization of the research conducted will be outlined followed by an overview of the findings. The final section discusses the implications and limitations of the study.

The Historical Foundation of the Assessment Movement

An examination of the assessment movement in higher education provides more context for the increasing use of, and debates surrounding, SETs in measuring student learning and teaching effectiveness. SETs have become a de facto tool for measurement with strong proponents of its benefits yet also with continued concerns for its shortcomings (Astin, 1977; Cashin, 1988; Marsh & Roche, 1993; Kite, 2012). Although the reliability and validity of these measurements has long been a matter of debate, the debates have generally evolved from whether or not SETs should be used to how they should be used more effectively (Marsh, 1982; Aleamoni, 1987; Marsh & Roche, 1993; McKeachie, 1997; Abrami, 2001; Gravestock & Gregor-Greenleaf, 2008; Kite, 2012). A richer appreciation for the different arguments surrounding SET use for both formative (pedagogy) and summative (employee review) reasons provide context for the importance of this study, as well as the increased inclusion of faculty perceptions and self-reported use of them.

Measuring Student Learning in Higher Education

Assessment as a measure of student learning in higher education has been in practice for decades, with increased attention since the 1980s. The concept of assessing “value-added” learning has its roots in a series of works by Astin (1977), who utilized longitudinal studies involving over 200,000 students and 300 institutions to examine the overall benefits of higher education for students. Among his many findings, Astin argues that student satisfaction—and thus, likely retention—is impacted not only by their

relationships with faculty and level of contact with them, but also in students' perceptions of their own learning and development. As Astin & Antonio (2012) illustrate, Harris' 1970 study argued that course grades were poor indicators of student learning. Martell and Calderon (2005) highlight the increased concern over the "quality of learning at higher education institutions" being largely driven in the early 1980s by discussions focusing on "the failure of higher education institutions as centers of learning" (pp. 1-2). Walvoord (2004) also attributes much of the "assessment movement" within higher education as being driven by the concerns of extrinsic groups—politicians, employers, learning communities, movements for problem-based learning, etc.—"who were disappointed with the quality of college graduates and the rising costs of higher education"(p. 5). These "external stakeholders" as Ewell (2002) refers to them, have grown to include accreditation agencies, market forces, and the media, all of which have become more "performance-conscious" and "data-hungry" than ever, making assessment for institutions of higher education "an unavoidable condition of doing business: institutions can no more abandon assessment than they can do without a development office" (p. 22). Assessment will continue to be a critical component of the academy, and likely, with only increased scrutiny and expectations as the costs of higher education and emphasis on graduation rates continue to rise.

SETs in Formative and Summative Evaluations

However, the use of SETs as a common component of assessment in measuring student learning has also led to much debate not only about their validity and reliability, but more recently, about their utility as a tool in formative and summative evaluations. Benton and Cashin (2012) contend most research has focused on debates over the

reliability and validity of SETs in measuring student learning and teaching effectiveness. As Gravestock and Gregor-Greenleaf (2008) summarize, collective research questioning reliability and validity of SETs claims numerous variables may influence how students rate a course and an instructor. Additionally, Berk (2013) notes the common contention that students themselves are not fully qualified to rate instructors which needs to be taken into consideration concerning SET use.

Yet as Gravestock & Gregor-Greenleaf (2008) note, the literature generally supports the validity and reliability of SETs and that, in fact, “It should be noted that any effect on overall ratings from any of these particular variables, even when statistically significant, is almost always very small—often changing the ratings by less than one-tenth of 1%,” (p. 44). Greenwald (1997) analyzed much of the research on the validity of SETs over a twenty-year period from 1975-1995, and indicated the majority of findings argued for their validity. Well-constructed and tested SETs appear to consistently measure specific components of teaching practice (Centra, 1993; Marsh, 1984; Aleamoni, 1999). McKeachie (1997) also reported SETs to be a valid method for indicating teaching effectiveness. Theall & Franklin (2001) argued that attempts to discredit SETs’ validity and reliability was akin to a “witch hunt” and instead, focus should be on properly utilizing their findings.

Concerns over use of SETs for summative evaluations have been voiced for years (Wilson, 1988) and have been a continuing concern for faculty and researchers (Schmelkin, Spencer, & Gellman, 1997; Zabaleta, 2007; Berk, 2013). Astin & Antonio (2012) argue the value of SETs as formative assessments and benefits to faculty pedagogy can be detracted by their simultaneous use as summative measurements. They

suggest using other data demonstrating student progress as well as products of students' academic work as more direct evidence of faculty teaching effectiveness ("faculty outcomes"), and state that if an institution uses student assessments of teaching, they should be primarily for the benefit of the instructor for formative reflection. Astin and Antonio (2012) also claim that one SET instrument should not be used for both formative and summative evaluations and that traditional motivations for assessing faculty are too often driven by purposes of hiring, tenuring, and promoting rather than enhancing the development of the faculty as researchers and teachers. In sum, the implications for a deeper understanding of perceptions and practices concerning use of SETs in formative and summative assessments have profound implications for higher education. Moreover, a framework for potentially applying faculty perceptions in an evaluation of SET use would likely be beneficial for all stakeholders, especially faculty.

Conceptual Framework and Methodology

Utilization-Focused Evaluation: An Inclusive, Action-Oriented Framework

Given scant research exists on understanding faculty perceptions of SETs and their potential utility for institutional improvement, this study utilized a formative evaluation approach based upon Patton's (2008) framework for "Utilization-Focused Evaluation" ("UFE"). Specifically, this study explores how UFE is a viable evaluation framework for informing continual institutional improvement concerning SET use at my own institution. UFE is well-suited for this type of on-going participatory study, as Patton (2008) described:

Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything that is done, from beginning to

end, will affect use. Use concerns how real people in the real world apply evaluation findings...Therefore the focus in utilization-focused evaluation is on intended use by intended users. (p. 37)

The UFE methodology was chosen for its inclusiveness of stakeholders most affected by the evaluation throughout the entire process thus perhaps enabling them a greater role in the process of how SETs are used in decision making, and primarily, for its “commitment to use” emphasizing “knowledge for action—finding out things that can be used” (Patton, 2008, p. 51). This methodological framework is conducive for inquiring about faculty’s perceptions and use of SETs, to what degree they do or do not utilize them to inform their own pedagogical practices, and to what extent faculty feel SETs should be used by administrators as a component of measuring their effectiveness in teaching.

Patton’s (2008) work suggests that UFE can allow participants to “engage” in the overall evaluative processes by infusing them into the organizational culture through training, and enhancing shared understandings of the evaluation itself through building processes that allow for the data to be tied to specific outcomes (e.g. teaching effectiveness, student’s self-report progress on learning through SETs). Another directly applicable aspect of UFE is Patton’s (2008) call for evaluations to increase engagement through empowering those involved through iterative reflective practice and self evaluation, which could be a foundation for inquiring as to how SETs may or may not be better utilized formatively as well as the extent to which they should be used for summative evaluations of teaching. A summary of the UFE framework adapted from Patton (2008) is illustrated in Figure 1 below:

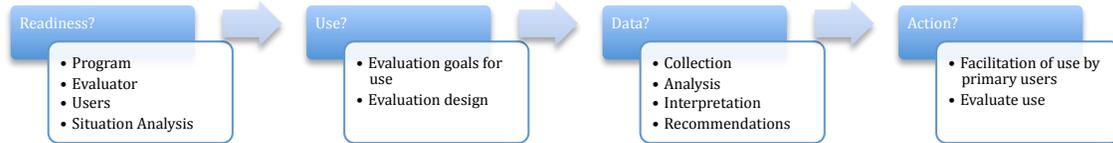


Figure 1: Author's Summary of Patton's (2008) UFE Framework

The UFE framework's concept of engaging evaluations framed the inquiry, design of the research questions, and the development, implementation, and analysis of the survey instrument. The primary purpose behind this study was to share the results with the faculty and administrators at the study site by gauging readiness for further inquiry and possible action concerning SETs and their perceived utility by those most deeply affected by them; the faculty. This study was designed to not only be an act of inquiry, but also to inform and initiate action through the UFE framework in further evaluating SETs, which is beyond the scope of this paper but is part of an ongoing UFE process at the study site.

Instrument and Data Collection Design

Nardi's (2005) work suggests that each method for collecting data has its own advantages and disadvantages, but that surveys are "ideal for asking about opinions and attitudes" and "better for sensitive and personal topics" (pp. 18-19). Sechrest and Sidani (1995) describe inquiry whereby a questionnaire and anecdotal evidence (e.g. "conversations") can provide what they refer to as "formulaic" (quantitative) and "clinical" (qualitative) approaches, respectively (pp. 82-84). Patton (2008) contends that, while both qualitative and quantitative data are beneficial to UFE, "in many cases, both

qualitative and quantitative methods should be used together” and “that there are no logical reasons why both kinds of data cannot be used together” (p. 438).

Therefore the survey was designed to capture both quantitative and qualitative data through scaled questions for descriptive statistical analysis as well as open-ended questions allowing respondents to clarify and describe their own experiences and opinions concerning SETs. Patton (2008) describes the merits of a mixed approach by stating that “qualitative inquiry involves small ‘purposeful samples’ of information-rich cases” that provide “rich” data that is “particularly useful” (p. 459). He also posits that while generalizability of both qualitative and quantitative findings within a UFE framework need to be taken into consideration because program evaluations (e.g. perceptions and reported use of SETs) are certainly context specific, he argues that “extrapolation” as Cronbach and others in the Stanford Evaluation Consortium (1980) define it is a more appropriate expectation as they are “modest speculations on the likely applicability of findings to other situations under similar, but not identical, conditions” (Patton, 2008, p. 459).

