2022

Comparing the Assessment Knowledge and Self-Efficacy of Science, Technology, Engineering, and Mathematics Faculty with Faculty from Other Disciplines

Karen Elizabeth Braley
Walden University

Walden University

College of Education and Human Sciences

This is to certify that the doctoral dissertation by

Karen Braley

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. John Flohr, Committee Chairperson, Education Faculty
Dr. Stephen Canipe, Committee Member, Education Faculty
Dr. Charlotte Redden, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2022

Abstract

Comparing the Assessment Knowledge and Self-Efficacy of Science, Technology, Engineering, and Mathematics Faculty with Faculty from Other Disciplines

by

Karen Braley

MS, Dartmouth College, 2009

BS, Trinity College, 1999

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

November 2022

Abstract

Based on reported success and retention rates, there is a need to improve education in 2-year colleges for science, technology, engineering, and mathematics (STEM) majors. Instructors' assessment knowledge and self-efficacy have not been studied in 2-year or STEM higher education. The problem addressed in this study was that scholars do not know the extent of differences in assessment knowledge and assessment self-efficacy of STEM and non-STEM faculty that could be a contributing factor in student success rates. The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment Survey at the 2-year college level. Deci and Ryan's self-determination theory served as the theoretical framework. The research questions were designed to determine if there was a significant difference regarding assessment knowledge and self-efficacy between 2-year college STEM faculty (n = 28) and non-STEM faculty (n = 35) based on the Yin Assessment Survey. Data were collected online from 2-year college in the Southeast using the Yin Assessment Survey. The resulting data were analyzed using an independent t test. Based on violations of assumptions, the Mann-Whitney U test was utilized. No statistically significant differences in assessment knowledge or self-efficacy scores were found, the average scores of the individual questions suggest faculty variance in formative assessment knowledge. The stakeholders that could benefit from this study are students, colleges, and the workforce. Positive social change may occur as STEM student retention increases and more graduates become available for employment in STEM fields.

Comparing the Assessment Knowledge and Self-Efficacy of Science, Technology, Engineering, and Mathematics Faculty with Faculty from Other Disciplines

by

Karen Braley

MS, Dartmouth College, 2009

BS, Trinity College, 1999

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

November 2022

Dedication

I would like to both acknowledge and dedicate this dissertation to my late husband, Michael Pierce. Throughout our 28 years together you were a constant source of encouragement for me. Without your influence and support, I would have never completed this level of education. You always believed in me and were so proud of my accomplishments. You never complained about the time I spend studying or writing. My life was blessed to have a true partner who stood with me as I set out to accomplish my dreams. There is no way to express how valuable your presence in my life was to me. I am grateful for every moment we had together. I wish that you could be here to see this final step in accomplishing my goals. I dedicate this to you, knowing that without you beside me for those 28 years this would never have happened.

Acknowledgments

First, I would like to acknowledge my committee members, Dr. John Flohr and Dr. Steve Canipe. The encouragement and support that they provided gave me the strength to continue despite the many bumps and hurdles. This dissertation would not have been possible without these two individuals. They were so valuable to me during the process. I hope someday to be able to be that type of mentor to a student who is working on their dissertation.

I would like to acknowledge my friend, Karen Koory. Thank you for always being willing to listen to me stress about this process. You encouraged me to keep going even when things were in chaos. I lost my biggest cheerleader halfway through this process, thank you for stepping up and assuming that role. It has meant a great deal to me to have your support, and I share this completed dissertation with you.

I want to acknowledge my co-workers who continued to cheer for me as I went through this process. Many of you also went through this process and you were a reminder that I could and should meet this goal. It was nice to feel like I was never alone in the process.

I would like to acknowledge the administration where I teach for their encouragement. Having the support and belief of individuals with such knowledge in the field of education was inspirational. I value greatly your role in my development as an educator.

Table of Contents

List of Figures v

Chapter 1: Introduction to the Study 1

Background 2

Problem Statement 4

Purpose of the Study 5

Research Questions and Hypotheses 5

Theoretical Framework 6

Nature of the Study 8

Definitions 9

Assumptions 9

Scope and Delimitation 10

Limitations 10

Significance 11

Summary 12

Chapter 2: Literature Review 13

Literature Search Strategy 13

Theoretical Foundation 15

The Origins of the SDT 15

Theoretical Propositions of the SDT 15

The SDT in Educational Literature 16

Rationale and Relatability for the SDT 18

Literature Review of the Study Variables and Concepts 20
Yin Assessment Survey 20
Methodology 21
The 2-Year College Setting 22
Student Success in 2-Year Colleges 23
Student Success in College STEM 25
Assessment 27
The Rationale for the Selection of Variables 32
The 2-Year College Setting 33
STEM and Non-STEM Faculty 33
Assessment in 2-Year College Courses 35
Faculty Self-Efficacy 36
Faculty Perceptions and Self-Efficacy 36
Studies With Related Research 39
Summary 40
Chapter 3: Research Method 43
Research Design and Rationale 43
Methodology 46
Population 46
Sampling and Sampling Procedures 46
Recruitment, Participation, and Data Collection 48
Instrumentation 49

Operationalization of Variables 50
Data Analysis Plan 51
Ethical Procedures 54
Summary 55
Chapter 4: Results 56
Data Collection 57
Results 59
Research Question 1 60
Research Question 2 67
Group Data 74
Summary 77
Chapter 5: Discussion, Conclusions, and Recommendations 78
Interpretation of Findings in Relation to the Theoretical Framework 79
Interpretation of Findings in Relation to the Empirical Literature 79
Limitations of the Study 81
Recommendations 82
Implications for Positive Social Change 84
Conclusion 85
References 86
Appendix A: Demographic Survey 98
Appendix B: Permission to Use Survey 99

List of Tables

Table 1. Operationalization of Variables 51

Table 2. Research Question 1: Descriptive Statistics 60

Table 3. Research Question 1: Test of Normality 61

Table 4. Research Question 1: Independent Samples Test 64

Table 5. Research Question 1: Hypothesis Test Summary 66

Table 6. Research Question 1: Mann-Whitney U Test 67

Table 7. Research Question 2: Descriptive Statistics 68

Table 8. Research Question 2: Test of Normality 70

Table 9. Research Question 2: Independent Samples Test 71

Table 10. Research Question 2: Hypothesis Test Summary 73

Table 11. Research Question 2: Mann-Whitney U Test 74

Table 12. Mean by Question 76

List of Figures

Figure 1. Research Question 1: Boxplot 61

Figure 2. Research Question 1: Mann-Whitney U Test 66

Figure 3. Research Question 2: Boxplot 68

Figure 4. Research Question 2: Mann-Whitney U Test 73

Chapter 1: Introduction to the Study

A current social educational problem is student success in 2-year public colleges. Cohen and Kelly (2019) described as part of their introduction how effectiveness or success is measured at community colleges. They stated that it is measured by completing a degree within 150% of the time that the degree should be completed. The rate of transfer to a 4-year college is also a factor that should be included in the measure (Cohen & Kelly, 2019). Bruck and Bruck (2018) justified the study based on the information that science, technology, engineering, and mathematics (STEM) majors often change majors early in college. As part of the introduction for their study, Scott et al. (2017), wrote that STEM completion rates are low based on data that stated less than half of the students who declare a major in STEM "actually complete the degree within 6 years". In summary, the above literature suggests that STEM completion rates are low.

Due to the low STEM success rates, the purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment survey at the 2-year college level. The Yin Assessment Survey measures assessment self-efficacy and knowledge regarding the practice of assessing students.

Chapter 1 includes information about the purpose of the study and a description of the significance of the problem. The self-determination theory (SDT) that served as the theoretical framework for the study is also described. Additionally, I provide the research questions that guided the study, the nature of the quantitative study, and the study limitations in the remaining sections of the chapter.

Background

The background section includes information about faculty classroom assessment self-efficacy and knowledge in 2-year colleges as measured by the Yin Assessment Survey. I present information about the success rates in this section to show the rationale for the study design. The percentages for completion at 2-year colleges are low in general, with about 40% of enrolled students completing a degree in any field (Shapiro et al., 2017). Soricone and Endel (2019) cited data from the National Center for Educational Statistics showing that "more than 20% of the students enrolled at a Community College will declare a STEM major at some point" (p. 6).

One area where the content areas have a discrepancy is comparing the success rates or course completion rates of students in developmental math (i.e., STEM) versus English (i.e., non-STEM) courses. In general, undergraduate STEM discipline student completion rates are 16% lower than in other disciplines (Malcom & Feder, 2016). The present study, like the work of Yin et al. (2015), was designed to learn more about faculty assessment self-efficacy and knowledge. Unlike the work of Yin et al., I did not include an intervention but instead compared STEM and non-STEM faculty.

In existing studies, improvements in classroom assessment have resulted in an increase in student success and/or learning (Abraham et al., 2018; Akkaraju et al., 2019; Keller, 2017; Laverty et al., 2016). The STEM courses have lower student success rates than non-STEM courses (Malcom & Feder, 2016). Malcom and Feder (2016) mentioned ways for improving STEM courses by making changes in the teaching practice. Some of their suggestions were including formative feedback and using problems or activities that are more genuine. Abraham et al., (2018) also emphasized the importance of assessing applied knowledge using the example of clinically based questions instead of recall in early courses to improve retention of knowledge. Keller (2017) illustrated how formative classroom assessment can be designed to promote student success, showing that the use of formative assessment in 2-year college science classes was an effective way to prepare students for a summative assessment as illustrated by student scores. In addition, formative assessment provides the faculty with a measure of student understanding prior to the summative assessment (Keller, 2017). The faculty member can adjust their approach and highlight student challenges. The addition of formative assessment to a 2-year college biology lab led to greater proficiency in student microscope skills and student success when compared to labs that did not have a formative assessment as part of the course (Keller, 2017).

In this study, I determined the difference in assessment knowledge and self-efficacy of the faculty as measured by the Yin Assessment Survey in STEM and non-STEM instructors to collect information that could have been useful to improve student success. Barton et al. (2020) stated that the design and practice of assessing students has been shown to play a role in student success in higher education. The reason for the lack of student success could be due to the use of classroom assessment strategies that reflect the traditional multiple-choice exams. Koretsky et al. (2016) showed that asking for a written explanation of the answer helped students answer questions correctly. Koretsky et al. also found that written explanations also give faculty a means for understanding how students reason. After a literature search that included multiple sessions with Walden

University librarians, I found no studies that compared the classroom assessment self-efficacy or classroom assessment knowledge of the two groups. With this study, I attempted to provide information that scholars and the field do not know regarding faculty assessment self-efficacy and knowledge.

The self-efficacy of educators and how that translates into the classroom are described in the works of Bandura (1997). Researchers have found a relationship between self-efficacy and online teaching experiences (Hardy et al., 2017; Horvitz et al., 2014). Other STEM faculty self-efficacy studies have found a relationship between self-efficacy and classroom practice or pedagogy (Daou, 2016; Lai et al., 2018). Finally, a positive relationship was found between faculty self-efficacy and working with students with disabilities (Wright & Meyer, 2017). The studies in the extant literature support the concept that self-efficacy relates to practice in the classroom. The Yin Assessment Survey measures the assessment self-efficacy and assessment knowledge of the instructor Yin et al. (2015). In this study, I used the measurement tool created by Yin.

The gap in the literature was supported by numerous literature searches in the databases and holdings of the Walden University Library where I could not find a comparison of faculty assessment knowledge and self-efficacy in non-STEM areas, such as history, English, social science, and humanities. This study was needed to provide information about assessment that does not currently exist in the literature.

Problem Statement

The problem addressed in this study was that scholars do not know the extent of differences in assessment knowledge and assessment self-efficacy of STEM and non-

STEM faculty, which could be a contributing factor in student success rates. STEM programs have low success and retention rates as measured by students failing to obtain a STEM or other degree (Soricone & Endel, 2019). In community colleges, the withdrawal rates from courses in STEM fields, such as science and engineering, is 15.2% and in math, including technology, the withdrawal rate is 17.9% (Cohen & Kelly, 2019). Cohen and Kelly (2019) also reported from other courses, such as English and humanities, the withdrawal rate was 9.2%. Differences in the instructor's assessment knowledge and the assessment of self-efficacy could be a one of the potential factors for the differences in student success.

Purpose of the Study

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment survey at the 2-year college level. This study expanded on the known information about the differences between STEM and non-STEM faculty by measuring assessment knowledge and self-efficacy using the Yin Assessment Survey. The independent variable was the field taught by the faculty member. The dependent variables were the scores on the Likert-style survey developed by Yin et al. (2015).

Research Questions and Hypotheses

Research Question 1: To what extent is there a significant difference in mean scores regarding knowledge about assessment between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey?

$H_0$1: There will be no significant difference in mean scores between STEM faculty and non-STEM faculty scores for knowledge about assessment on the Yin Assessment Survey.

$H_1$1: Non-STEM faculty scores will be significantly higher than STEM faculty for knowledge about assessment on the Yin Assessment Survey.

Research Question 2: To what extent is there a significant difference in mean scores regarding assessment self-efficacy between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey?

$H_0$2: There will be no difference in mean scores between STEM faculty and non-STEM faculty scores for self-efficacy on the Yin Assessment Survey.

$H_1$2: Non-STEM faculty scores will be significantly higher than STEM faculty for assessment self-efficacy on the Yin Assessment Survey.

Theoretical Framework

The theoretical basis for this study was the SDT (Deci & Ryan, 2002). The premise of the SDT is that an individual needs to have three psychological needs met: competence, autonomy, and relatedness for motivation and self-efficacy (Deci & Ryan, 2002). The aspect of this theory that served as a framework for this study was competence. Competence relates to the effectiveness or perceived effectiveness of the individual. Part of the concept of competence is that the individual can adapt and change with new situations over time. Another aspect was the individual's perceived competence. Perceived competence can be influenced by the experiences of the individual (Deci & Ryan, 2002). The need for perceived self-efficacy is part of the core of SDT (Deci & Ryan, 2002).

Furthermore, I selected the SDT because it aligned with this study in that competence and self-efficacy could be positively related with assessment knowledge as measured by the survey tool. Kaygisiz et al. (2018) studied the predictive relationship between self-efficacy and teaching methods and found that self-efficacy was a predictor of teaching methods utilized by the instructors. Eather et al. (2019) reported that preservice teacher perceived confidence and self-efficacy in teachers were not statistically different. Using SDT, specifically competency as a lens to examine self-efficacy and knowledge, provides a view of assessment as part of teaching methods.

The theoretical framework for the Yin et al. (2015) study that developed the survey tool being used in this current study was the social learning theory by Bandura. The work of Bandura (1997) supports the idea that educators need competence to build self-efficacy. Bandura stated self-efficacy or perceived competence can be influenced in the educational setting by an increase in training. Based on Bandura's ideas, Yin et al. measured self-efficacy and knowledge based on a professional development intervention. In the current study, I measured baseline self-efficacy and knowledge. The competence portion of the SDT serves as a lens for examining self-efficacy instead of the social learning theory. Social learning is used as a framework for studies with an intervention, this study did not include an intervention, so the SDT was used.

Nature of the Study

In this study, I employed a quantitative survey design to determine if there was a difference in mean scores between 2-year college STEM faculty and 2-year college non-STEM faculty in Florida on the Yin Assessment Survey. A quantitative focus was consistent with the goal of comparing the mean scores for assessment knowledge and assessment self-efficacy between the STEM and non-STEM faculty. In the original publication of the survey instrument, Yin et al. (2015) used a quantitative study to compare pre- and postintervention. Additionally, in another study, a quantitative survey allowed for distribution to occur via email and reduce bias due to anonymity (Frankfort-Nachmias & Nachmias, 2008).

Although there may be other differences in teaching between STEM and non-STEM faculty, the scope of this current study was limited to assessment knowledge and self-efficacy. The independent variable for the experiment was the field (i.e., STEM or non-STEM) taught by the faculty member. The dependent variable was the score on the Likert-style Yin Assessment Survey (Yin et al., 2015) to measure the instructor's assessment knowledge and self-efficacy. Permission for the use of the survey was obtained from the author (see Appendix B). I included a separate question to determine the content area taught by the faculty (see Appendix A). After the Institutional Review Boards (IRBs) provided approval, I distributed the survey via email to the voluntary participants. Data collected from the study were analyzed using IBM Statistical Package for Social Sciences (SPSS) software obtained from Walden University.

Definitions

Formative assessment: An assessment measure given to provide the student and the instructor with information about the student's level of understanding. Formative assessments are often intermittent (Offerdahl et al., 2018).

Perceived self-efficacy: An individual's belief in their ability to perform a task to produce the desired outcome (Deci & Ryan, 2002).

Summative assessment: An assessment measure given to determine the amount of learning that has taken place over a period of time. Summative assessments are higher stakes assessments (Dolin & Evans, 2018).

Assumptions

The first assumption of the study was that the participants would have completed the survey honestly. The faculty members may have had preconceived notions about how they should answer the questions. Faculty bias could have also influenced the answers given. A large sample size can be used to reduce bias in the results (Frankfort-Nachmias & Nachmias, 2008).

The second assumption was that faculty had knowledge of classroom assessment techniques, including formative assessment. The survey used as part of this study was designed to collect data about faculty members' assessment knowledge and included statements about their use or practice of the assessments.

I also assumed that the participants' knowledge of formative assessment that was addressed in the survey translates into their use of formative assessment in the classroom. The scope of this study was to look at ways faculty can contribute to student success, and assessment knowledge and self-efficacy is one aspect of teaching practice that could be helpful in classroom practice.

Scope and Delimitation

The study's scope consisted of a Likert-style survey that I used to measure the assessment knowledge and assessment self-efficacy of STEM and non-STEM instructors regarding the practice of assessing students. The participants were from a midsized, 2-year community college in Central Florida. I send participants a request to participate in the survey via email. The demographic portion of the survey also included information regarding the faculty content area. The demographic survey is available in Appendix A.

The delimitations of the study included the choice of the target study population. I focused on 2-year college faculty in this study because the faculty at community colleges see their primary job as teaching, and the open enrollment environment contributes to low success. Four-year college faculty were excluded from this study because some faculty may not see teaching as a primary job. Other delimitations for the study were the variables chosen (i.e., the STEM versus non-STEM content areas and the assessment knowledge and self-efficacy as measured by the Yin Assessment Survey).

Limitations

I identified a limitation of the study related to internal validity and the responses obtained from the faculty. A limitation that influenced the internal validity was the unequal sampling from the two study groups. A larger response from one of the groups representing the independent variable was a limitation that was considered via the statistical method selected for the study. Another limitation that was a factor for external validity was the study was conducted regionally and may not be representative of the STEM and non-STEM groups globally. A final limitation was the influence of the COVID-19 pandemic on the modality that the faculty members currently use to teach. Teaching in a format that is unfamiliar to a faculty member could have influenced their perceived self-efficacy and the classroom assessment strategies used. The COVID-19 pandemic was not an issue for the distribution of the survey, however, because it was conducted electronically, and no interviews were required. At the time the survey was conducted, the faculty had returned to the classroom and were not teaching strictly online.

Significance

The problem addressed in the study was that scholars do not know the extent of differences in assessment knowledge and self-efficacy of STEM and non-STEM faculty that could be a contributing factor in student success rates. This study is significant because the information gathered about faculty assessment knowledge and self-efficacy as measured by the Yin Assessment Survey could be used to improve completion rates in STEM. There are two ways the study findings advance the knowledge in the STEM discipline.

The study advances knowledge regarding STEM education through contributing to the success of students early in their STEM course work. One reason that STEM course success is important is that students entering a STEM major will experience an initial requisite course, such as algebra or biology, early on in their education. The assessment method, or pedagogy, can be influential in the student experience or success (De & Arguello, 2020; Krulatz, 2017; Rodriguez et al., 2018).

The findings from the present study about faculty assessment self-efficacy and knowledge provide information that can be used to improve STEM education and student success. The stakeholders for this change are not only the students but also the faculty, administrators, and researchers. An increase in students successfully completing a STEM degree could help fill the job market needs (Doerschuk et al., 2016). The implications for positive social change associated with this study are the possible improvements in student success that could benefit the job market and would be beneficial to society due to the millions of STEM jobs that cannot currently be filled (see White & Shakibnia, 2019).

Summary

Chapter 1 includes information about STEM student success and the need to improve success rates. I detailed the purpose of this quantitative study and provided a brief discussion of the nature of the study and design.

In Chapter 2, I present a literature review that begins with a list of terms and databases used for the search. The first section of Chapter 2 includes the origins and application of the SDT as a framework for research studies. In the next section, I provide background on the constructs used in the study before moving on to a discussion of extant literature that supports the selection of the variables, STEM faculty, non-STEM faculty, assessment, and self-efficacy. The final section of the chapter contains information about previous studies with similar designs.

Chapter 2: Literature Review

The problem addressed in this current study was that scholars do not know the extent of differences in assessment knowledge and self-efficacy of STEM and non-STEM faculty that could be a contributing factor in student success rates. The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment survey at the 2-year college level.

In the literature review, I describe the SDT that served as a framework for the study. The theory relates to the meeting of physiological needs to influence motivation (Deci & Ryan, 2002). In the review, I provide examples of other educational studies that also used the theory as a framework for educational studies. The chapter also includes a review of relevant literature related to the concepts of the study, including 2-year colleges, student success, and STEM. The current literature illustrates the challenges with the concepts of the study: success in 2-year colleges and STEM. Additionally, the chapter contains a review of the literature that relates to the variables for the study: STEM faculty, non-STEM faculty, assessment, and faculty self-efficacy regarding the practice of assessing students. The review also includes a discussion of the rationale, background for the study, the methodology, and the Yin Assessment Survey. The literature review begins with a description of the literature search strategy.

Literature Search Strategy

I used the Walden University Library to conduct searches for literature sources. The select all databases function in Walden's online library was used. Searches were also conducted using the Education Source and ERIC databases and the Google Scholar search engine. I initially conducted searches for literature published in or after 2016. When those searches did not yield many sources, the date searched was expanded to include articles published in 2010 or after. The following keyword search terms were used: *assessment, science, faculty, student, success, college, class, motivation, training, factors, chemistry, biology, retention, 2-year, strategies, self-determination theory, alternative, active learning, community college, STEM, self-efficacy, assessment practices, assessment knowledge, formative assessment, summative assessment, 2-year college assessment, 2-year college performance measures, 2-year college student performance, 2-year academic evaluation, academic assessment, STEM faculty, liberal arts faculty, higher education, post-secondary, instructor motivation, knowledge, faculty surveys, STEM formative assessment, humanities, social sciences* and *non-STEM*. Books were used as sources, when needed, to supplement the information published in journals. The reference lists of the books were also used to find additional journal articles in the literature. Some of the references have publication dates that fall outside the 5-year range due to the lack of availability of sources.

I conducted this literature search described over an extensive period and consulted with the Walden University librarians for four searches to exhaust all potential sources. After the completion of the literature searches, I found there was an abundance of research related to the assessment practices in the Prekindergarten through grade 12 (PK-12) system, but there was also a lack of literature regarding assessment in STEM fields. The literature that met the time parameters of being published within the last 5 years was included in the review. The focus of the extant studies was on specific assessment strategies as opposed to faculty assessment knowledge and self-efficacy. There was also a lack of studies comparing STEM faculty assessment and self-efficacy with those of non-STEM faculty. Although there are several references to STEM success rates being lower than non-STEM success rates, there is a lack of comparative studies examining the specific reasons for the discrepancy at the college level.

Theoretical Foundation

The SDT served as the theoretical framework for the current study. The premise of the SDT is that having three psychological needs met (i.e., competence, autonomy, and relatedness) is a drive for motivation (Deci & Ryan, 2002). The SDT also considers the role of intrinsic and extrinsic motivation.

The Origins of the SDT

Deci and Ryan (2002) are credited with having the first practical evidence of the SDT. The original work was conducted in the 1970s, but their research was not accepted as a theory until the 1980s (Deci & Ryan, 1985). Since the theory was published over 20 years ago, this area of study has developed and been applied to education (Deci & Ryan, 2008).

Theoretical Propositions of the SDT

One of the aspects of the SDT is that the three equal basic psychological needs must be met for optimal development (Deci & Ryan, 2002). One need is competence. Competence is the need for the individual to experience mastery and feel control for the outcome. The next need is autonomy, which is described as having a feeling of independence without necessarily being independent or isolated from others. The last need is relatedness. This refers to a feeling or having a connection to other individuals (Deci & Ryan, 2002).

In addition to having the three basic psychological needs met, the SDT also includes motivation as part of the theory. Intrinsic motivation is the internal drive or desire to seek out and try new things (Deci, 1995). Intrinsic motivation has been described as related to the individual's self-efficacy (Bandura, 1997). Extrinsic motivation comes from outside of the individual and can be related to various forms of regulation (Deci & Ryan, 1985). Regulation that is based on an external reward or requirement is one type of extrinsic motivation. Other types of extrinsic motivation, such as wanting to show success or reach a goal are considered extrinsic but with a component of autonomy (Deci & Ryan, 1985).