An Internet survey was selected for several reasons; confidentiality, anonymity, and convenience for respondents, and attempting to garner as much qualitative and quantitative data within my case as possible. Given the potentially sensitive nature of faculty self-reporting their perceptions and use of SETs, the fact that SETs are utilized by administration for formative and summative purposes, and my role as a faculty-administrator within my division, I felt a strong need for the faculty to trust the claimed intention and use of my survey and tried to convey that in my survey invitations and instructions. Lastly, site access was obtained through permission from the Provost and

Office of Academic Affairs at my institution, and I ensured that my research project and survey instrument were approved by IRBs at both my host university (as part of my doctoral research) and the research site institution. These approvals were conveyed to prospective respondents.

The majority of questions were developed as closed-ended with ordinal scales to measure gradations of “opinions, attitudes, behaviors, and attributes” (Dillman, Smyth, & Christian, 2009, p. 135-136), with response categories limited to no more than five to allow a realistic “continuum” (p. 137) in the respondents minds without overburdening them and thus potentially negatively impacting their response or participation in each question, and the scales were also constructed with an intent to provide category balance. Additionally, several of the closed-ended ordinal scale questions allow respondents an open-ended text box and thus opportunity to clarify or make additional comments should they wish to do so (as opposed to requiring such a response or the instrument containing more open-ended questions). I surveyed the entire population of full- and part-time faculty within my division precisely because I wish to utilize the findings as an initial part of implementing a UFE process related to SETs within my division. As the survey population of all full- and part-time faculty is relatively small (n=30), the survey attempted to avoid sampling error by collecting data from all of the faculty within the division.

Participatory Action and Iterative Data Analysis

As Glesne (2006) describes, data analysis at its essence involves the synthesis and interpretation of data in order to make sense of what might be learned through descriptions, explanations, and positing of hypotheses (p. 147). Similar to Miles &

Huberman's (1994) method of data reduction, display, and conclusion drawing, Patton's (2008) four-stage model first calls for a description and analysis of both qualitative and quantitative data into a format revealing patterns specifically so that intended users can readily understand and interpret the results (pp. 478-481). Patton's (2008) second step involves "interpreting" the findings by "going beyond the data" to add "context, determine meaning, and tease out substantive significance" (p. 478). Patton's (2008) four-stage model of engaging evidence (analysis, interpretation, judgment, and recommendations) was utilized by the primary evaluator (the author) for initial analysis, to be subsequently shared with the primary stakeholders (faculty respondents) as part of an iterative UFE process.

An essential part of Patton's UFE framework for engaging evidence as discussed above relates to evaluator (researcher) bias, and thus I utilized an impartial peer debriefer who as Creswell (2007) contended played an invaluable role as a "devil's advocate" (p. 208) in checking not only my interpretations but also in my biases (they did not have access to the raw data). In full disclosure of my bias, I tend to agree with the "protagonists" mentioned in the literature review section of this study who argue that SETs and garnering students' perceptions of their own learning is an important part of measuring student learning and teaching effectiveness, and can be very beneficial formatively in perpetuating or modifying pedagogy in the classroom (I personally utilize SETs in my own reflections of my teaching practices). However, I also agree with some of the skeptics' and protagonists' cautions that SETs not be used as solitary measurements of student learning nor teaching effectiveness, and should not be the sole basis for summative evaluations.

Limitations

Some of the primary limitations of this study are related both to the site institution and its use in this study as the single case. The site is a private four-year, career-oriented college the northeastern United States. Moreover, the institution is focused primarily on a practitioner-scholar model placing much higher expectations for faculty on teaching than research, with a normal teaching load of eight courses per academic year. The institution does not offer tenure, but rather, one-, three-, and five-year rank appointments. Additionally, the faculty surveyed teach primarily in business and business related disciplines. Another limitation is the very nature of respondents' self-reporting. Despite all efforts in transmitting anonymity and confidentiality to all potential respondents, it is possible that part-time faculty may feel more vulnerable to SET ratings and their summative use, although the site institution does not offer tenure to full-time faculty. Primary delimitations of this study are restricting the survey to 30 full- and part-time undergraduate faculty within my current division at my institution to focus on my division as a single case study, and restricting my survey questions and invited population to undergraduate teaching.

Yet these limitations also make this study's survey population very relevant to further inquiry concerning SETs precisely because of the heavy emphasis on, and culture of, teaching. The data captured within this study are integral to not only my gaining a deeper understanding of faculty perceptions towards SETs, but moreover, are critical foundational steps towards the development of an on-going evaluation process based on the UFE framework. Despite valid concerns of overgeneralization from inferring conclusions based on a small sample, this study will still provide invaluable information

for me, my division, and perhaps my institution and its practices involving SETs. This study and its findings may also provide some added value to the research and practice related to SETs at other institutions, which others may find analogous or beneficial to their contexts.

Faculty Perceptions of SETs

Generally three central themes emerged from an analysis of the survey. Overall, faculty were generally very supportive of SETs for formative assessments, and strongly reported their importance and use for evaluating their own pedagogy. However, they also reported relying primarily upon the students' written comments rather than answers to scaled-questions pre-determined upon criteria from the externally-generated SETs and ensuring diagnostic reports, and indicated a desire for a different process providing more institutional support in interpreting and incorporating all feedback. Faculty also reported the importance of SETs as part of their own summative evaluations, yet also expressed concern about overreliance upon them and indicated a desire for a different process. These findings, discussed below, indicate that more of an action-oriented, inclusive evaluation of the overall SET process within my division is needed, and that Patton's UFE methodology is a viable framework. Before the findings are presented a background on the SET instrument currently utilized will be provided as well as demographic data on the respondents.

Background on SET Instrument Used at Research Site

The SET instrument currently used at the study site is the IDEA Student Ratings of Instruction. IDEA states more than 365 institutions use their services through more

than 5 million surveys across 250,000 classes each year and claims, “No other nationally available student ratings instrument provides more evidence of validity and reliability” (IDEA Center). The IDEA Center self-describes its Student Ratings of Instruction as different from most other SETs in factoring out “extraneous circumstances,” focusing on “student learning of 12 specific objectives” and its design for “soliciting students’ feedback on their own learning progress, effort, and motivation, as well as their perceptions of the instructor’s use of 20 instructional strategies and teaching methods. In addition, the system surveys instructors regarding their overall goals and highlights for them in the analysis and report” (IDEA Center). Faculty and administrative receive a “Diagnostic Report” that includes a comparison of student perceptions of the course’s objectives, as determined by instructing faculty through their selection of which pre-determined IDEA “Relevant Learning Objectives” are “Essential” or “Important” to each course. The Report also includes unadjusted and adjusted averages, with the latter incorporating students’ responses to questions about their own interest and effort put forth in each course, and provides the opportunity for students to write feedback through prompted and open-ended comments.

Demographics of Respondents

Out of a total population of 30 possible respondents, 27 (90%) participated in the survey. While $n=30$ is too small to draw conclusions using inferential statistics, the high response rate does provide a confidence level of 90%, indicating participant responses are likely successful in capturing the population’s characteristic 90% of the time. The high response rate also perhaps indicates that SETs are an important topic for all faculty, and that efforts to solicit high response rates and convey confidentiality and anonymity were

likely successful. Of particular note is the 100% response rate of all full-time faculty (n=17) within the division, comprising 63% of the survey population. There was also a high response rate (69%) of part-time faculty (n=9) participating, with 1 respondent choosing not to identify their status.

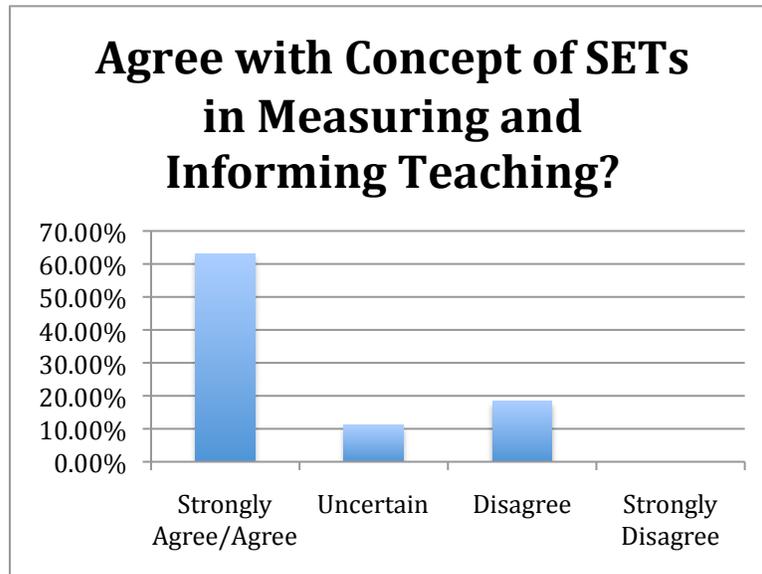
Of the 27 respondents, there was a very good distribution in terms of years of teaching undergraduate college courses; 22% indicated between 1-7 years of teaching experience, 41% with 8-14 years, and 37% with 15 or more years. This distribution indicates that an oversampling of one category was not present, and that faculty with a range of teaching experience provided subsequent answers. When asked to describe what level of undergraduate courses they primarily taught, 56% chose the category of “Mixture: 100-400 level courses,” meaning that more than half of respondents had experience in teaching students of varying class years and in classes of varying content breadth and depth.

Faculty Generally Support Concept of SETs

One major finding is that in general, the majority of faculty in this study self-reported both a perceived value in SETs as well as utilization of them to inform their own teaching practices. Several questions allowing for faculty perceptions related to both SETs in general and the IDEA tool in particular provided some very interesting responses as they related to faculty’s self-reported perceptions of SETs. As evidenced in Table 4.1 in response to the question *“In general, do you agree with the overall concept of student evaluations of teaching in measuring and informing teaching practices?”* 25 out of 26 respondents participated (1 non-response). Of particular note is that 65% of respondents

agree to some extent, while no respondents strongly disagreed and only 19% disagreed somewhat.

Table 4.1

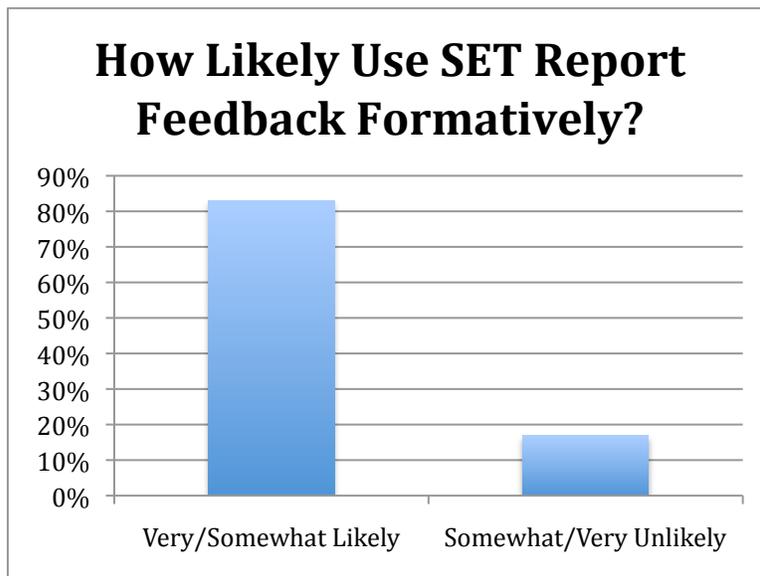


Concerning providing optional qualitative answers to this question, 58% of respondents also wrote comments, the majority of which were supportive of SETs in concept. While a few written responses indicated caution about students’ ability “to know what’s best for them” or that it “required a mature and motivated student to complete the evaluation with honesty and integrity,” the majority was supportive of SETs in general. One respondent wrote that they “Do find written comments and observations by students of value” and several others indicated value or potential value from SETs. Another commented, “In cases where the professor is really falling short of professional standards the evaluations can help inform and hopefully correct a situation.” One faculty member stated that while they agreed with SETs, “gathering written responses is more valuable to influencing how I teach than answers to standardized questions,” and that getting students to engage even further in written evaluations would “compel them to be more engaged in the process.”

This particular response also indicated that “having the criteria (both the faculty and the students) before being evaluated against it would be ideal.”

In addition to their perceptions of SETs in general, faculty were also asked to self-report their overall use of the IDEA Diagnostic Reports which accompany the feedback given to each faculty, and how likely “they are to use the report’s feedback to inform their own teaching practices.” Another significant finding from the survey as indicated in Table 4.2 is that 83% of respondents stated they were in fact likely (either “somewhat” or “very” likely) to use the reports (with 23 out of 27 responding) formatively.

Table 4.2



Faculty’s qualitative feedback mirrored their sentiments towards SETs in general, whereby the majority valued in student feedback through the IDEA tool, particularly in the students’ written comments, with several faculty reporting their extensive use of feedback. One faculty reported “I have made changes based on overall themes I see” while another reported “seeing far more value” and desire to use them even further. One response indicated their use but desire for more qualitative feedback, as they were “very likely to use this information because it’s the best tool we currently have, however, I do

believe it could be improved upon by specifically using written responses to less but targeted questions.” The results and sentiments from these two questions provide invaluable information conducive to initiating a UFE process. First, in indicating general faculty support of SETs and their suggestions towards providing more opportunity and incentive for students’ qualitative feedback, and secondly in their appeal for more instruction about the process of SETs itself for both faculty and students. These suggestions are examples of precisely what the UFE process of analysis, interpretation, judgment and action aims to facilitate.

Current SET Instrument Useful, Yet Indications a Better Process More Beneficial

Faculty responses to two additional questions gauging their perceptions further indicated the finding that the majority sees value in feedback from SETs, including through the IDEA tool’s pre-determined course learning objectives. As part of the IDEA Student Ratings of Instruction, faculty were asked to rate the level of importance towards predetermined teaching goals as course objectives (e.g. gaining factual information, learning to apply course material, etc.). Students then self-report their perception of the courses’ objectives, and the IDEA tool provides a comparison of these. When asked, *“how important is it to compare students’ perceptions of course objectives with your own determination of course objectives,”* 100% of respondents indicated placing some degree of importance on this comparison, with 78% respondents indicating a high-degree of importance placed (with 23 out of 27 responding). When faculty were then asked the extent to which they utilized this course-objective comparison *“to evaluate their own teaching practices,”* 65% indicated they utilized the information to some or a large extent

with only 1 respondent indicating they did not use the information at all (with 23 out of 27 responding). There was no opportunity for qualitative feedback on these questions.

The IDEA instrument also asks students to consider certain “Teaching Methods and Styles” used by instructors under a section entitled “Improving Teaching Effectiveness” with questions claiming to gauge the instructor’s ability to stimulate student interest, foster collaboration and involvement, establish rapport, and provide a structured classroom experience. One question in the survey asked faculty to self-report the extent to which they utilized this particular information in evaluating their own teaching practices. 61% of respondents reported utilizing this information to some or a large extent, 35% claiming they did not use the information much, and one respondent stating they did not use the information at all (with 23 out of 27 responding). In their qualitative responses, several faculty respondents reported using this information as “consideration” for future classes and that it was “helpful to make minor changes to the courses.” One faculty member stated they “really should utilize this information” and would in the future “be reviewing in much more detail” this section’s reports to determine “what they could learn from them.” Thus, overall, responses further indicated faculty consideration and use of SETs formatively, although to a lesser degree than other survey questions gauging value and use. This is perhaps related to the faculty’s concerns about students ability to evaluate effective teaching, awareness of their own learning progress within any given course, or possibly the IDEA criteria versus what faculty themselves may deem to be important measures of teaching effectiveness. One faculty member suggested review by third parties would be better measures of teaching

effectiveness, with another stating that more support for interpreting the IDEA results would be beneficial.

Faculty also indicated concern over the IDEA tool itself in their responses to another question which asked to what extent they agreed with IDEA's claim that its SET is rooted in a "Student Learning Model" which helps assess *"how specific teaching behaviors influence certain types of student progress under certain circumstances."* Of the 23 respondents, results were quite mixed with approximately 40% stating they somewhat agreed with IDEA's claim, 30% neither agreeing nor disagreeing, and 30% disagreeing. No respondents indicated agreeing strongly. The perceived benefits and concerns of this question could likely be uncovered during a divisional-level discussion as part of a UFE process, which could provide more support for the faculty seeking to utilize the information more while still considering faculty concerns.

Interestingly, faculty's perceived value in, and use of, SETs, was reported numerous times through the responses above despite mixed feelings as to students' ability to accurately measure their own learning. As to whether or not faculty agreed with students' ability to measure their own learning through SETs, 36% of faculty agreed to some extent, with 48% disagreeing somewhat (25 out of 27 responses to this question). Comments indicated that students are somewhat able to measure their own learning and "elements of their own progress," but that they likely are not "able to recognize what 'learning' can mean" especially vis-à-vis the SET objectives and criteria and likely "evaluate their progress via feedback from assignments and exams." Delving deeper into this perception, since it contrasts with the strongly reported sentiment that SETs are indeed valuable and used by the majority of faculty, further conversation through a UFE

process could shed further light on the sentiments behind, and impact of, this juxtaposition.

Faculty responses to these questions about both SETs in general and their self-reported use of information from the IDEA tool specifically indicate not only the value they see in SETs, but also a need for a more inclusive SET process involving their input. Faculty concerns about the instrument's pre-defined criteria, desire for more emphasis on opportunities for written student feedback, and call for a better process for assisting them in reflecting upon SET feedback are important points for further discussion in a UFE framework. Responses such as these generally support findings of SET proponents in the literature. In their meta-analysis of research on SETs, Penny and Coe (2004) concluded that SETs do have considerable utility formatively, especially with proper consultation of feedback including active involvement of instructors in the learning process and adequate time for reflection and dialogue. Moreover, Kember, et al. (2008) also emphasized the importance of instructor reflection on SET feedback, with their last, and most effective stage of "critical reflection" when the instructor not only relate the results to their experiences, but takes "transformative" action to incorporate necessary changes or augment successful practices.