The key assumption in the SDT is that the individuals have intrinsic motivation (Hodgins & Knee, 2002). In the theory, it is also assumed that motivation is related to the individual's behavior and that people are motivated to progress and develop in the varying aspects of their lives (Deci & Ryan, 2016). I applied the assumptions of the SDT to the current study because I also assumed that educators are self-motivated to improve their assessment knowledge and self-efficacy.

The SDT in Educational Literature

Zwart et al. (2015) used the SDT as part of their rationale for examining development in PK-12 teachers in the Netherlands. The goal of the mixed-methods study was to measure the self-efficacy, autonomy, competence, and relatedness of development. As part of the study, the teachers were given the opportunity to coach others as well as learn from peers. The results indicated that coaching and interaction with other teachers resulted in an increased self-efficacy and autonomy.

A similar study was conducted by Karaarslan et al. (2013) measuring the "needs of the of preservice teachers" by using environmental science (considered part of the STEM discipline) course activities designed with SDT. The course had embedded a framework to ensure the three needs of the SDT (i.e., competence, autonomy, and relatedness) were included as part of the training. The study results showed preservice teachers having their needs met by the course resulted in increased self-confidence, self-initiation, and awareness of self-behaviors. The data from the qualitative portion of the mixed methods study suggested cognitive and instruction support could be helpful for satisfying the needs of the participants (Karaarslan et al., 2013). Their study is an example of the SDT being used in a study with a population of STEM educators and illustrates that the components of the SDT are important in the educator being able to make changes.

Hu and Zhang (2017) applied the SDT to examine the role of the psychological need of autonomy in the learning process among students in an English as a second language course. The researchers used a mixed-method approach with surveys and interviews and created an action program to make sure that the need autonomy was met for the learners (Hu & Zhang, 2017). They found that the learner's satisfaction and progress were improved by the implementation of the action program; however, the use of the action program for 1 year was not enough to result in a complete feeling of autonomy for the learners.

Sanchez-Olivia et al. (2017) conducted a study using the SDT to explore the training of physical education PK-12 teachers. The aim of the study was to compare the students of teachers that participated in a training program to improve the student's autonomy, relatedness, and competence with the students from teachers that did not participate. The results reported from the study indicated the teachers that had participated in the training for both autonomous motivation and perceived need support reported a positive experience (Sanchez-Olivia et al., 2017).

The studies of Hu and Zhang (2017) and Sanchez-Olivia et al. (2017) were outside the STEM discipline; however, the studies are examples of the use of the SDT in educational research. The need for a sense of competence is part of the design of the SDT (Deci & Ryan, 1985). Karaarslan et al.'s study (2013) was conducted within the STEM discipline, and the results supported the need for competence. I did not find any STEM studies directly linked to the SDT that were published in the last 5 years. Due to the lack of current studies in the literature, the Karaarslan et al. study was included as an example of STEM research that applied the SDT.

Rationale and Relatability for the SDT

The studies described above, the original work of Bandura (1997), and the work of Deci and Ryan (1985) regarding teacher self-efficacy served as the rationale behind the use of the SDT as the framework for this study. The specific aspect of the SDT that was applied to the current study is competence. Other ideas included in this study are self-efficacy being a sense of self-competence and knowledge correlating to competence. Although there were not specific previous studies measuring assessment through the lens of the SDT, other studies have been conducted in the educational environment measuring other aspects of teaching and learning. Zwart et al. (2015) and Karaarslan et al. (2013) both used SDT to examine the role of training in self-efficacy and competence. These studies showed that improvement in competence and knowledge relates to the individual's self-efficacy. In the current study, I did not specifically look at any intervention but focused on the baseline assessment knowledge and self-efficacy of the instructors.

Bandura (1997) stated that a teacher's ability to create an optimal learning environment is related to the self-efficacy of the teacher. The teacher's self-efficacy is related to their approach and the activities that they utilize (Bandura, 1997). The needs of the instructor being met can directly translate to the needs of the students being met (Deci & Ryan, 2016). Thus, the self-efficacy of the teacher can play a role in the success students experience in the classroom. In the current study, I examined the assessment knowledge and self-efficacy of the instructors using the need for competence from the SDT theory. In summary, the SDT theory may not have been previously used specifically for a study of assessment competence in instructors, but it has been used as a framework for general classroom practice as described above. Assessment is part of the general classroom practice.

The overall rationale for the use of SDT in this study was related to the competence portion of the SDT. If the instructor has a higher assessment knowledge and self-efficacy, they will have the sense of competence that is necessary, according to the SDT. The SDT also aligns with the work of Bandura (1997) who stated that self-efficacy is important to classroom practice. If the hypothesis of this study was correct and there was a difference in the knowledge and self-efficacy of instructors, the findings could be used as the foundation for a study like that of Sanchez-Olivia et al. (2017) where training was implemented to improve the competence need of the instructors with the goal of improving the classroom experience. If the null hypothesis from this study was shown as valid, it would suggest that competence is not likely the factor. The other aspects of the SDT theory could then be applied to assess if there are differences in autonomy, relatedness, or motivation.

Literature Review of the Study Variables and Concepts

In this section of the literature review, I provide background information regarding the variables of the study, including assessment, self-efficacy, STEM faculty, and non- STEM faculty, as well as the rationale for the selection of the variables. The concepts of student success and 2-year colleges are also detailed.

Yin Assessment Survey

I used the Yin Assessment Survey (Yin et al., 2015) that was designed to determine the outcome of two professional development programs for math teachers. The researchers used Likert-style surveys to measure the teacher’s assessment knowledge and self-efficacy. The survey instrument was used four times through the professional development activity. Their conclusion was that there was increase in assessment knowledge and self-efficacy after professional development. The work by Yin et al.

differs from the current study because they used the survey to measure faculty responses after an intervention, and I measured faculty responses without an intervention.

In a mixed-method study, Alotaibi (2019) examined assessment knowledge and self-efficacy using an assessment survey as the tool. The study examined the relationship between teacher perceptions and the use of formative assessment, with the teachers reporting using formative assessment to determine student progress in learning but not for the measurement of outcomes (Alotaibi, 2019). The perceptions of formative assessment difficulties were higher with older teachers, and confidence was higher with tenured teachers (Alotaibi, 2019). Alotaibi compared the responses of teachers as opposed to college instructors. Unlike the Alotaibi study, I compared results from the survey based on the content area taught in the current study.

Methodology

I chose the independent *t* test for statistical analysis in this study. The independent *t* test has been used for other published studies that focus on comparisons. For example, Suwito et al. (2020) used a quasi-experimental design to examine the effect of a textbook on learning outcomes in geography among two groups: an experimental group who used the 5E learning cycle (a learning model) and the control group who used the traditional method. The researchers administered pre- and posttests to the students. Data from the pretest was used to determine homogeneity of the sample population. The researchers used the independent *t* test to compare the results from the pre- and posttests, finding that students had an improved gain of knowledge in the experimental group. The Suwito et al. study varies from the current study because it includes pre- and posttests and intervention. Suwito et al.'s study also was conducted using a student population. Like the current study, however, Suwito et al. used the *t* test for comparing two groups.

Another study similar to the Suwito et al. study (2020) was conducted to examine the flipped classroom pedagogy for teaching science skills in a sixth-grade classroom (Shana & Alwaely, 2021). In this study, there was also a control and an experimental group. Both groups were given a pretest and posttest. The experimental group was taught using the flipped classroom method and the control group was taught using the traditional method. The pretest scores from the study were analyzed and compared using the *t* test. The conclusion from the study was that the flipped classroom had a beneficial effect on learning (Shana & Alwaely, 2021). Although this study also differs from the study in the use of an intervention, a pretest, and a student population it does show the use of the *t* test for data analysis between two group.

The 2-Year College Setting

The study utilized 2-year college-instructors as the population. The unique aspects of the 2-year setting are documented in the literature (Pratt, 2017; Taylor & Giani, 2019; Torres et al., 2018). Public 2-year schools often have an open enrollment policy that does not require students to have a previous record of academic success (Pratt, 2017). The faculty at 2-year colleges are thus additionally tasked with teaching unprepared students (Pratt, 2017). Although the open enrollment policy allows for the inclusion of anyone wanting to attend college, the students may not be successful. Based on this, 2-year colleges have been criticized for low completion rates of students (Torres et al., 2018). In some instances, the low completion rate is also accompanied by debt. The task of increasing student success and completion without compromising education is one problem 2-year colleges are addressing (Taylor & Giani, 2019).

Another unique aspect of the 2-year college setting is the effectiveness of the institution. Researchers at the national level use actual and predicted graduation rates as a means for measuring effectiveness. A quantitative study was conducted to expand on the variables used to determine the effectiveness and compare the effectiveness of 2-year colleges across the country (Horn et al., 2019). The study included completion rates, structural characteristics, demographic characteristics, financial and contextual attributes, and support services. The data were analyzed using linear mixed regression analysis (p. 159). Conclusions that were drawn from the study are that learning gains are not measured by effectiveness. Effectiveness measures do not include preparation for workplace and citizenship upon graduating (Horn et al., 2019). The final issue was the effectiveness measures did not adequately account for completion rates in Michigan and Kansas (Horn et al., 2019). One of the problems with the study by Horn et al. (2019), focused on completion rates for degree and did not have student transfer rates from all of the colleges. Transfer rates are considered a performance indicator for 2-year college students. The data reported by Horn et al. is an example of the need for more studies that can help understand student success at the 2-year college. The current study examined one aspect, assessment, at the 2-year college.

Student Success in 2-Year Colleges

Numerous studies examined factors that influence student success have been conducted. Elements that can be used to predict student success in college are enrollment in nonremedial courses only, completing all credits attempted, and persistence to complete the first year and enroll the second year (Stephan et al., 2015). Other studies determined that a consideration that predicts student success in college is the student's preparedness for the rigor of the courses. Only 28% of the students enrolled in remediation courses are likely to graduate, but 43% of students who do not require remediation are likely to graduate. Remediation courses can increase both the time to degree completion and cost (Benken et al., 2015).

Colleges are challenged to find ways to improve success and retention; one of the methods used is a general college success course. Interventions that can aid students in skill building such as reading course material have been shown to improve student success (Shaffer et al., 2015). Assessment studies at the community college level focused on faculty and not specific assessments strategies are limited. A study by O'Gara et al. (2008) used a qualitative design to look at student persistence for first semester community college students. Two schools offered student success courses as part of the study. Students reported that the course was a helpful tool for navigating the resources available at the college but did not care for the manner they were made available. The students stated that the course professors were also helpful for developing a course plan. The professors in the introductory courses were reported to be more knowledgeable about individual advising than the general advisors (O'Gara et al., 2008). The result of a relationship with a professor can enhance student success. Finally, the course was developed to enhance relationships with other students (O'Gara et al., 2008). The study illustrates how professors at 2-year colleges may need to guide students in general academic skills. The study results point to a need for students to receive support to adapt and be successful in college. Requiring a success course can be a problem due to credit number restrictions and the lack of advising at a community college. This study focuses on existing classes and not require any changes to student schedules or programs. The study is designed to examine existing practices that could relate to student success.

Another approach that has been successful at 2-year community colleges is the use of tutoring centers for math, which is part of the STEM courses, to help students (Jaafar et al., 2016). Data collected from students showed that 45% of them would have dropped or potentially dropped the course without tutoring. Students failing remedial math courses are at risk for dropping out (Jaafar et al., 2016). This paper shows an alternative approach that could be useful for multiple disciplines.

Student Success in College STEM

This study was designed to compare STEM and non-STEM faculty assessment knowledge and self-efficacy. Adapting to the college environment can be difficult for students and students may not be prepared for the expectations of college science courses (Schwartz et al., 2008). In addition, students struggle with math courses in college (Jaafar et al., 2016). The information presented supports the need for interventions.

Westat et al. (2015) provide information about STEM reform in their paper. They provided details for areas that could be considered for change. Some of the ideas that are related to the classroom are curriculum, pedagogy, and assessment. This current study looked at assessment knowledge and self-efficacy in 2-year college faculty.

As stated in the introduction to their work, retention issues are common in the STEM discipline at the community college (Bruck & Bruck, 2018). Various challenges that students face during the college years can influence if they remain a STEM major. One quantitative study focused on the resources available to students in general chemistry which is a course within STEM at a 2-year college. The study found that students preferred instructor created resources, such as notes and practice problems (Bruck & Bruck, 2018). The study illustrates the importance of instructor engagement and its role in student success. The students' preference for materials created by the instructor over the textbook supplied resources suggests that faculty might need to be more involved in the creation of materials.

Lysne and Miller (2017) conducted a study in a 2-year college biology course, part of the STEM umbrella. The treatment group received more activities and assessment styles compared to the control group. They did not report a significant difference in performance in the control or treatment group. The treatment group did have better course retention by 5% (Lysne & Miller). This study illustrated how changes in the course can lead to other positive outcomes other than grades, such as retention.

Another study targeted at improving a STEM course, examined student success in an introductory biology course by measuring engagement and achievement based on the class size (Scott et al., 2017). Student success, retention, and engagement were compared between a large class with 70-80 students and smaller classes with 20-30 students. The classes compared were taught by the same instructor. The findings were that the students in the smaller classes were more likely to be engaged than the larger classes. Based on the data obtained from the *t* tests used to compare the two groups, the students in the smaller class were also more successful on their assessments. The implication is that the smaller class size is beneficial to the students. The authors did acknowledge that this is not always a feasible option, and other means for improving student success may be necessary (Scott et al., 2017). The study was designed to collect information about assessment which is another aspect of improving success.

The studies by Lysne and Miller (2017) and Scott et al. (2017) focused on specific courses and did not examine the STEM field as a whole. There was also no comparison with non-STEM fields. Their studies did not look at the influence of the faculty members views or knowledge about assessment. The study by Scott et al. mentioned that class size could be a factor in the biology class, but without a comparison to other class sizes inside and outside of the STEM field the conclusion does not indicate that class size is the sole contributing factor. The study was unique because it compared the STEM and non-STEM faculty assessment knowledge and self-efficacy to help elucidate the lower STEM success rates when compared to other content areas. This study examined STEM as a whole and not focus on individual courses or student groups like previous studies.

Assessment

First, the literature related to STEM assessment will be outlined. This discussion will include recent trends in STEM assessment. Formative assessment from the Yin et al. (2015) study will be discussed Finally, formative assessment in STEM will be outlined.

Assessment is included in the construct description because the survey used in the study relates to self-efficacy and assessment knowledge. The survey also includes formative assessment specific questions and thus formative assessment will also be included in the review. The assessment practice in STEM classes is described in this portion of the literature review. Assessment refers to both the use of assessment to determine student mastery of the material and the use of student assessment data to improve the course.

The goal of a STEM graduate degree is focused on research with little if any training for educational concepts like pedagogy (Baiduc et al., 2016). STEM faculty often have limited or even no classroom experience when they start teaching in the college classroom and feel ill prepared. This lack of training related to education could also mean that assessment is another area where the faculty are ill prepared. The reported lack of training for STEM instructors is part of the background behind the comparison of STEM and non-STEM faculty in the study.

One of the important concepts in science assessment are the difference between assessment of skills versus assessment as a mean for measuring outcomes (Hanauer & Baruerle, 2012). In science, a multiple-choice exam where students are required to pick the exact answer. This type of assessment can be focused on recall. Some of the focus on new assessment methods in science have been focused on assessing higher order thinking and not recall (Hanauer & Baruerle, 2012). The information presented suggests that assessment is an area within science and potentially STEM that could benefit from further study.

The lack of training and the unprepared feeling of faculty reported by Baiduc et al., (2016) could mean that the faculty have lower self-efficacy. There is a reliance in science on traditional methods that measure recall (Hanauer & Baruerle, 2012). The use of recall assessment could be part of the problem for other STEM disciplines and suggests more information about assessment could be informative. Types of assessments could account for success differences in STEM versus non-STEM. The goal of this current study was to study the assessment differences.

A STEM assessment specific study conducted by Cotner and Ballen (2017) examined the role of mixed method assessment in creating an equitable STEM classroom. The goal was to determine if the emphasis on high stakes exams as the sole determinant of student success would lead to more equitable assessment. Low-stakes assessments, such as group participation and modeling were used in the evaluation of the students (Cotner & Ballen, 2017). They found a gender gap in performance on high-stakes assessment with women attaining lower scores. They did not find evidence that lower scores achieved by the women was related to class size. Another aspect of the active learning class is there is a need for students to be prepared for class due and that preparation is rewarded (Cotner & Ballen, 2017). Based on the literature, assessment is important for success, but the comparison of assessment for STEM and non-STEM is not available. Information about the assessment beliefs in STEM versus non-STEM could be useful for determining if the need for assessment reform is discipline specific or broad based.

A majority of the research focused on STEM assessment does not examine faculty knowledge or self-efficacy. Varsavsky and Rayner (2013) alternative assessments that required students to use critical review of a paper was implemented in biology. Advanced laboratory exercises that required students to work in groups were also added to chemistry. The students were surveyed to get their perceptions of the alternative assessments. The students reported a more in-depth understanding as well as a more stimulating experience with the alternative assessment (Varsavsky & Rayner, 2013). Studies like the one described above, have shown that assessments can influence understanding by implementing new strategies. The study was seeking to understand more about the faculty and their basic knowledge and self-efficacy regarding assessment. Understanding the faculty could be part of the process of encouraging change in the classroom. The studies related to STEM assessment have been focused on assessment in the one discipline of aspect of STEM.

Another idea in college science assessment is to move to assessing conceptual understanding and higher-order thinking (Hanauer & Baruerle, 2012). One pedagogical technique that has been shown to help students build knowledge and understanding is collaborative learning (Lawrie et al., 2014). First year science classes at the college level can be large. The size of the class can make it difficult for an instructor to implement, manage, and assess group learning (Lawrie et al., 2014). A mixed-method study by Lawrie et al. (2014) utilized a web-based task management software to complete group activities, peer assessment, and analysis of artifacts. The study showed that in well-functioning groups, students developed interdependence. The study also determined that the student’s technology skills and being supported were contributing factors to the benefit of the method (Lawrie et al., 2014).

Another assessment mixed method study by Siegel et al. (2015) used collaborative group assessments. The data indicated that students in the group assessments performed better with conceptual items. Students were interviewed and reported that they felt an increase in engagement, less anxiety, and that their thinking was stimulated using the group assessment (Siegel et al., 2015). The faculty that was interviewed stated that they felt the students learned the materials based on the process. The faculty also reported that they observed students' collaboration skills and that that strategy was helpful in a diverse classroom. The problem that the faculty reported with the assessment was the lack of resources available to them to implement the strategy (Siegel et al., 2015). The studies described above from the literature are focused on a specific assessment design or method. The studies showed that diversity and alternative assessment strategies can be useful for measuring success and inclusive for students. The studies do not provide adequate details about the lack of use of assessment methods in the classroom. The present study could elucidate differences in the STEM faculty and the non-STEM faculty that influence their assessment practices.

Formative Assessment

The Yin Assessment Survey includes questions specific to formative assessment. The role of formative assessment is described in this section. An example of the effectiveness of formative assessment in the college science classroom is the use of formative assessment to improve student skills when using a microscope (Keller, 2017). The study was conducted in the biology lab at a community college. The instructor utilized a series of formative assessments including questions, drawings, and reflections for students to assess their skills prior to the summative assessment. Although the sample size was too small to be conclusive, students that regularly attended did have a better understanding of their skill set before the summative exam (Keller, 2017). This study did not conclusively show that formative assessment improved success due to the sample size, but the information that is lacking is the use of formative assessment in the STEM college classroom. This study provides information the knowledge of regarding formative assessment in the STEM and non-STEM disciplines.

Formative assessment has been established as an important aspect of significant learning and understanding (Kim et al., 2019). Formative assessment is often missing in STEM classes for a number of reasons, such as class size. The study by Kim et al. (2019) was focused on a specific formative assessment method using a learning researcher, an undergraduate trained to collect evidence. They reported the experience provided them with a better sense of the student experience and could be used to improve teaching (Kim et al.). The use of formative assessment in STEM and non-STEM at the college level is not documented. The literature described provides evidence that it is important for the learning and assessment process. The knowledge and faculty self-efficacy regarding formative assessment is needed to determine if this is an area that should be explored for college STEM assessment reform.

The Rationale for the Selection of Variables

The following section will discuss the reasoning for the selection of the study settings and participants. The rationale of the selection of the variables is included.

Information from the literature is provided as part of the rationale for the selected variables.

The 2-Year College Setting

There is a need to improve education for the STEM field, especially during the first 2 years (Malcom & Feder, 2016). The 2-year college can serve as a feeder for the 4-year schools for many STEM programs. Community colleges can survive as a starting point due to attributes that include low cost and an open-door policy.

A second reason that the 2-year college setting was selected is because of issues with completion rates in general for students. One of the reasons is students have difficulty with courses such as college algebra (Gaze, 2018). It is being considered algebra may not be the most appropriate gateway course (Gaze, 2018). These are some of the reasons the 2-year college was selected for the study.

The rationale for this setting was the diverse group of students that are represented in this setting could benefit from faculty with more training and diversity in assessment practice. The 2-year college has some of the most diversely prepared students due to their open enrollment policies and have the lowest completion rates. The faculty in this setting could benefit the most from increased skills and knowledge.

STEM and Non-STEM Faculty

The independent variable for this study was if the faculty member is STEM faculty or non-STEM faculty. The faculty were divided into groups based whether they were STEM or non-STEM faculty. STEM faculty were chosen for two reasons. The lack of training for STEM educators was documented by Baiduc et al. (2016), but direct comparisons with other disciplines were not available.

English as foreign language faculty value formative assessment as a learning tool (Guadu & Boersma, 2018). A case study was conducted by Rohrbacher (2017) examined the views of humanities professor's conception of assessment. The study was conducted at two universities. The results indicated that the humanities professors were enthusiastic about using assessment as a tool for informing them about teaching and learning. The faculty were less enthusiastic about general education assessment (Rohrbacher, 2017). These examples of assessment views are limited to one course and do not consider differences between disciplines.

The open beliefs of non-STEM faculty are supported by the research by Guadu and Boersma (2018) that was designed to examine the beliefs and practices of English instructors and formative assessments. The study was designed to focus on English as second language professors. The data from the mixed-method study indicated that the faculty had a positive perception of using writing as a formative assessment. The survey data also illustrated that despite their perceptions of formative assessment as valuable, it did not reflect in their practice of the assessment (Guadu & Boersma, 2018). The data from the survey suggested that faculty in some of the disciplines in non-STEM have a knowledge of formative assessment, which is one of the variables in the Yin Assessment Survey tool.

Assessment in 2-Year College Courses

The dependent variable in the study was the faculty members scores on the Yin Assessment Survey. The survey measures assessment knowledge and self-efficacy. The rationale for the dependent variable will be described in this section. Assessment in STEM courses has traditionally centered around recall (Hanauer & Baruerle, 2012). In the study conducted by Cotner and Ballen (2017), the reform of the assessment practices in the classroom the use of low stakes assessments was helpful for decreasing the gender performance gap. The study design implemented the use of a mixed-methods assessment. The researchers were unable to compare high stakes environments with low stakes environments. The performance gap based on gender was not observed in the low stakes assessments (Cotner & Ballen, 2017). This modification is an example of how assessment can influence the success of students and how faculty knowledge about the types of assessments could be useful for improving performance. The results from published studies show that reviewing course materials can improve the course experience such as the use of authentic work (Gray et al., 2017) and using self-assessment (Sit & Brudzinski, 2017). Another study showed the benefit of examining the pre-existing math skills and knowledge of students in early in their enrollment engineering and calculus, disciplines of STEM (Nortvedt & Siqveland, 2018). Another rationale for studying assessment was to determine if the faculty understand how assessment can be used to improve a course or course materials.

Faculty Self-Efficacy

Self-efficacy was included as a dependent variable because it is a concept that was measured by the Yin survey instrument. Researchers have reported that improvements in self-efficacy ad knowledge in graduate assistants can occur through training programs (Reeves et al., 2018). Additionally, Bandura's (1997) social learning theory includes the value of learning communities and training in self-efficacy.

Faculty Perceptions and Self-Efficacy

The self-efficacy of the instructor can directly influence the planning of educational strategies and goals (Bandura, 1997). A study conducted by Velthuis et al., (2014) that showed that training at the preservice level has a positive role on the self-efficacy of teachers. Based on the design the data illustrated the benefits of training on teacher self-efficacy, but not the self-efficacy of college professors. The study goal was to expand upon what is described above regarding assessment by adding information about self-efficacy to the assessment literature.

Faculty-self efficacy and its role in teaching have been studied within the online teaching format. A study conducted by Horvitz et al. (2014) examined the self-efficacy of instructors that were teaching online. Other factors that they found that influenced the instructors online self-efficacy scores were association with a professional school, and the number of semesters taught online (Horvitz et al., 2014). Another study focused on online teaching was conducted by Hardy et al. (2017). This study looked at the role of self-efficacy in part-time teachers online teaching. Hardy et al. found a positive correlation between self-efficacy with the instructor satisfaction with teaching. This study showed that chronological age also influenced the instructor's perception of their mastery. Teachers with a higher age perceived a higher level of preparedness to teach their content (Hardy et al., 2017).