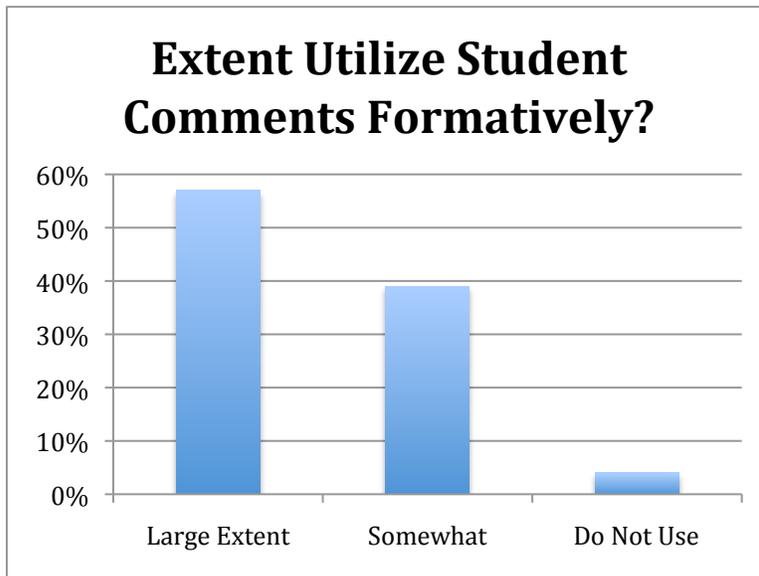
Changes to pedagogy are often seen as overwhelming or difficult. McGowan and Graham (2009) argue that "meaningful change" does not require complete overhauls of teaching practices, but rather, are achievable changes related to creating more "active" learning, increasing student-teacher interactions, establishing expectations and maintaining high standards, being properly prepared for class and examining grading practices and feedback given to students. Such suggestions indicate that faculty

appreciated the value of SET feedback but that perhaps a more inclusive process of assisting them in better utilizing such feedback for action might be conducive to higher levels of acceptance, reflection, and utilization in assessing their own teaching practices. A UFE process involving faculty suggestions for such active reflection and involvement of faculty would appear to potentially beneficial in this regard and could foster deeper conversations about faculty concerns over the IDEA tool itself and either possibly adjustments to it and/or possibly conversations about another tool.

“Qualitative Comments Are Most Beneficial”

While some faculty freely offered their perspective on the importance of qualitative feedback from students in their responses to the questions mentioned above, faculty were provided a specific opportunity to report to what extent they “*utilize students’ comments to evaluate and inform*” their own teaching practices in a separate question. Their responses also powerfully indicated a receptivity and use of student comments formatively. As noted in Table 4.3, of 23 respondents (four non-responses), 96% reported utilizing student comments, with 57% stating they utilized the information to a “large extent.”

Table 4.3



While a few faculty commented on needing to take students' comments "with a grain of salt," the vast majority of comments strongly indicated faculty utilization of student comments. One faculty comment strikingly reflects the quantitative data, "Whatever teaching style I have evolved to in close to three decades of teaching has been driven by comments on student evaluations more than anything else." Another faculty stated, "I find the qualitative comments students offer to be the most beneficial for me to either improve or maintain certain practices. This seems to be the best unfiltered feedback directly from students that clearly details what they valued and did not from the course." Another response affirmed, "Absolutely this is where the most tangible and actionable information is derived," while another stated, "The qualitative comments are the most useful for me as I consider how I did that past semester and how I hope to improve for the next." Other comments further indicated perceived importance and use of student feedback, but also concern that current structure of the IDEA tool as used in my division did not provide enough opportunity for student feedback; "The comments not only provide meaningful feedback, but seems to highlight the major perceptions of

the student which can sometimes get lost in the IDEA evals.” While yet another faculty member acknowledged “The problem is that the IDEA format reduces the volume of comments—far less than our previous format produced.”

The results and qualitative feedback from this specific question are informative, as it further suggests that faculty self-report they utilize student comments for formative purposes, and desire even more qualitative feedback from students. It thus further highlights the importance of initiating a larger conversation within my division about SETs in general and the IDEA tool in particular, as it is clear from the faculty responses that more targeted opportunities for students to provide qualitative feedback would not only be utilized, but are desired. Patton’s (2008) UFE methodology argues for evaluations to focus on the utility of the information for action and decision-making. The results of the survey strongly indicate a readiness for foundational groundwork within my division in attempting to develop a process (program) that explicitly connects SET and IDEA data and analysis with enhancing achievement through feedback and use of the data to inform decisions and increase understanding of SETs and faculty’s perceptions of them. Patton (2008) also describes a primary intention for his UFE model as “building capacity” of those involved (p. 157), which directly applies to faculty and students. This could be incredibly powerful in developing more inclusive models of SETs and their formative use by faculty, beginning with a possible collective determination to solicit more opportunity for student qualitative feedback through my divisions SET process, as well as including a better framework for understanding the IDEA report information. Patton’s (2008) UFE framework would definitely provide a mechanism for learning more

about faculty perceptions of SETs precisely by including survey results with them through a process of collective analysis and recommendations.

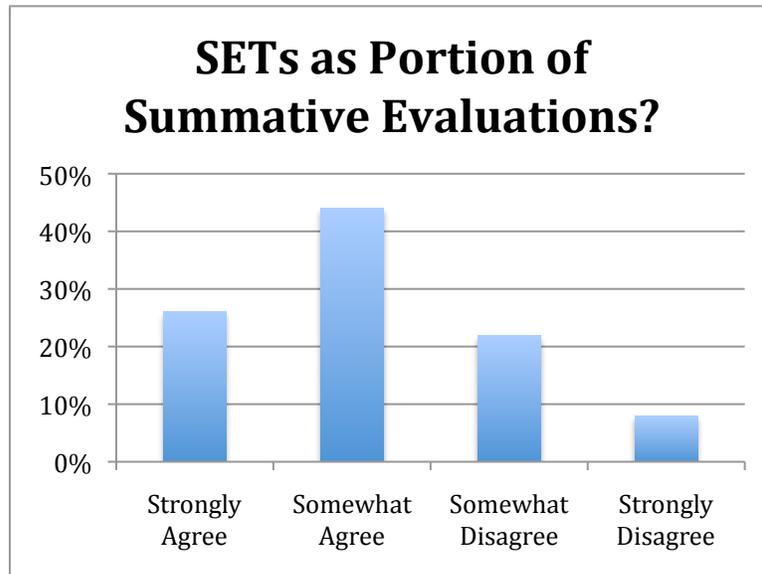
Both the quantitative and qualitative data strongly indicate Patton's (2008) UFE framework would be conducive for faculty recommendations made in light of "improvement-oriented, formative evaluations" concerning SETs. Moreover, collectively exploring SETs' "strengths and weaknesses with the expectation that both will be found and each can be used to inform an ongoing cycle of reflection and innovation" (Patton, 2008, p.116) has the potential to benefit faculty within my division. Faculty do indeed self-report utilizing SETs, and perceiving their merit, in helping to reflect upon and perhaps improve their own teaching practices, and their responses indicate a need for process improvement based on Patton's UFE.

Valuing SETs, But Wanting "A Better Way"

While the primary focus of this inquiry was designed to gauge self-reported faculty perceptions of SETs for formative use, two questions in the survey provided opportunity for faculty to report their perceptions of SETs for summative use. The findings from these questions are also quite informative in indicating faculty desire for a more inclusive evaluation of the current SET process within my division through a UFE framework. In one question, faculty were asked "*In general, do you agree with the practice of including student evaluations as a portion of your overall teaching evaluations?*" This question explicitly stated to consider their answer within the overall context of SETs used for summative purposes and not the IDEA tool specifically. As indicated in Table 4.4, 70% of respondents indicated they "agreed" to some extent with

SETs being used as part of their summative evaluations, while 8% indicated they strongly disagreed (with 23 out of 27 responded to this question).

Table 4.4

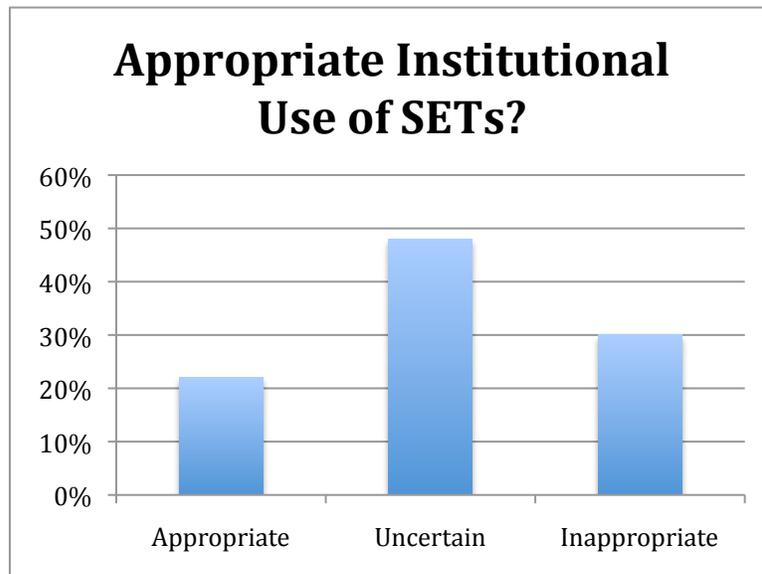


One faculty member responded, “A good student evaluation process is necessary to see how well the material was taught,” while another commented, “Yes I agree. We are there to serve the students and their feedback is essential.” While still supportive of SETs as a portion of summative evaluations, another faculty member stated “All feedback is important, but must be kept in perspective of the class dynamic and other evaluation tools,” while another reported “I see it as being a portion but not the entire basis of my teaching evaluation.” One faculty member commented that “While I understand that at an administrative level there needs to be some type of way to assess teaching and learning...there has to be a better way” and expressed concern about their scores on SETs being directly “tied to their pay.”

While faculty were quite supportive of SETs as a portion of their summative evaluations, the concept of a “better way” came to light in response to the second question, “*In your opinion, how appropriately does the institution use the IDEA Student*

Ratings of Instruction and Summary Reports?” With 23 out of 27 participants responding, the results were quite mixed; as noted in Table 4.5, 22% felt they were used appropriately to some extent, while 30% felt they were used inappropriately to some extent and 48% indicated they were uncertain.

Table 4.5



The qualitative comments mirrored these results, with affirmative comments such as “I think the Administration and College in general are very realistic about the use of student ratings” to “I believe my institution uses them appropriately. The information is used as a development and feedback tool.” Yet faculty also stated being uncertain of the process, “I really am unsure of how they are used and whether they are used properly,” as well as being concerned as to a lack of an overall “pattern” and that each “Dean looks at them differently.” One faculty was explicit in their stated concern that too much emphasis was placed on the IDEA results in summative reviews and suggested that a new process was needed to enhance their formative benefit;

I believe our institution uses them very inappropriately as they seem solely used for faculty performance reviews (with very little other data that goes into that...no peer evaluations, etc.). We also do no systematic processing of the results,

leaving me as an individual faculty to determine how I can improve in the areas suggested by the evaluation. It would be helpful to have specific resources/coaching available on areas where improvements are needed.

Thus, while faculty self-report being supportive of SETs as a portion of summative evaluations, it is clear that their uncertain or perceived overuse for summative evaluations combined with a general sentiment of underutilization institutionally formatively, needs to be part of a larger conversation within an extension of a UFE process. Such a process would likely help capitalize on general faculty sentiment that SETs should indeed be a part of summative evaluations, yet address and uncover their differing sentiments concerning current use of the IDEA and provide more opportunity for their suggestions as to better processes. Indeed, the last faculty comment summarizes an important conversation through the UFE process of analysis, interpretation, judgment, and recommendation concerning SETs and summative evaluations within my division. There is general consent among the faculty as to the importance of SETs inclusion for summative evaluations, yet their apparent concern as current processes implies feeling excluded. Thus, there may be a very rich opportunity for administration to capitalize on faculty support of SETs by opening up such a conversation that also allows dialogue about their specific concerns for the current process and/or the IDEA tool itself.

Discussion

The findings from faculty suggest the utility of the development of an inclusive evaluation of the formative and summative components of SETs practice using Patton's (2008) UFE process. As Patton (2002) describes, a primary tenet of his UFE framework is for intended users of evaluation processes to be actively engaged in the evaluation process itself;

A psychology of use undergirds and informs utilization-focused evaluation: intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they've been actively involved; by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way (p. 1).

Based on the analysis of faculty perceptions, it is clear that an evaluation of current SET practice, the IDEA tool, and the formative and summative use of them has the potential to be tremendously beneficial for all stakeholders as faculty are generally supportive of SETs in concept for both formative and summative purposes, yet expressed concerns as to implementation and current practices concerning both uses. Faculty are supportive of SETs but indicate a desire for more ownership, which would likely lead to more robust and informative use of them.

In one of the few studies drawing primarily upon faculty perceptions of SETs, Simpson & Siguaw (2000) report that “Overall, faculty appeared to believe that SETs encourage instructors to lower educational standards, serve as a tool for student revenge, encourage overreliance on ratings in performance evaluations, and are rife with measurement issues” (p. 209). Their study also indicated that given these perceptions, faculty responses to SETs could be categorized as either ignoring them, consciously or unconsciously adopting lower course and grading standards to appease students, or actually implementing teaching practices aimed at positively improving ratings rather than for increases in student learning. The findings within this study indicate, however, that there is not necessarily such widespread resistance nor dismissal of SETs for formative use. On the contrary, faculty respondents clearly reported perceived value and self-report utilizing SETs formatively, and as one faculty reported, “It needs to be part of

the discussion.” Yet, according to the faculty in my division, there needs to be more training for faculty as users of SETs to “reinforce the intended utility” of them formatively.

Patton’s (2008) UFE is conducive for better including teaching faculty as primary stakeholders through a combination of “designing instruments...overseeing implementation, and interpreting findings” (p. 77). One fundamental purpose of evaluations as Patton (2008) defines them can be for “formative improvement and learning,” which allows administrators, staff, and participants who are regularly involved in the program to all be the primary users (p.139). This model would potentially make UFE—through a lens of formative improvement learning—a very viable framework for interpreting and incorporating the results of faculty perceptions of SETs into a more inclusive model; a framework through which all stakeholders become more involved through making adjustments, acting upon feedback, and enhancing implementation and outcomes. Incorporating aspects of Patton’s UFE through this lens asks stakeholders to examine what is and isn’t working as well, discover strengths and weaknesses, and to inquire more deeply into participant (student) reactions—to discern if certain patterns emerge among different groups (e.g. classes or majors)—in an attempt to increase outcomes (e.g. teaching effectiveness and student learning). Participant feedback, reflective practice and appreciative inquiry are common approaches in this particular practice of UFE, and Patton describes the key factors in using UFE for formative learning and improvement as “creating a learning climate, openness to feedback and change” and “trust” that lead to use of the findings in “actionable” steps (p. 140). The data from this study indicate that faculty might indeed be open to beginning a process that leads to more

actionable steps addressing their perceived benefits and concerns of SETs' formative utility.

There also clearly needs to be a conversation within my division about faculty's support of SETs in summative evaluations but within the context of their very real concerns as to how IDEA evaluations are currently being used for such purposes. While the majority of faculty report SETs as a necessary component of their summative evaluations, they report being concerned about their overuse in this regard. As one faculty member reported, "There needs to be a more consistent and constructive way to determine course evaluations." Beran, Violato, & Kline (2007) stress the importance of multiple sources of information in evaluating teaching effectiveness, and Benton & Cashin (2012) are very clear in their argument that "writers on faculty evaluation are almost universal in recommending the use of multiple sources of data" (p. 1) and that SETs alone do not provide "sufficient information to make a valid judgment about an instructor's overall teaching effectiveness" (pp. 1-2).

The UFE framework would also be ideal in bringing these faculty perceptions to light within the overall summative evaluation process, again as part of Patton's (2008) principle of engaging evaluations through shared understandings of the evaluation process itself (p. 158). It is also important to include faculty perceptions of SETs as "summative, judgment-oriented evaluations" that help determine the "overall merit, worth, significance, or value" of SETs (Patton, 2008, p. 113) by including the faculty as primary stakeholders in a process that profoundly impacts them; SETs as a tool used in summative evaluations within my division. Faculty responses to the survey questions related to their perceptions of SETs as summative evaluations could be instructive for

next steps for my division to consider, and perhaps would lead to a process (and culture) change designed to shift more of the focus—and administrative use—of SETs and the IDEA evaluation towards “formative improvement and learning” as Patton describes it above. As faculty report seeing merit in SETs but believe they are overused as summative measurements, perhaps this information can help administrators and my institution better discern how they’d like to use SETs. Does the institution wish to use SETs primarily for summative purposes, or might there be a receptive opportunity to assist faculty in utilizing them more formatively (and if so, how might the institution decouple/discharge current summative use from the formative use SETs provide)?

My division and institution have a very real opportunity to begin engagement in a collective evaluation of SETs and the IDEA instrument, beginning with an evaluation of faculty’s perceived formative and summative utility of them from this study. Patton’s UFE process calls for the initial steps of determining readiness of participants to begin an evaluation, identification of primary intended users, and identification of primary intended uses of an evaluation (Patton, 2002). Based off feedback from this study, the faculty within my division appear more than ready to begin exploring the prospect of evaluating SET use and the IDEA instrument, clearly identify themselves as primary stakeholders (intended users), and express a desire to evaluate current uses of them for both formative and summative purposes. Given their reported investment in SETs and to some degree the IDEA instrument, faculty also appear quite willing to play a role in the UFE framework’s subsequent steps of actively determining the priorities and design of further evaluating current practices and envisioning better processes. By doing so, faculty would become larger stakeholders in the intended use of SETs as well as an

increased role as intended users of them, and would likely feel more ownership of a collectively redesigned SET process, and thus likely individually utilize findings even more.

Conclusion

In sum, demands for accountability of student learning and teaching effectiveness in higher education are intensifying from multiple stakeholders. As a result, institutions are using SETs more than ever as an answer (or “the” answer) for demands of accountability of student learning and teaching effectiveness, yet perhaps without a deeper understanding of faculty’s self-reported perceptions and practices concerning SETs. This study is consistent within its findings that increased concerns or resentment towards SETs from faculty, especially when they are factored into summative evaluations of teaching, might in turn be overshadowing any potential formative utility that may be perceived. The current and potential uses of SETs as one measurement of student learning and of teaching effectiveness remain complex issues. This study suggests that UFE is one potential way to explore and perhaps effective change concerning these issues related to SETs in formative use and summative evaluations in future research and evaluations. Use of SETs is likely only to increase, so including faculty perspectives seems of great importance to not only broaden and enrich the discussion surrounding SETs, but for how their results may—or may not—be best utilized. Additional feedback from faculty concerning their perspective and self-reported use of SETs, as part of an inclusive and iterative UFE process, could lead to some very beneficial conversations at any divisional or institutional level.