Lastly, a study by Gosselin (2016) examined the self-efficacy and development of instructors teaching online. The self-efficacy of the faculty was shown to increase in relation to time. The researchers reported in the study that selecting technical resources and virtual interactions were two of the areas where less self-efficacy was observed, and development activities could be useful (Gosselin, 2016). The studies by Horvitz et al., Hardy et al. (2017), and Gosselin all studied self-efficacy in relation to online teaching. The current study was examining self-efficacy and assessment and not online teaching. The studies were included in the review even though they do not measure assessment because the data illustrates that self-efficacy can be related to faculty classroom practice. Additional studies designed to examine the importance of educator self-efficacy in areas of teaching are described in the literature and included below.

Other studies which were examined, looked at instructor self-efficacy and aspects of the classrooms. A study was conducted to determine the role of instructor self-efficacy and the self-disclosure of students with disabilities (Wright & Meyer, 2017). The study results indicated that when reading the scenarios in the study the faculty had a decrease in empathy and flexibility the professor had regarding accommodations. It was also reported that if the faculty self-reported student disability, they reported an increase in perceived self-efficacy in relation to making the accommodation (Wright & Meyer, 2017). The above study is another example of the importance of self-efficacy and educator practice.

The inclusion of self-efficacy is part of the justification of self-efficacy as a variable in the study.

Faculty self-efficacy and utilizing alternative classroom practices, such as a flipped classroom have been studied (Lai et al., 2018). The study also looked at motivation and opportunity. The study was designed to examine the faculty's continued use of the flip model classroom at the college level. The researchers found the faculty with a high-level motivation and a high reported self-efficacy were likely to persevere and continue using the flipped classroom. They also found that faculty who received supportive resources were more likely to continue the flipped format (Lai et al., 2018).

Another area where the instructor's self-efficacy and proficiency can influence the classroom is the use of technology (Daou, 2016). Faculty self-efficacy was improved using training model. The availability of training regarding technology suggests the instructor is more likely to implement technology into the classroom (Daou, 2016). The studies by Wright and Meyer (2017), Lai et al. (2018) and Daou also indicated that faculty self-efficacy was associated with classroom practice. Wright and Meyer found self-efficacy important for faculty who needed to make accommodations for students. The other studies found self-efficacy related to the persistence and use of classroom tools such as flipped classrooms and technology (Daou; Lai et al.). The studies were included because they further illustrate the importance of faculty self-efficacy in the practice of teaching. This study examined assessment self-efficacy. The studies described in the literature review by Wright and Meyer, Lai et al., and Daou examined self-efficacy but not assessment self-efficacy. The above studies do support the importance in self-efficacy in educators which was a variable in the present study.

The role of instructor self-efficacy and how it translates into the classroom has also been shown in several studies. In students with disabilities, the higher the self-efficacy of the faculty member, the more confident with making accommodations (Wright & Meyer, 2017). The higher the technology self-efficacy of the faculty was influenced by their proficiency, and this was important for their use of technology in the classroom (Daou, 2016). Faculty with higher self-efficacy are more likely to continue to use a new classroom methodology such as the flipped classroom (Lai et al., 2018). Finally, faculty with higher self-efficacy are comfortable with the online teaching modality (Gosselin, 2016; Hardy et al., 2017). The idea that the improvements and practice of teaching is influenced by the instructor's self-efficacy is rationale for the selection of self-efficacy as a variable. Research suggests self-efficacy impacts technology and other aspects of teaching.

Studies With Related Research

The studies described below are similar to the present study. A study completed by Simpson (2016) examined assessment in multiple disciplines in the United Kingdom. The researcher collected information by discipline such as art, biology, and mathematics. The data indicated that the assessments used varied by discipline (Simpson, 2016). The study design did not include the faculty knowledge or the self-efficacy as a variable but focused on types of assessment and grades. There was no information provided about the level of autonomy that faculty had regarding the development of assessment. The study published did not have a lot of background information about why the assessment methods were used. The lack of data suggested that it might be useful information to reexamine the differences between the content area from another perspective. The study examined the knowledge and self-efficacy of the faculty and be independent of factors such as development of assessment.

Another example from the literature that compares to the study was conducted to compare the creative process skills for students by discipline (Daly et al., 2016). The data were collected using a survey instrument administered to students near the end of the course. The results indicated that the creative process was greater in the humanities, social sciences, and art. The STEM discipline of engineering had less creativity reported (Daly et al., 2016). This study does not examine assessment, but it does compare STEM to non-STEM content courses. The study by Daly et al. (2016) was included, although it is not assessment is an example of a study that was conducted with the goal of comparing disciplines. The information from these studies supports the idea that potential differences could exist between STEM and non-STEM discipline and suggests the hypothesis used in the study of assessment differences between the disciplines was valid.

Summary

Based on the literature review, numerous factors can be used to predict a college student's success and persistence to remain enrolled in school. Briefly summarizing the literature, a variety of approaches to increase student success and retention at the college level. Additional problems exist for student success and retention in STEM fields, such as science. Although students play a key role in their success, the literature shows that faculty and classroom strategies can have an influence. There are individual studies detailing the use of alternative assessment strategies and formative assessment tools. However, there is little evidence that these tools are used widespread. The literature supports the idea of reforming the science education process to be more inclusive (Cotner & Ballen, 2017). Teaching includes developing techniques and expanding your knowledge (Rawn & Fox, 2018,). Studies focused on the self-efficacy of faculty were described, but the self-efficacy studies did not address assessment. The literature review provides brief details regarding differences in STEM and non-STEM courses in higher education. The studies reported general differences in learning and assessment.

The literature review details what is known from studies conducted in the areas of self-efficacy, knowledge, assessment, and STEM versus non-STEM. The self-efficacy and knowledge of faculty has been studies in other aspects of the education process, such as the use of technology. The research reports indicated that increased knowledge let to an increase in self-efficacy. The literature supported differences in STEM and non-STEM in regard to critical thinking and creativity. A study conducted comparing assessment in STEM and non-STEM examined metacognitive differences, but no studies were found in the literature examining differences in the faculty.

There were no studies that compared assessment directly between STEM and non-STEM. There is a gap in the literature in the assessment knowledge and self-efficacy of STEM faculty compared to non-STEM faculty knowledge. The present study was designed to help addresses the gap that exists about faculty and assessment knowledge.

The third chapter of this study addresses the method, approach, and analysis that was used to conduct this study to examine this gap in the literature.

Chapter 3: Research Method

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment survey at the 2-year college level. I used a Likert-style survey to assess STEM and non-STEM faculty assessment knowledge and self-efficacy regarding the practice of assessing students. In this chapter, I describe the quantitative survey tool used, the setting, and the participants. In addition, the data collection and the statistical analysis processes used to analyze the data are provided. Finally, the threats to validity of the study are included in the discussion.

Research Design and Rationale

I employed a quantitative survey design in this study. The independent variable for the study was STEM faculty and non-STEM faculty. Demographic information was obtained from the survey to classify faculty that teach STEM or non-STEM courses. The dependent variable was the mean score on the Likert-style survey developed by Yin et al. (2015) to measure assessment self-efficacy and assessment knowledge regarding the practice of assessing students. Permission for the use of the survey was granted via email with the agreement that the authors would be referenced in all published and unpublished work

The use of the Likert-Style questionnaire allowed for the quantification of the data. I examined the relationship between the variables through the statistical methods and described in the Data Analysis section. The data were analyzed using an independent *t* test and the Mann-Whitney U test. The rationale for the use of an independent *t* test was that the *t* test is designed to be used when comparing two groups or independent variables (see Laerd Statistics, 2020). The *t* test is appropriate to compare the mean scores between two groups (Laerd Statistics, 2020). The three basic assumptions for the *t* test are two independent variables, a continuous dependent variable, and independence of observation (Laerd Statistics, 2020). In the current study, I used a Likert-style survey for the continuous dependent variable and each participant was only able to complete the survey once. The independent *t* test requires that there are no significant outliers, and this assumption can be examined using a boxplot (Laerd Statistics, 2020). This was addressed in the current study using simple box plots to determine if there were any outliers. Another assumption that must be met for the independent *t* test is the dependent variable should be normally distributed for both groups. I addressed this using the Shapiro-Wilk test for normality. Lastly, the homogeneity of variances was the final assumption, and this assumption was examined by the use of Levene's test for equality of variances (see Laerd Statistics, 2020). There were violations of assumptions for both Research Question 1 and Research Question 2. Laerd Statistics (2020) advised using a nonparametric test based on the violation of assumptions, so the Mann-Whitney U test was completed. Chapter 4 contains a report of the data for testing the assumptions, the independent *t* test, and the Mann-Whitney U test.

There were some potential anticipated issues with time or resource constraints in this study. One potential issue was if a low sample size was received from the first college, the survey might have needed to be sent to a second college. The use of a second college could have meant the completion of a second IRB approval process at the additional college. I did not recruit a second college for the study due to time constraints with dissertation completion and because the survey would be sent during the summer semester when faculty are not required to be on campus.

The survey instrument was consistent with the one used by Yin et al. (2015) and no questions were altered. I used the Yin Assessment Survey to establish faculty assessment knowledge and self-efficacy regarding the practice of assessing students. Unlike the study conducted by Yin et al., this study did not include any type of intervention. I used a demographic questionnaire in place of the intervention as the dependent variable.

The following research questions and hypotheses guided this study:

Research Question 1: To what extent is there a significant difference in mean scores regarding knowledge about assessment between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey?

$H_0$1: There will be no significant difference in mean scores between STEM faculty and non-STEM faculty scores for knowledge about assessment on the Yin Assessment Survey.

$H_1$1: Non-STEM faculty scores will be significantly higher than STEM faculty for knowledge about assessment on the Yin Assessment Survey.

Research Question 2: To what extent is there a significant difference in mean scores regarding assessment self-efficacy between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey?

$H_0 2$: There will be no difference in mean scores between STEM faculty and non-STEM faculty scores for self-efficacy on the Yin Assessment Survey.

$H_1 2$: Non-STEM faculty scores will be significantly higher than STEM faculty for assessment self-efficacy on the Yin Assessment Survey.

Methodology

This section contains details about the methodology of the study. The population, sample, setting, data collection and instrumentation used for the study design are discussed. In this section, I also describe the data analysis plan.

Population

The study population was STEM and non-STEM 2-year college instructors who were currently teaching at least one course at a 2-year college. Faculty who taught in a variety of platforms were eligible for participation in the study. Faculty who were considered part of the STEM population teach in one or more of the fields of science, technology, engineering, and mathematics. All faculty teaching other content areas were classified as non-STEM.

Sampling and Sampling Procedures

I used a single-stage random sampling method by distributing emails to all faculty at the study site institution. Based on feedback from the Walden University IRB office, the distribution occurred via Survey Monkey from my Walden University email address and the invitation did not come as an attachment from a third party. Although public email addresses were used, the study site IRB was contacted because all the sampling would occur at this one institution. The target institution requested to be the IRB of record, and this was approved by the Walden University IRB office. The Walden University IRB office had an abbreviated IRB for the study. At the time of the study, the target institution had a total of 827 instructors teaching either full-time or part-time. There were 227 faculty that had a continuing contract at the time of the study. The sampling process required subjects to be separated into two groups after completion of the survey. I based the selection on the content area taught as reported by the subjects in the demographic survey. All faculty teaching science, engineering, mathematics, and technology were classified as STEM, and faculty teaching in other content areas were classified as non-STEM. For inclusion criteria in the study, the invitation and survey instrument tool were only sent to faculty at the specific study site institution.

I classified the subjects as STEM or non-STEM faculty prior to the completion of the statistical analysis. A large sample size was targeted. Based on analysis using G*Power 3.1.9.7, the target sample was a total of 102 participants to ensure the most representative group. The actual sample size that was obtained was 68, but five surveys were not used because the respondents did not complete all the questions. If an equal sample between the two groups was not obtained, an equal random sample from each group could have been used for analysis. Due to the lower than desired sample size, I used the 28 STEM responses and the 35 non-STEM responses for analysis. G-power 3 analysis was used assuming a medium effect for Cohens *f*. The power used for the calculation was 0.80 and an alpha of 0.05 was used. An effect size of 0.50 is considered a medium effect sample size (Geert van de Berg, n.d.). A target sample of 102 was not obtained and with the goal for this study to have a medium effect sample size, I used G*Power 3.1.9.7 to recalculate the sample size using 0.80 for power, an alpha of 0.05, and an effect size of 0.68. The change to the sample effect size meant that it was still within the medium effect range, but the sample size number was reduced to 56.