Journal Article References

- Abrami, P. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *Directions for Institutional Research* , 109, 59-87.
- Aleamoni, L. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education* , 1, 111-119.
- Arum, R., & Roska, J. (2010). *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.
- Astin, A. W. (1977). *Four Critical Years: Effects of College on Beliefs, Attitudes, and Knowledge*. San Francisco: Jossey-Bass.
- Astin, A., & Antonio, A. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Rowman & Littlefield.
- Benton, S., & Cashin, W. (2012). *Student Ratings of Teaching: A Summary of Research and Literature*. Manhattan, KS: The IDEA Center.
- Beran, T. V. (2007). What's the 'Use' of Student Ratings of Instruction for Administrators? One University's Experience. *Canadian Journal of Higher Education* , 17 (1), 27-43.
- Berk, R. (2006). *Thirteen Strategies to Measure College Teaching Effectiveness*. Sterling, VA: Stylus.
- Berk, R. (2013). Top Five Flashpoints in the Assessment of Teaching Effectiveness. *Medical Teacher* , 35, 15-26.
- Cashin, W. (1988). *Student Ratings of Teaching: A Summary of the Research*. Kansas State University, Center for Faculty Evaluation and Development. Manhattan, KS: KSU.
- Centra, J. (1993). *Reflective Faculty Evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L. &. (1980). *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method 3/E*. Hoboken, NJ: John Wiley & Sons.
- Doubleday, J. (2013, December 2). With an Eye Toward Retention, Colleges Amp Up Advising. *The Chronicle of Higher Education* , pp. 1-6.
- Ewell, P. T. (2002). An Emerging Scholarship: A Brief History of Assessment. In T. W. Banta, *Building a Scholarship of Assessment* (pp. 3-25). San Francisco: Jossey-Bass.

Glasne, C. (2006). *Becoming Qualitative Researchers: An Introduction* (3rd ed.). Boston: Pearson.

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.

Greenwald, A. (1997). Validity Concerns and usefulness of student ratings of instruction. *American Psychologist* , 52 (11), 1182-1186.

IDEA Center. (1999, 1 1). *The IDEA Center*. Retrieved 7 10, 2014, from The IDEA Center: www.theideacenter.org

Kember, D., McKay, J., Sinclair, K., & Wong, F. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education* , 33 (4), 363-379.

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What Matters to Student Success: A Review of the Literature*. Washington, DC: National Postsecondary Education Cooperative.

Marsh, H. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology* , 52, 77-95.

Marsh, H., & Roche, L. (1993). The Use of Students' Evaluations and Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal* , 30 (1), 217-251.

Martell, K., & Calderon, T. G. (2005). Assessment in Business Schools: What It Is, Where We Are, and Where We Need to Go Now. In K. Martell, & T. G. (eds.), *Assessment of Student Learning in Business Schools: Best Practices Each Step of the Way* (Vol. 1, pp. 1-26). Tallahassee, FL: The Association for Institutional Research and AACSB International.

McCarthy, M. (2012). Using Student Feedback as One Measure of Faculty Teaching Effectiveness. In M. Kite, *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*. Society for the Teaching of Psychology.

McGowan, W., & Graham, C. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education* , 34, 161-171.

McKeachie, W. (1997). Student Ratings: The Validity of Use. *American Psychologist* , 52 (11), 1218-1225.

Miles, M. & Huberman, A. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: SAGE Publications.

- Nardi, P. M. (2005). *Doing Survey Research, 2/E*. Boston: Ally & Bacon.
- National Center for Public Policy and Higher Education. (2008). *Measuring Up 2008*. Washington, D.C.: NCPPHE.
- National Commission on Accountability in Higher Education. (2005). *Accountability for Better Results: A National Imperative for Higher Education*. State Higher Education Executive Officers and the Ford Foundation.
- Patton, M. Q. (2002). *Utilization-Focused Evaluation Checklist*. Available at www.betterevaluation.org/resource/guide/UFE_checklist
- Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Penny, A., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: Meta-analysis. *Review of Educational Research* , 74, 215-253.
- Richardson, J. (2005). Instruments for Obtaining Student Feedback: A Review of the Literature. *Assessment and Evaluation in Higher Education* , 30 (4), 387-415.
- Schmelkin, L. P., Spencer, K., & Gellman, E. (1997). Faculty Perspectives on Course and Teacher Evaluations. *Research in Higher Education* , 38 (5), 575-592.
- Sechrest, L., & Sidani, S. (1995). "Quantitative and Qualitative Methods: Is There an Alternative? *Evaluation and Program Planning* , 18 (1), 77-87.
- Seldin, P. (2006). *Evaluating Faculty Performance*. Bolton, MA: Anker.
- Seldin, P. (1993, July 21). The Use and Abuse of Student Ratings of Professors. *The Chronicle of Higher Education* , p. A40.
- Simpson, P., & Siguaw, J. (2000). Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. *Journal of Marketing Education* , 22 (3), 199-213.
- Smith, B. (2007). Student Ratings of Teaching Effectiveness: An Analysis of End-of-Course Faculty Evaluations. *College Student Journal* , 41 (4), 788-800.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research* , 109, 45-56.
- U.S. Department of Education. (2011). *Meeting the Nation's 2020 Goal: State Targets for Increasing the Number and Percentage of College Graduates with Degrees*. Washington, D.C.: U.S. Department of Education.

Wilson, R. (1998, January 16). New Research Casts Doubt on Value of Student Evaluations of Professor. *The Chronicle* .

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education* , 12 (1), 55-76.

REFERENCES

- Abrami, P. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *Directions for Institutional Research* , 109, 59-87.
- Accreditation Council for Business Schools and Programs. (1998). *ACBSP Standards and Criteria for Demonstrating Excellence in Baccalaureate/Graduate Degree Schools and Programs*. Overland Park: ACBSP.
- ACT. (2010). *2010 Retention/Completion Summary*. Iowa City: ACT.
- Adair, V. (2001). Poverty and the (Broken) Promise of Higher Education. *Harvard Educational Review* 71:2 , 217-239.
- Aleamoni, L. (1999). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education*, 13, 153-166.
- Algozzine, B. (2004). Student Evaluation of College Teaching: A practice in search of principles. *College Teaching* , 52 (4), 134-141.
- American Association for Higher Education and Accreditation. (1996). *AAHE Assessment Forum: 9 Principles of Good Practice for Assessing Student Learning*. Washington, DC: AAHEA.
- Angelo, T. A. (2002). Engaging and Supporting Faculty in the Scholarship of Assessment: Guidelines from Research and Best Practice. In T. W. Banta, *Building a Scholarship of Assessment* (pp. 185-200). San Francisco: Jossey-Bass.
- Arreola, R. (2007). *Developing a Comprehensive Faculty Evaluation System* (3rd ed.). Bolton, MA: Anker.
- Arum, R., & Roska, J. (2010). *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.
- Astin, A. W. (1977). *Four Critical Years: Effects of College on Beliefs, Attitudes, and Knowledge*. San Francisco: Jossey-Bass.
- Astin, A., & Antonio, A. (2012). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education* (2nd Edition ed.). Plymouth, UK: Rowman & Littlefield.
- Athvale, M., Davis, R., & Myring, M. (2008). "The Integrated Business Curriculum: An Examination of Perceptions and Practices". *Journal of Education for Business* , 83 (5), 295-301.

Baker, T. L., & Velez, W. (1996). Access to and Opportunity in Postsecondary Education in the United States: A Review. *Sociology of Education* , 82-101.

Balam, E. M. & Shannon, D.M. (2010) Student Ratings of College Teaching: A Comparison of Faculty and Their Students. *Assessment & Evaluation in Higher Education*, 10 (2), 172-184.

Banta, T. W. (2002). *Building a Scholarship of Assessment*. San Francisco: Jossey-Bass.
Bennis, W. (2004). The Crucible of Authentic Leadership. In J. Antonakis, A. Cianciolo, & R. Sternberg, *The Nature of Leadership* (pp. 331-342). Thousand Oaks, CA: Sage Publications.

Beran, T., Violato, C., & Kline, D. (2007). What's the 'use' of Student Ratings of Instruction for Administrators? One University's Experience. *Canadian Journal of Higher Education*, 17 (1), 27-43.

Benton, S., & Cashin, W. (2012). *Student Ratings of Teaching: A Summary of Research and Literature*. Manhattan, KS: The IDEA Center.

Berk, R. (2006). *Thirteen Strategies to Measure College Teaching Effectiveness*. Sterling, VA: Stylus.

Berk, R. (2013). Top Five Flashpoints in the Assessment of Teaching Effectiveness. *Medical Teacher* , 35, 15-26.

Birkland, T. A. (2011). *An Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making* (3rd ed.). Armonk, NY: M.E. Sharpe.

Bishop, T., Vaughn, T., Jensen, G., Hanna, N., & Graf, D. (1998). A Cross-Functionally Integrated Undergraduate Business Core Curriculum. *Mid-American Journal of Business*, 13 (1), 65-71.

Bloomberg, L., & Volpe, M. (2012). *Completing Your Qualitative Dissertation*. Thousand Oaks, CA: SAGE Publications.

Bowles, S. (1971). Unequal Education and the Reproduction of the Social Division of Labor. *Review of Radical Political Economics* , 3.

Burke, W. W. (2008). *Organizational Change: Theory and Practice* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Campbell, N., Heriot, K., & Finney, R. (2006). In Defense of Silos: An Argument Against the Integrative Undergraduate Business Curriculum. *Journal of Management Education* , 30 (2), 316-332.

Cashin, W. (1990). *Student Ratings of Teaching: A Summary of the Research*. Kansas State University, Center for Faculty Evaluation and Development. Manhattan, KS: KSU.

Centra, J. (1993). *Reflective Faculty Evaluation*. San Francisco: Jossey-Bass.