Recruitment, Participation, and Data Collection

I used Survey Monkey as the recruitment tool and sent the survey out as an invitation to participate. Faculty received a total of three emails regarding the request for their participation in the study. A total of 68 participants were recruited, and 5 surveys were eliminated because they were not complete. There were 28 STEM responses and 35 non-STEM responses for analysis. The survey did not require any debriefing or follow up with the participants.

Participation was voluntary, and no incentives were provided. Surveys were distributed using Survey Monkey and consent was obtained through the survey using the first question as a consent. I did not obtain signed consent forms based on the request of the study site IRB which required total anonymity for the participants. This anonymity also included that the institution was not named in the dissertation and that the advisor to the student and the Walden University IRB be the only external sources identified.

The demographic data collection occurred using a demographic survey, and I collected the assessment knowledge and self-efficacy data via the Yin Assessment Survey. The survey details and the survey questions used for each of the research questions are provided in the following sections of this chapter. All the collected data were analyzed using SPSS.

Instrumentation

I used the Yin Assessment Survey (Yin et al., 2015), a Likert-style survey instrument, to measure the participants' assessment knowledge and self-efficacy. The survey was developed by Yin et al. (2015) in a study to compare types of professional development. The process of the development and validation of the survey by Yin et al. (2008) included examination by experts and pilot testing. The pilot test utilized 50 teachers that would not be participating in the published study. The original survey instrument was edited based on feedback from the pilot. The reliability statistics for the survey are published within that study and indicated that the items were reliable with alpha coefficient scores ranging from .69 to .97. Any items that were not within the acceptable range were removed from the instrument by Yin et al. during the development of the final survey tool.

The survey was designed to measure the general assessment knowledge, the formative assessment knowledge, and the assessment self-efficacy of the faculty regarding the practice of assessing students. The participants ranked their perceived knowledge or efficacy on a 1–5 Likert scale. On the scale of the survey, 5 = *strongly agree* and 1 = *strongly disagree* (Yin et al., 2015). The scores on the survey were then compared for the STEM and non-STEM faculty by conducting both the independent *t* test and the Mann-Whitney U test as described in the Data Analysis plan. I separated the participants into groups as outlined in the Sampling section of the chapter and the questions used for analysis are described in the following subsection.

I obtained permission for the use of the survey from Dr. Yin in 2019, with the understanding that the authors would be cited and credited in all works. The demographic questions were unique to this study (see Appendix A). The information from the demographic survey was used to group participants into the STEM and non-STEM categories based on content area taught.

Operationalization of Variables

In this quantitative study, I originally used the *t* test to determine the difference in the mean scores on the Yin Assessment Survey. The nine survey questions that were used for assessment knowledge are the first nine questions in the Yin survey. In the survey that was distributed for this study, the assessment knowledge questions were numbered 3–11. For this study, Question 1 was the consent and Question 2 was the demographic question (Appendix A). The second survey question had assessment self-efficacy as the dependent variable. The Yin Assessment Survey had 11 Likert-style questions that address assessment self-efficacy. In the survey that was distributed for this study, the assessment self-efficacy questions were numbered 12–22. Once the survey distribution ended, I downloaded the raw data into SPSS, and all STEM faculty were coded a 1 and all the remaining faculty were coded a 2. For assessment knowledge, SPSS calculated the total score for each individual on Questions 3-11. As an example, if Participant 22 had a 5 on Question 3, 3 on Question 4, 3 on Question 5, 4 on Question 6, 4 on Question 7, 3 on Question 8, 5 on Question 9, 4 on Question 10, and a 2 on Question 11, the total would be determined by SPSS to be 33. I compared the total scores from each of the STEM participants and non-STEM participants using the independent *t* test and the Mann-

Whitney U test. The self-efficacy questions followed the consent, demographic question, and the assessment knowledge questions. The independent variable for both research questions was the content area taught by the faculty and that information was self-reported in the demographic survey. Table 1 displays information about the operationalization of the variables.

Table 1

Operationalization of Variables

Description	Variable	Type	Range of Scores
IV	Content area taught	Categorical	STEM or Non-STEM
DV 1	Mean difference in assessment knowledge	Interval	1 to 5
DV 2	Mean difference in assessment self-efficacy	Interval	1 to 5

Data Analysis Plan

I first analyzed the data using an independent *t* test as originally proposed after the raw data were collected by the survey software and collated to a downloadable file. The data collected were analyzed using the SPSS statistical software Version 27 provided through Walden University. For Research Question 1, the 95% confidence interval for the mean of the STEM group was 37.92 for the lower bound and 41.51 for the upper bound. For the non-STEM group, the 95% confidence interval for the mean was 36.54 for the lower bound and 39.91 for the upper bound. Table 2 in Chapter 4 contains the results for the descriptive statistics. For Research Question 2, the 95% confidence interval for the mean for the STEM group was 42.77 for the lower bound and 48.66 for the upper bound. For the non-STEM group, the 95% confidence interval for the mean was 42.63 for the lower bound and 46.63 for the upper bound. Table 7 in Chapter 4 contains the results for the descriptive statistics. The data is presented through charts and graphs created through Microsoft Word and SPSS software.

The independent *t* test was used to compare the mean scores for assessment knowledge and assessment self-efficacy between STEM and non-STEM faculty. The use of the independent *t* test is the same analysis method used by the other researchers comparing two groups (Shana & Alwaely, 2021; Suwito et al., 2020). The independent *t* test was the original statistical method proposed for the study. Due to assumption violations detailed below and in Chapter 4 a Mann-Whitney U test was also conducted.

The assumptions related to validity associated with the independent *t*-test include having a continuous dependent variable, the independent variable is two groups, and there is independence of observation (Laerd Statistics, 2020). Other assumptions are that there are no significant outliers and that the dependent variable is normally distributed (Laerd Statistics, 2020). These assumptions were addressed with a simple boxplot and Shapiro-Wilk test. Finally, the assumption is that there is homogeneity of variance, and this was assessed using Levene’s test (Laerd Statistics, 2020). The homogeneity of the sample size was assessed prior to the statistical analysis of the data. The tests for assumptions were completed and violations for assumptions were found for both research questions. The numeric test results for Research Question 1 are shown in Figure 1, Table 3, and Table 4 in Chapter 4. The numeric test results for Research Question 2 are shown in Figure 3, Table 8, and Table 9 in Chapter 4. The difference the violations had in on this study was that a non-parametric test would need to be conducted Laerd Statistics (2020). The Mann-Whitney U test was therefore completed in addition to the independent *t*-test. The data from the tests conducted appear in Chapter 4, the results chapter.

Threats to Validity

Both internal and external validity threats can influence the credibility of a study (Creswell & Creswell, 2018). The external validity of the study was the generalization of using human subjects. The data may not reflect all STEM and non-STEM faculty or teachers. The study was conducted at one college located in one state. The results may not be reflective of other 2-year colleges. This study only examined 2-year college faculty and that data cannot be generalized to instructors at the 4-year college or in the PK-12 setting.

Internal selection bias can occur through the use of a convenience sample (Frankfort-Nachmias & Nachmias, 2008). Faculty that participated in this study may or may not have done so based on an interest in classroom assessment or improving the college educational process. Another internal validity issue is human error. Participants can make mistakes or skip questions when completing the survey. Incomplete surveys were excluded to eliminate part of this validity issue. Any surveys that had skipped questions were not included in the data analysis. Additionally, participants might have answered questions in the manner they assume is correct rather than based on their actual feeling.

To ensure credibility of the analysis descriptive statistics were reported. Conclusions that are drawn from the data are linked to the research questions by the study and limit assumptions. Additionally, conclusions are supported by best practices as determined by the literature (Creswell & Creswell, 2018).

The proposal called for a total of 102 participants to ensure the most representative group. A total of 68 participants responded and 65 of the response met the criteria for inclusion. There were 28 STEM responses and 35 non-STEM responses for analysis. The statistical analysis included averages, and the reliability of these numbers are shown using standard deviation from the mean calculations.

Ethical Procedures

Information obtained from attending IRB office hours indicated that using publicly available email addresses for faculty would be considered an ethical approach for recruitment. Recruitment was occurring at one target institution therefore the institution was contacted to ensure that any IRB needs or concerns for the institution were addressed. All data collected for the study was done with IRB approval from the target institution as the IRB of the record (Approval No. 22-03-01) and Walden University with a secondary IRB (Approval No. 03-25-22-0242485). I collected data by emailing faculty directly and not using a third party. The survey was completed using Survey Monkey. Participation was voluntary, and faculty did not receive more than three requests to eliminate the perception of harassment. The only information collected about the participants was if they teach STEM or non-STEM courses. No other identifying information was collected, and the participants remained anonymous. The data was stored in a secure file and for the time, as outlined in the IRB approval.

Summary

The Yin Assessment Survey was utilized in a quantitative study which measures assessment knowledge and self-efficacy regarding the practice of assessing students. The dependent variable was the score on the Yin Assessment Survey. The mean scores for questions 3-11 were used to for Research Question 1, assessment knowledge. The mean scores for questions 12-22 were used for Research Question 2, assessment self-efficacy. The independent variable is the self-reported content area taught by the instructors. Faculty were recruited via their email associated with the 2-year college. Participation was voluntary, and no interventions were part of the study.

The method of statistical analysis used in this study were the independent *t* test and the Mann-Whitney U test. The data included analysis for reliability as part of the data analysis process. The data obtained from the study was addressed in Chapter 4 of this manuscript.

Chapter 4: Results

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment survey at the 2-year college level. The research questions and hypotheses for this study were:

Research Question 1: To what extent is there a significant difference in mean scores regarding knowledge about assessment between 2-year college STEM faculty and 2-year college non-STEM faculty in Florida on the Yin Assessment Survey?

$H_0$1: There will be no significant difference in mean scores between STEM faculty and non-STEM faculty scores for knowledge about assessment on the Yin Assessment Survey.

$H_1$1: Non-STEM faculty scores will be significantly higher than STEM faculty for knowledge about assessment on the Yin Assessment Survey.

Research Question 2: To what extent is there a significant difference in mean scores regarding assessment self-efficacy between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey?

$H_0$2: There will be no difference in mean scores between STEM faculty and non-STEM faculty scores for self-efficacy on the Yin Assessment Survey.

$H_1$2: Non-STEM faculty scores will be significantly higher than STEM faculty for assessment self-efficacy on the Yin Assessment Survey.

Chapter 4 includes information about the collection process, descriptive statistics, and the data collected in the study.

Data Collection

Data collection began after I received approval from the target institutions and the Walden University IRB. The survey was administered using Survey Monkey. I emailed the survey to participants from my Walden University email account. The first question was an acknowledgement of consent. The second was the demographic question (Appendix A), and the remaining questions were from the Yin Assessment Survey without any alterations. The initial research plan was to leave the survey open for 3 to 4 weeks and to ensure the distribution took place at least 3 days after the faculty returned from spring break. Due to the complications detailed in this chapter, the survey remained open to participants for 8 weeks.

In Chapter 3, I originally stated that the survey would only be emailed once to the participants. The target institution had received a large phishing attempt prior to the distribution of the survey. Due to the phishing attempt, the institution increased their email filtering from outside email addresses. The first distribution attempt did not reach a majority of the participants. I contacted the institution's Information Technology department to make them aware that I had IRB approval from the study site to distribute the survey from an email address not associated with the institution. After confirming with the study site IRB office and Walden University IRB office that the active, approved IRB plan did not prohibit a second distribution, the survey was redistributed to increase the number of participants receiving the email. Administrators of the study site institution advised the chairs of each academic department that their faculty could complete the survey. The number of participants per academic department was thus dependent on the department chairs notifying their faculty of the legitimacy of the survey. Due to the failure of the proper distribution of the survey, the participation window was increased to accommodate information regarding the legitimacy of the study to be relayed to the participants. The study site IRB stated that I would not be actively recruiting at the institution, thus the reliance was on the college to disseminate this information.

I conducted the initial calculation regarding sample size using G*Power 3.1.9.7, which resulted in a minimum sample size of 102 participants with 51 participants from each of the groups, making up the independent variable. The initial calculations were completed using a power of 0.80, an alpha of 0.05, and an effect size of 0.50. An effect size of 0.50 is considered a medium effect sample size (Geert van de Berg, n.d.). The range for the medium effect sample size is 0.50 to 0.80 (Geert van de Berg, n.d.). Based on the complications that occurred during the distribution of the survey, I used G*Power 3.1.9.7 to recalculate the sample size using 0.80 for power, an alpha of 0.05, and an effect size of 0.68. The change to the sample effect size meant that it was still within the medium effect range, but the calculation reduced the sample size number to 56. The reduced sample size would mean that for the STEM and the non-STEM groups, a minimum sample size of 28 would be required. The reduced sample size number was more appropriate for the study based on the technological limitations.