Centre for the Study of Higher Education. (2002). *Core Principles of Effective Assessment*. Melbourne: Centre for the Study of Higher Education.

College Board. (2009). *Trends in College Pricing 2009*. Retrieved 2013 28-November from The College Board: http://www.trends-collegeboard.com/college_pricing/index.html

Colman, J. (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Creswell, J. W. (2007). *Qualitative Inquiry and Research Design: Choosing Among the Five Approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Cronbach, L.J. & Associates (1980) *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass.

DeBerard, M. S., Spielmans, G., & Julka, D. L. (2004 March). Predictors of Academic Achievement and Retention Among College Freshman; A Longitudinal Study. *College Student Journal* , 1-12.

DeVillis, R. F. (2003). *Scale Development: Theory and Applications*. Thousand Oaks, CA: SAGE Publications.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method 3/E*. Hoboken, NJ: John Wiley & Sons.

Donahue, P. (2000). Evaluating Teaching. *ADE Bulletin* , pp. 46-47.

Donaldson, S., Gooler, L., & Scriven, M. (2002). "Strategies for Managing Evaluation Anxiety: Toward a Psychology of Program Evaluation". *American Journal of Evaluation*, 23 (3), 261-73.

Doubleday, J. (2013 2-December). With an Eye Toward Retention, Colleges Amp Up Advising. *The Chronicle of Higher Education*, pp. 1-6.

Ewell, P. T. (2002). An Emerging Scholarship: A Brief History of Assessment. In T. W. Banta, *Building a Scholarship of Assessment* (pp. 3-25). San Francisco: Jossey-Bass.

Feinberg, W., & Soltis, J. (1985). *School and Society*. NY: Teachers College Press.

Freyd, M. (1923). The Graphic Rating Scale. *Journal of Educational Psychology* , 14, 83-102.

Fryer, B. (2009 29-January). Is College Working? *Harvard Business Review* , pp. 1-2.

Fullan, M. (1999). *Change Forces: The Sequel*. Oxon, UK: RouledgeFalmer.

Genc, I. H., & Bekmez, S. (2008). Designing an Integrated Business Curriculum with Students' Success In Mind. *International Journal of Management Education* , 7 (3), 81-86.

Glesne, C. (2006). *Becoming Qualitative Researchers: An Introduction* (3rd ed.). Boston: Pearson.

Goldschmid, M. (1978). The evaluation and improvement of teaching in higher education. *Higher Education*, 7 (2), 221-245.

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student Course Evaluations: Research, Models and Trends*. Toronto: Higher Education Quality Council of Ontario.

Greenwald, A. (1997). Validity Concerns and usefulness of student ratings of instruction. *American Psychologist* , 52 (11), 1182-1186.

Heckert, T., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal* , 40 (3), 588-596.

Hutchins, P. (2010). *Opening Doors to Faculty Involvement in Assessment*. National Institute for Learning Outcomes Assessment. Champaign: NILOA.

IDEA Center. (1999 1-1). *The IDEA Center*. Retrieved 2010 йил 1-12 from The IDEA Center: www.theideacenter.org

Ishitani, T. (2006). Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States. *The Journal of Higher Education* , 861-885.

Jares, T. E., & Alexander, J. F. (2005). Pursuing A Quality-Based Strategy: A Case Study of the First Business School to Earn the Baldrige Award. In K. Martell, & T. Calderon, *Assessment of Student Learning in Business Schools: Best Practices Each Step of the Way* (pp. 66-81). Tallahassee: Association for Institutional Research.

Kelderman, E. (2009 22-January). "Risk Managers Extend Their Turf to Every Corner of the Campus." *The Chronicle of Higher Education* .

Kember, D., McKay, J., Sinclair, K., & Wong, F. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education* , 33 (4), 363-379.

Keown, A., Martin, J. D., & Petty, J. (2007). *Foundations of Finance: The Logic and Practice of Financial Management*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Kite, M. E. (2012) (Ed.). Effective evaluation of teaching: A guide for faculty and administrators. Available on the Society for the Teaching of Psychology web site: <http://teachpsych.org/ebooks/evals2012/index.php>

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What Matters to Student Success: A Review of the Literature*. Washington, DC: National Postsecondary Education Cooperative.

Kuh, G. (1993). In Their Own Words: What Students Learn Outside the Classroom. *American Educational Research Journal* , 30 (2), 277-304.

Lipman-Blumen, J. (2000). *Connective Leadership: Managing in a Changing World*. New York: Oxford Press.

Mallinckrodt, W., & Sedlacek, B. (1987). Student Retention and the Use of Campus Facilities. *NASPA* , 24, 28-32.

Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77-95.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.

Marsh, H., & Roche, L. (1993). The Use of Students' Evaluations and Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal* , 30 (1), 217-251.

Martell, K., & Calderon, T. G. (2005). Assessment in Business Schools: What It Is, Where We Are, and Where We Need to Go Now. In K. Martell, & T. G. (eds.), *Assessment of Student Learning in Business Schools: Best Practices Each Step of the Way* (Vol. 1, pp. 1-26). Tallahassee, FL: The Association for Institutional Research and AACSB International.

McCarthy, M.A. (2012). Using Student Feedback as One Measure of Faculty Teaching Effectiveness. In Kite, M. E. (2012) (Ed.). Effective evaluation of teaching: A guide for faculty and administrators. Available on the Society for the Teaching of Psychology web site: <http://teachpsych.org/ebooks/evals2012/index.php>

- McGowan, W., & Graham, C. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education* , 34, 161-171.
- McGrath, M., & Braunstein, A. (1997). The Prediction of Freshman Attrition: An Examination of the Importance of Certain Demographic, Academic, Financial and Social Factors. *College Student Journal* , 31 (3), 396-408.
- McKeachie, W. (1997). Student Ratings: The Validity of Use. *American Psychologist* , 52 (11), 1218-1225.
- Miller, R., & Leskes, A. (2005). *Levels of Assessment: From the Student to the Institution*. Washington, DC: AAC&U.
- Nardi, P. M. (2005). *Doing Survey Research, 2/E*. Boston: Ally & Bacon.
- National Center for Education Statistics. (2006). *Digest of Higher Education Statistics*. NCES.
- National Center for Education Statistics. (2013). *Fast Facts*. Retrieved 2013 November 25-11 from National Center for Education Statistics:
<http://nces.ed.gov/fastfacts/display.asp?id=372>.
- National Center for Education Statistics. (2013). *Projections of Education Statistics to 2021*. Washington, DC: U.S. Department of Education.
- National Center for Public Policy and Higher Education. (2008). *Measuring Up 2008*. Washington, D.C.: NCPPHE.
- National Commission on Accountability in Higher Education. (2005). *Accountability for Better Results: A National Imperative for Higher Education*. State Higher Education Executive Officers and the Ford Foundation.
- National Task Force on Institutional Accreditation. (2012). *Assuring Academic Quality in the 21st Century: Self-Regulation in a New Era*. American Council on Education.
- Ory, J., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Teaching and Learning* , 87, 3-15.
- Pace, R. C. (1979). *Measuring the Outcomes of College*. San Francisco: Jossey-Bass.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Penny, A., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: Meta-analysis. *Review of Educational Research* , 74, 215-253.

- Pew Center for Research. (2009). *College Enrollment*. Pew Center.
- Richardson, J. (2005). Instruments for Obtaining Student Feedback: A Review of the Literature. *Assessment and Evaluation in Higher Education* , 30 (4), 387-415.
- Rifkin, T. (1995). *The Status and Scope of Faculty Evaluation*. Los Angeles: ERIC Clearinghouse for Community Colleges.
- Schiro, M. S. (2008). *Curriculum Theory: Conflicting Visions and Enduring Concerns*. Thousand Oaks, CA: Sage Publications.
- Schmelkin, L. P., Spencer, K., & Gellman, E. (1997). Faculty Perspectives on Course and Teacher Evaluations. *Research in Higher Education* , 38 (5), 575-592.
- Scott, W. R. (2001). Constructing an Analytical Framework I: Three Pillars of Institutions. In W. R. Scott, *Institutions and Organizations* (2nd ed., pp. 47-70). Thousand Oaks, CA: Sage Publications.
- Sechrest, L., & Sidani, S. (1995). "Quantitative and Qualitative Methods: Is There an Alternative? *Evaluation and Program Planning* , 18 (1), 77-87.
- Seldin, P. (2006). *Evaluating Faculty Performance*. Bolton, MA: Anker.
- Seldin, P. (1993 21-July). The Use and Abuse of Student Ratings of Professors. *The Chronicle of Higher Education* , p. A40.
- Simpson, P., & Siguaw, J. (2000). Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. *Journal of Marketing Education* , 22 (3), 199-213.
- Smith, B. (2007). Student Ratings of Teaching Effectiveness: An Analysis of End-of-Course Faculty Evaluations. *College Student Journal* , 41 (4), 788-800.
- Smith, K. (1995). Individually and Collectively, Schools Need to Start the Journey Towards Transformation. *AACSB 1995 Annual Meeting*. AACSB.
- Sojka, J., Ashok, K., & Dawn, R. (2002). Student and faculty perceptions of student evaluations of teaching. *College Teaching* , 50 (2), 44-49.
- Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications.
- State Higher Education Executive Officers. (2005). Accountability for Better Results--A National Imperative for Higher Education. *Network News* , 24, pp. 1-4.

Stone, D. (2002). *Policy Paradox: The Art of Political Decision Making*. New York: WW Norton & Company.

Study Group on Conditions of Excellence in American Higher Education. (1984). *Involvement in Learning: Realizing the Potential of American Higher Education*". Washington, DC: Department of Education and National Institute of Education.

Study Group on the Conditions of Excellence in American Higher Education. (1984). *The Progress of an Agenda: A First Report*. Washington, DC: National Institute of Education.
Taut, S., & Brauns, D. (2003). "Resistance to Evaluations: A Psychological Perspective". *Evaluation* , 9 (3), 247-69.

The National Center for Public Policy and Higher Education. (2009). *Measuring Up 2008: The National Report Card on Higher Education*. San Jose: The National Center for Public Policy and Higher Education.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research* , 109, 45-56.

U.S. Department of Education. (2011). *Meeting the Nation's 2020 Goal: State Targets for Increasing the Number and Percentage of College Graduates with Degrees*. Washington, D.C.: U.S. Department of Education.

U.S. Department of Education. (2014). *White House College Scorecard*. Retrieved from <http://www.whitehouse.gov/issues/education/higher-education/college-score-card>.

Wachtel, H. (1998). Student Evaluation of college teaching effectiveness. *Assessment & Evaluation in Higher Education* , 29 (2), 119-121.

Walvoord, B. E. (2004). *Assessment Clear and Simple: A Practical Guide for Institutions, Departments, and General Education*. San Francisco: Jossey-Bass.

Wilson, R. (1998, January 16). New Research Casts Doubt on Value of Student Evaluations of Professor. *The Chronicle* .

Yin, R. (2003). *Case Study Research: Design and Method* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education* , 12 (1), 55-76.

APPENDIX A

Survey Instrument: Faculty Perceptions of Student Evaluations of Teaching

I am trying to gain a better understanding of faculty perceptions and self-reported use of undergraduate student evaluations of teaching as part of my dissertation research. As someone with undergraduate teaching responsibilities, your opinions concerning student teaching evaluations will be very beneficial to my overall research. Questions will generally ask for your level of agreement or disagreement, and some provide space for you to make additional comments should you wish to do so. There are no right or wrong answers.

Your participation in this survey will be confidential and anonymous. The survey software (LimeSurvey) will anonymize all responses, and there is no way to link response data to participants.

I realize that time is a finite and precious resource, so thank you in advance for your participation in this survey!

Scott H. Baker

Note: As several questions within this survey refer specifically to the IDEA Evaluations and Diagnostic Report forms instructors receive, you may find it very beneficial to have a Diagnostic report available while completing this survey. [You may also access sample report forms by clicking here.](#)

The survey questions and LimeSurvey are approved by the Research Protections Office at the University of Vermont (UVM), and I received both UVM and this institution's IRB approval to conduct this survey. If you have questions about your rights as a participant in a research project, you may contact Nancy Stalnaker, the Director of the Research Protections Office at the University of Vermont (802-656-5040).

Your participation in this online survey is completely voluntary. Your consent to participate will be implied upon the completion of the survey.

There are 15 questions in this survey
Perceptions of Student Teaching Evaluations

1. Please indicate your current role as it relates to your undergraduate teaching responsibilities:

Please choose only one of the following:

- Full-time faculty
- Part-time faculty

2. How many years of experience do you have teaching undergraduate college courses?

Please choose only one of the following:

- 1-7 academic years
- 8-14 academic years
- 15+ academic years

Please consider your total years of experience at any institution(s) where the majority of your classes taught per year are undergraduate college level courses.

3. What level of undergraduate courses do you primarily teach?

Please choose only one of the following:

- Primarily 100-200 level
- Primarily 200-300 level
- Primarily 300-400 level
- Mixture: 100-400 level

In any given academic year, what is/are the primary level(s) of undergraduate courses within your teaching load?

4. In general, do you agree with the overall concept of student evaluations of teaching in measuring and informing teaching practices? If you would like to clarify your response, please feel free to do so in the space provided.

*

Please choose only one of the following:

- Strongly agree with the overall concept
- Agree with the overall concept
- Uncertain about the overall concept
- Disagree with the overall concept
- Strongly disagree with the overall concept

Make a comment on your choice here:

Note, subsequent questions will provided opportunity to comment specifically about the IDEA evaluations.

5. In general, to what extent do you agree that students are able to accurately measure their own learning progress in any given course through student evaluations of teaching? If you would like to clarify your response, please feel free to do so in the space provided.

Please choose only one of the following:

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Make a comment on your choice here:

Note, subsequent questions will provided opportunity to comment specifically about the IDEA evaluations.

6. IDEA claims the Student Ratings of Instruction are rooted in a “Student Learning Model” whereby it is important to assess how “specific teaching behaviors influence certain types of student progress under certain circumstances.” To what extent do you agree or disagree with this claim?

Please choose only one of the following:

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

You may find it very beneficial to see a copy of an IDEA Diagnostic Report while answering several of the remaining questions. Please [click here](#) for a sample report.

7. As part of the IDEA Student Ratings of Instruction, instructors are asked to fill out a Faculty Information Form. This form allows instructors to rate predetermined Objectives for each course as either “Essential,” “Important,” or “Minor or No Importance.” IDEA describes this as instructors’ goal attainment related to teaching goals and objectives for each class.

In your opinion, how important is it to compare students' perceptions of course objectives with your own determination of course objectives?

Please choose only one of the following:

- Very important
- Important
- Of little importance
- Very unimportant

8. A major part of IDEA Diagnostic Report Form instructors receive back is the "Student Ratings of Learning on Relevant Objectives." This is a summary (Progress on Relevant Objectives, page 1) and per-objective breakdown (page 2) of students' perceptions as to the importance of the Learning Objectives the instructor marked as "Essential" or "Important" for the particular course.

To what extent do you utilize this information to evaluate your own teaching practices/approach to the particular course?

*

Please choose only one of the following:

- I utilize the information to a large extent
- I utilize the information somewhat
- I do not utilize the information much
- I do not utilize the information at all

IDEA's description student responses concerning relevant objectives:

Average unadjusted (raw) and adjusted progress ratings are shown below for those objectives you identified as "Important" or "Essential." Progress on Relevant Objectives (also shown on page 1) is a weighted average of student ratings of the progress they reported on objectives selected as "Important" or "Essential" (double weighted). The percent of students rating each as "1" or "2" (either "no" or "slight" progress) and as "4" or "5" ("substantial" or "exceptional" progress) is also reported. These results should help you identify objectives where improvement efforts might best be focused.

9. Another major category on the IDEA Student Ratings of Instruction is the "Improving Teaching Effectiveness" section. In this part of the report, IDEA summarizes a comparison of the Teaching Methods and Styles questions asked of students, links to Relevant Objectives, and suggests action such as retaining or increasing use.

To what extent do you utilize this information to evaluate and inform your own teaching practices/approach to the particular course? If you would like to clarify your response, please feel free to do so in the space provided.

*

Please choose only one of the following:

- I utilize the information to a large extent
- I utilize the information somewhat
- I do not utilize the information much
- I do not utilize the information at all

Make a comment on your choice here:

Note, this particular section is only available in the "long" diagnostic report from IDEA (not in the report for the "short form" version. Please see the [Sample Diagnostic Form Report](#) if beneficial.

10. In addition to the Diagnostic Forms, instructors also receive students' comments for each course. To what extent do you utilize students' comments to evaluate and inform your own teaching practices/approach to the particular course? If you would like to clarify your response, please feel free to do so in the space provided.

*

Please choose only one of the following:

- I utilize the information to a large extent
- I utilize the information somewhat
- I do not utilize the information much
- I do not utilize the information at all

Make a comment on your choice here:

11. Considering all of the feedback instructors receive from the IDEA Diagnostic Report (long or short versions), and the sections mentioned in previous questions, in general, how likely are you to use the Report's feedback to inform your teaching practices?

If you would like to clarify your response, please feel free to do so in the space provided.

Please choose only one of the following:

- Very likely
- Somewhat likely
- Somewhat unlikely
- Very unlikely

Make a comment on your choice here:

Note: By "inform your teaching practices" it is meant how likely are you to incorporate the student feedback into your teaching philosophy or practice, either in general or in particular courses. To what extent do you "use" the student evaluations to retain or modify practices?

12. If you have any ideas or suggestions as to how student evaluations of teaching (the process, the tool itself, etc.) might be more helpful to you as an instructor in terms of informing your teaching practices, please feel free to comment here.

Please write your answer here:

13. In general, do you agree with the practice of including student evaluations as a portion of your overall teaching evaluations? Note; this question pertains to your perception of the use of student evaluations in evaluating teaching, not the IDEA instrument specifically. Subsequent questions will provide opportunity to answer/comment on the IDEA instrument/process.

If you would like to clarify your response, please do so in the space provided.

*

Please choose only one of the following:

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Make a comment on your choice here:

14. In your opinion, how appropriately does the institution use the IDEA Student Ratings of Instruction and Summary Reports? If you would like to clarify your response, please do so in the space provided.

Please choose only one of the following:

- Used very appropriately
- Used somewhat appropriately
- Uncertain
- Used somewhat inappropriately
- Used very inappropriately

Make a comment on your choice here:

15. If you have any additional comments (or if you'd like to expand upon a previous comment) or suggestions about student evaluations of teaching in general, or the IDEA Evaluations in particular, please feel free to include them here.

Please write your answer here:

Thank you very much for your time in completing this survey! Again, your participation in this survey will be confidential and anonymous.

Again, many thanks for your time and effort in completing this survey.

Sincerely,

-Scott