In the initial proposal, I stated that another comparable institution could be used if the desired sample size was not reached. The most compatible institution had an IRB that only meets twice per year. The approval process would not have begun until fall of 2022. Additionally, the approval timeline from the other institutions would have meant the survey would be distributed during the summer semester. Many of the teaching faculty do not work the entire summer semester, thus potentially resulting in a limited response. The option of using a second institution would have increased the timeline for this study outside the approved window. Therefore, based on the noted limitations, I deemed it not feasible to use a second similar institution.

The sample sizes between the two groups of STEM and non-STEM were not equal. A total of 68 faculty participated in the survey, with 28 of the participants from the STEM disciplines and 40 from non-STEM disciplines. There were five participants from the non-STEM group that did not complete all the questions in the survey, and their data were excluded from analysis. The final statistics were conducted using 28 STEM and 35 non-STEM participants. The population only represents 2-year college faculty at one institution. The data may not represent faculty at other institutions and in other geographical locations.

Results

In this section, I present the results for Research Questions 1 and 2. The descriptive statistics, *t* test, and means are included for each of the research questions. The results from the *t* test indicated that the assumption of normality was violated. If the assumption of normality is not met, a nonparametric test must be conducted (Laerd Statistics, 2020). Based on the violation, I performed a Mann-Whitney U test, and the results are provided. Information regarding the individual questions for the groups are also provided in this section.

Research Question 1

For this question, the two independent variables were STEM and non-STEM faculty. The participants were classified into one of the two groups based on self-reported content area taught: STEM (*n* = 28) and non-STEM (*n* = 35). I classified STEM faculty into Group 1 and non-STEM faculty into Group 2. Data for the STEM group is shown as mean ± standard deviation (*n* = 28, 39.71 ± 0.88). Data for the STEM group is shown as mean ± standard deviation (*n* = 35, 38.23 ± .83). For the STEM group, the minimum value of the variable was 30.00 and the maximum value 45.00. For the non-STEM group, the minimum value of the variable was 26.00 and the maximum value 45.00. The complete descriptive statistics are found in Table 2.

Table 2

Research Question 1: Descriptive Statistics

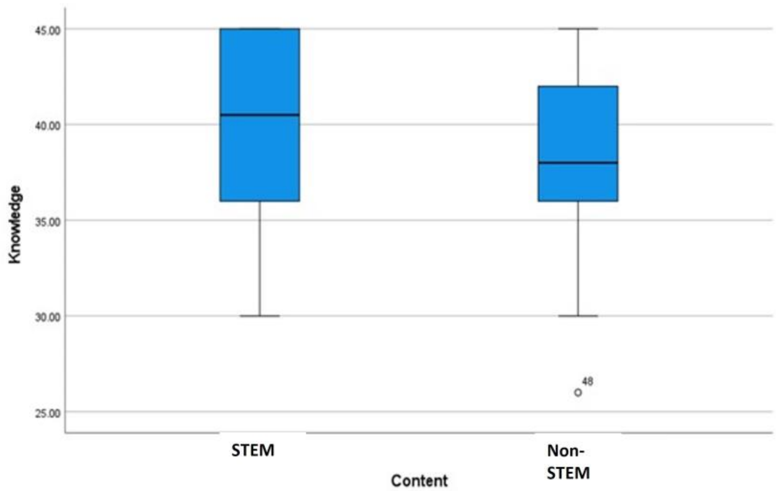
					95% confidence interval for mean			
	N	*M*	*SD*	Std. error	Lower bound	Upper bound	Minimum	Maximum
STEM	28	39.71	4.63	0.88	37.92	41.51	30.00	45.00
Non-STEM	35	38.23	4.91	0.83	36.54	39.91	26.00	45.00

The fourth assumption for the independent *t* test is the presence of no significant outliers in the data. This was tested by completing a boxplot (see Figure 1). The two groups representing the independent variable are located on the X axis and labeled content. The independent variable represents the content area taught by the participants.

The Y axis represents the total scores for each individual. It is their total score for all of the questions in the dependent variable added together. The questions that relate to Research Question 1 are Questions 1–9 on the Yin Assessment Survey. There was one outlier in the non-STEM faculty (i.e., Group 2 illustrated in the boxplot in Figure 1). I made the decision to retain the outlier and proceed with the independent *t* test.

Figure 1

Research Question 1: Boxplot



The fifth assumption for the independent *t* test is that the dependent variable has normal distribution for each of the independent variable groups. The Shapiro-Wilkes test results are presented in Table 3 and discussed below.

Table 3

Research Question 1: Test of Normality

	Shapiro-Wilk		
	Statistic	*df*	Sig.
STEM	0.88	28	.005
Non-STEM	0.95	35	.080

There are six assumptions that need to be met for the independent *t* test. In the following list, I describe the assumptions, how they were assessed, and the outcome.

1. The first assumption was that there was one dependent variable, and it was measured on a continuous scale. The dependent variable was assessment knowledge, and the measurement used the mean score for the first nine questions in the Yin Likert scale survey. This assumption was met.
2. The second assumption was that there was one independent variable that consists of two independent groups. The assumption was met because there were two independent groups: STEM, and non-STEM faculty.
3. The third assumption was independence of observations. None of the participants in the survey indicated that they taught a STEM and a non-STEM course; therefore, this assumption was met.
4. The fourth assumption for the independent *t* test was that there should be no significant outliers in the data. There was one outlier in the non-STEM faculty (i.e., Group 2 illustrated in the boxplot in Figure 1). Examination of the outlier's answers in the survey did not indicate any missed questions. The

answers tended to be lower than other faculty, which could be the actual assessment knowledge of the participant. The outlier was retained and used in the completion of the independent *t* test.

5. A fifth assumption for the independent *t* test was that the dependent variable is normally distributed for each of the groups represented in the independent variable. The Shapiro-Wilkes test results are presented in Table 3. The results from the Shapiro-Wilkes test ($p < 0.05$) determined assessment knowledge scores were not normally distributed for the STEM group; however, the non-STEM had normal distribution of assessment knowledge scores ($p > 0.05$). The sample sizes were under 50 so a normal Q-Q plot was not used to assess distribution. The data were not transformed because only one of the groups was not normally distributed. Thus, the *t* test was completed with the violation of the assumption reported.

6. The sixth assumption was that there is homogeneity of variances, or the variances in each group are the same. I conducted Levene's test of homogeneity of variances as part of the independent *t* test and presented the results in Table 4. The *p* values of 0.11 and 0.22 indicated homogeneity of variance. The assumption of homogeneity was not violated as assessed by Levene's test. I analyzed the *t* test results using the values for equal variances assumed. Based on the violation of normal distribution for the STEM group, a nonparametric analysis was used.

Table 4

Research Question 1: Independent Samples Test

	Levene's test for equality of variances		*t* test for Equality of means						95% confidence interval of the difference	
					Significance					
	F	Sig	*t*	*df*	One-Sided *p*	Two-Sided *p*	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.002	.96	1.2	61	0.11	0.22	1.49	1.21	-0.94	3.91
Equal variances not assumed			1.2	59.28	0.11	0.22	1.49	1.21	-0.93	3.90

The Levene's test indicated that there was no violation of homogeneity. Due to the assumption being met, I used the values in the first row (i.e., equal variances were assumed) for the independent samples *t* test. Based on the results of the independent *t* test, there was no statistically significant difference in mean assessment knowledge scores between STEM and non-STEM faculty, *t* (61) = 1.23, *p* = 0.23. Based on the lack of a statistically significant difference in mean scores, I failed to reject the null hypothesis. Based on the violation of assumption for normal distribution, the nonparametric Mann-Whitney U test was completed, and the results are provided below.

The Mann-Whitney U Test has four assumptions that need to be met. In the following list, I describe the assumptions, how they were assessed, and the outcome.

1. The first assumption was that there needed to be one dependent variable that is measured on a continuous scale. The dependent variable was assessment

knowledge, and the measurement used was a Likert scale survey, so this assumption for the Mann-Whitney U test was met.

2. The next assumption required that there was one independent variable and that it consists of two groups independent of each other. This assumption was met because there were two independent groups: STEM, and non-STEM faculty.
3. For the Mann-Whitney U test, the third assumption was that there is independence of observations. None of the participants in the survey indicated that they taught a STEM and a non-STEM course; therefore, this assumption was met.
4. The fourth assumption was that the distribution of scores for both the groups making up the independent variable must be determined. Based on the values obtained, the distribution patterns are not similar, so the mean rank calculated as part of the Mann-Whitney was reported.

Figure 2

Research Question 1: Mann-Whitney U Test

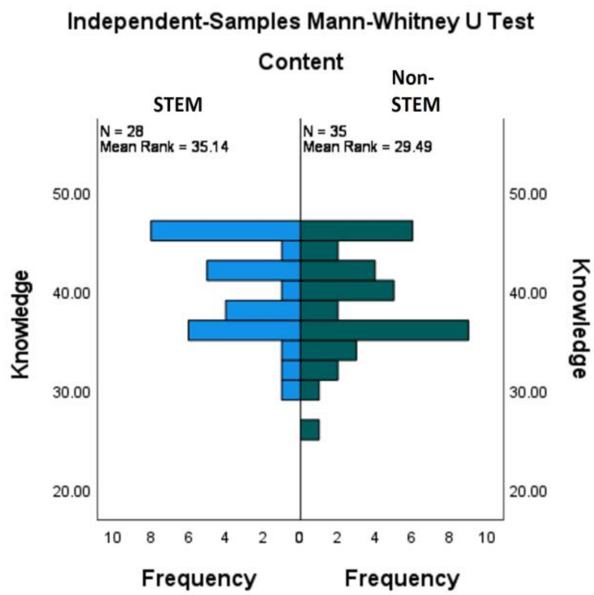


Table 5

Research Question 1: Hypothesis Test Summary

Null Hypothesis	Test	Sig	Decision
The distribution of knowledge is the same across categories of content taught.	Independent-samples Mann Whitney U test	.22	Retain the null hypothesis

Table 6

Research Question 1: Mann-Whitney U Test

Total *N*	63
Mann-Whitney U	402.00
Wilcoxon W	1,032.00
Test statistic	402.00
Standard error	71.33
Standardized test statistic	-1.23
Asymptotic sig (2-sided test)	0.217

The distribution of assessment knowledge scores between STEM and non-STEM faculty were not similar on the Mann-Whitney U. The mean rank was calculated using the Mann-Whitney U test. The mean rank scores are presented in Figure 2. The difference in means Assessment knowledge for STEM faculty (mean rank = 35.14) and non-STEM faculty (mean rank = 29.49) were not statistically significant, U= 402, z = -1.234, p = 0.217. There was no significant difference based on the statistics from the Mann-Whitney U nor the independent *t* test. Based on the lack of a statistically significant difference in mean scores the null hypothesis was retained.

Research Question 2

Research question 2 for the study was to what extent is there a significant difference in mean scores regarding assessment self-efficacy between 2-year college STEM faculty and 2-year college non-STEM faculty on the Yin Assessment Survey? For this question, there were two groups in the independent variable. The two groups were STEM and non-STEM faculty. The participants were classified into one of the two groups based on self-reported content area taught: STEM (*n* = 28) and non-STEM (*n* = 35). Data for the STEM group is shown as mean ± standard deviation (*n* = 28, 45.71 ± 7.60). Data for the STEM group is shown as mean ± standard deviation (*n* = 35, 44.32 ± 5.83). For the STEM group the minimum value of the variable was 31.00 and the maximum value 55.00. For the non-STEM group, the minimum value of the variable was 29.00 and the maximum value 55.00. The complete descriptive statistics are found in table 7.

Table 7

Research Question 2: Descriptive Statistics

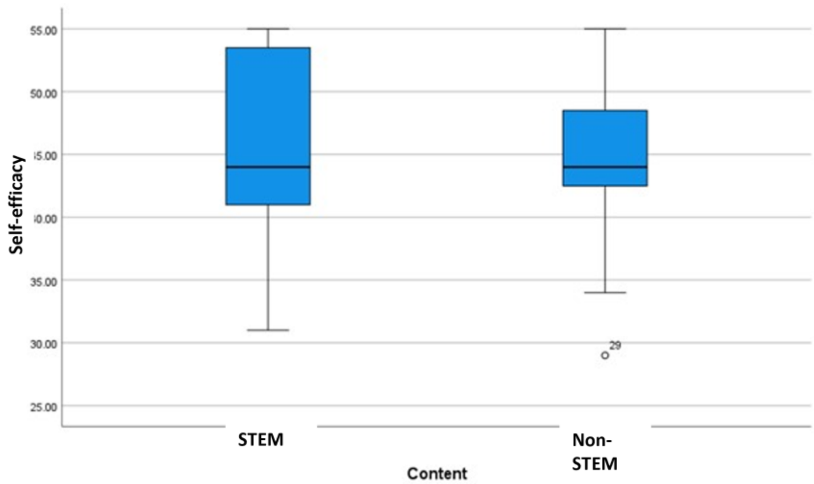
					95% confidence interval for mean			
	N	M	SD	Std. error	Lower bound	Upper bound	Minimum	Maximum
STEM	28	45.71	7.60	1.44	42.77	48.66	31.00	55.00
Non-STEM	35	44.63	5.83	0.99	42.63	46.63	29.00	55.00

As previously discussed in Research Question 1, the fourth assumption for the independent *t* test was the presence of no significant outliers in the data. A boxplot was examined to assess this assumption and shown in Figure 3. As before, the two groups representing the independent variable are located on the X axis labeled content. The independent variable represents the content area taught by the faculty. The Y axis is the total score for all of the questions in the dependent variable added together for each participant. The questions used for Research Question 2 are the last 11 questions on the

Yin Assessment survey. There was one outlier in the non-STEM faculty, Group 2 illustrated in the boxplot in Figure 3. The decision to retain the outlier was made and to proceed with the independent *t* test.

Figure 3

Research Question 2: Boxplot



The Shapiro-Wilkes test was conducted to assess the fifth assumption for the independent *t* test which is that the dependent variable is normally distributed for each of the groups represented in the independent variable. The Shapiro-Wilkes test results are presented in Table 8.

Table 8

Research Question 2: Test of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
STEM	0.92	28	0.03
Non-STEM	0.94	35	0.04

As earlier described for Research Question 1 there are six assumptions that needed to be met for the independent *t* test. The assessment and the breakdown of the assumptions for Research Questions 2 are provided below. The first three assumptions were met in the manner as described in Research Question 1.

Another assumption for the independent *t* test was that there should be no significant outliers. As with Research Question 1, there was one outlier in the non-STEM faculty, Group 2 illustrated in the boxplot in figure 3. The outlier in Research Question 2 is not the same outlier as in Research Question 1. As before in Research Question 1, the answers of the outliers were examined, and it was determined that there were no missed questions. As before, the answers tended to be lower than other faculty which could be the actual assessment self-efficacy of the participant. The decision to retain the outlier was made and to proceed to with the independent *t* test.

The next assumption was that dependent variable should be normally distributed for the groups in the independent variable. The Shapiro-Wilkes test was conducted for Research Question 2 and the results are presented in Table 6. The assessment self-efficacy scores for the STEM group were assessed by the Shapiro-Wilkes test ($p < 0.05$)

and were not normally distributed. The self-efficacy scores for the non-STEM group were also assessed using the Shapiro-Wilkes test ($p < 0.05$) and the results indicated that they were not normally distributed. The sample sizes were under 50 so a normal Q-Q plot was not used to assess distribution. Based on the violations for Research Questions 1 and the recommendation to conduct a nonparametric test, the data were not transformed. The option to proceed was selected. The *t* test was completed with the violation of the assumption reported.

The sixth assumption was that there is homogeneity of variances. Levene's test of homogeneity of variances was conducted and presented in Table 7. The assumption of homogeneity was violated, as assessed by Levene's test ($p < 0.05$). The *t* test was analyzed using the values for equal variances not assumed (second row of the *t* test). As with Research Question 1, because of the violations a nonparametric analysis was recommended.

Table 9

Research Question 2: Independent Samples Test

	Levene's test for equality of variances				*t*-test for Equality of means				95% confidence interval of the difference	
					Significance					
	F	Sig	*t*	*df*	One-Sided *p*	Two-Sided *p*	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	5.07	0.028	0.64	61	0.26	0.52	1.09	1.69	-2.29	4.47
Equal variances not assumed			0.62	49.67	0.27	0.54	1.09	1.74	-24	4.59

Due to the violation of homogeneity, the values used for analysis were obtained from the second row of the independent *t*-test (equal variances not assumed). Based on the independent *t* test there was no statistically significant difference in mean assessment self-efficacy scores between STEM and non-STEM faculty, t (50) = 0.62, p = 0.54. Based on the lack of a statistically significant difference in mean scores the null hypothesis was retained. The violations of assumptions are similar to Research Question 1. As with Research Question 1, a nonparametric Mann-Whitney U test was completed, and the results provided below.

As previously stated for Research Question 1, the Mann-Whitney U has four assumptions that must be met. There must be one dependent variable that is measured on a continuous scale. Additionally, there must be only one independent variable and it needs to consist of two independent groups. The last assumption is that there is independence of observations. These three assumptions were met as described in the results for Research Question 1. The fourth assumption is that the distribution of scores for both of the groups making up the independent variable must be determined. Based on visual interpretation, the distribution patterns are not similar, and the mean rank will be reported.

Figure 4

Research Question 2: Mann-Whitney U Test

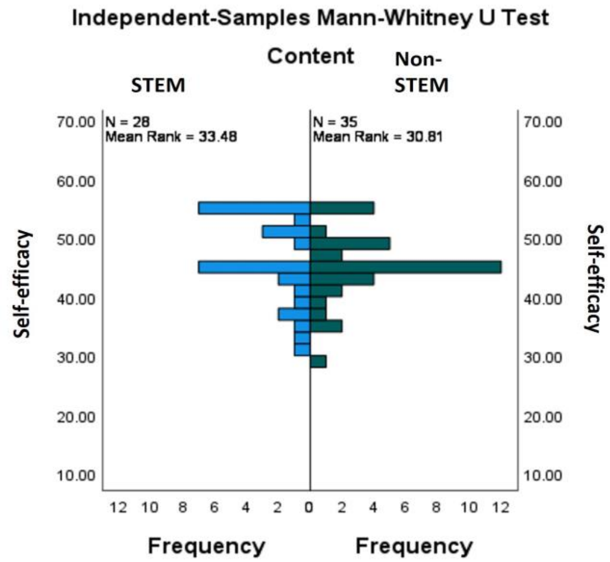


Table 10

Research Question 2: Hypothesis Test Summary

Null Hypothesis	Test	Sig	Decision
The distribution of knowledge is the same across categories of content taught.	Independent-samples Mann Whitney U test	0.56	Retain the null hypothesis

Table 11

Research Question 2: Mann-Whitney U Test

Total *N*	63
Mann-Whitney U	448.50
Wilcoxon W	1078.50
Test statistic	448.50
Standard error	71.62
Standardized test statistic	-579
Asymptotic sig, (2-sided test)	0.56

The distribution of assessment self-efficacy scores between STEM and non-STEM faculty were not similar. The mean rank was calculated using the Mann-Whitney U test. The mean rank scores are presented in Figure 4. The difference in means for assessment self-efficacy for STEM faculty (mean rank = 33.48) and non-STEM faculty (mean rank = 30.81) were not statistically significant, U = 448.5, z = -579, *p* = 0.562. Based on the lack of a statistically significant difference in mean scores from both the independent *t* test and the Mann-Whitney U test the null hypothesis was retained.

Group Data

Data from both the STEM and non-STEM faculty were analyzed as one group for each individual question. Although the variances of responses to the survey questions was not an original research question, I included the data because information from the total mean of the responses to the survey questions contributed to a potential future research direction for the researcher. The future direction is covered in the recommendation section of Chapter 5. The mean and standard deviation for the questions in the survey were determined using SPSS. Questions 3-8 were used for assessment knowledge, Questions 9-11 were also used for assessment knowledge, however they were specific to formative assessment knowledge. Questions 12-22 were used for assessment self-efficacy. The breakdown of means and standard deviation by question is provided.

Table 12

Mean by Question

Question Number	Question	*M*	*SD*
3	I know how to design effective tasks to assess student learning.	4.40	0.81
4	I can tell which assessment items will effectively assess student learning.	4.32	0.69
5	I can choose appropriate assessment tasks based on the purpose of the assessment.	4.41	0.71
6	I know whether an assessment item is aligned with a given standard.	4.44	0.59
7	I know what concept or skill an assessment item is testing.	4.52	0.56
8	I know how to improve existing assessments.	4.37	0.60
9	I am familiar with formative assessment strategies.	4.24	0.84
10	I know how to use formative assessments in teaching.	4.19	0.86
11	I have sufficient knowledge about formative assessment.	4.00	0.97
12	I can effectively implement formative assessment in my teaching.	4.13	0.79
13	I can use formative assessment to improve my student learning.	4.13	0.75
14	I can engage students in learning by using formative assessment.	4.06	0.72
15	I can use formative assessment to identify the gap between students' current understanding and the learning goals.	4.06	0.82
16	I am able to use formative assessment to figure out what students know and can do.	4.05	0.79
17	I can interpret the information from formative assessment.	4.13	0.71
18	I can figure out students learning needs by using formative assessment.	3.97	0.72
19	I can use information collected from formative assessment to guide my teaching.	4.14	0.64
20	I can take appropriate action to help students close the learning gap.	4.13	0.63
21	I can match instruction to the learning gap identified by formative assessment.	4.08	0.68
22	I can provide effective feedback to students when using formative assessment.	4.24	0.67

The Yin Assessment Survey utilized the following Likert scale: 1 *strongly disagree*, 2 *disagree*, 3 *neutral*, 4 *agree*, 5 *strongly agree*. Examination of the mean of each question shows that the average was in the agree range for all of the survey questions. In the survey, Questions 9-11 are specific to formative assessment knowledge. Although the three questions did not have the lowest means, they did have the larger standard deviations.

Summary

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment Survey at the 2-year college level. Changes to data collection were made based on the issues that arose with email distribution problems at the target institution related to policies in place based on phishing. This issue did contribute to an increase in the time to collect data and a need for a second email to be sent to participants.

An independent *t* test was completed for both of the research questions. Based on the violations of assumptions, a nonparametric test was also conducted. For Research Question 1, here is no significant difference in the mean assessment knowledge scores for STEM and non-STEM faculty. For Research Question 2, there is also no significant difference in the mean assessment self-efficacy scores for STEM and non-STEM faculty. Chapter 5 will include a discussion of the conclusions from the data presented in Chapter 4 of this study.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment Survey at the 2-year college level. The findings for Research Question 1 indicate that there was no statistically significant difference in mean assessment knowledge scores between STEM and non-STEM faculty. For Research Question 2, the findings show that there was no statistically significant difference in mean assessment self-efficacy scores between STEM and non-STEM faculty. Although not one of the research questions, the survey questions in the Yin Assessment Survey related to formative assessment knowledge as opposed to the general assessment questions contained some of the larger standard deviations.

My interpretation of the collected data is included in this chapter. The limitations related to the study are also detailed in the chapter. Finally, recommendations for future research associated with the study and the implications for positive social change are provided.

There was no statistically significant difference in mean assessment knowledge scores or assessment self-efficacy scores between STEM and non-STEM faculty. Despite the lack of a difference in mean scores, information from the study can be useful for continuing research in assessment in the 2-year college setting. The lack of statistically significant collected data could be based on the study design and population size. The Recommendations section in this chapter includes suggestions for improvement in study design to examine this question.

Interpretation of Findings in Relation to the Theoretical Framework

The framework I used for this study was the SDT. The SDT includes the need for a sense of competence (i.e., knowledge) for self-efficacy (Deci & Ryan, 2002). Results from the current study suggest that there is no significant difference in mean scores for assessment knowledge and self-efficacy between STEM and non-STEM faculty. There is no difference in the assessment knowledge between the groups suggesting both groups have a sense of competence. If both groups have the sense of competence with no differences, it is likely there would be no difference in mean assessment self-efficacy scores if knowledge is related to self-efficacy.

Many of the previous studies reviewed in Chapter 2 utilized training or interventions as a way to improve self-efficacy. An example is Zwart et al.'s (2015) study that used coaching as an intervention to improve self-efficacy (p. 591). Unlike Zwart et al., the current study did not utilize any training or intervention but measured mean scores for assessment self-efficacy for STEM and non-STEM faculty. The lack of an intervention or a training program means that results from the current study do not compare to those reported in the Chapter 2 literature that utilized the SDT as a framework.

Interpretation of Findings in Relation to the Empirical Literature

The results indicated that there is no difference between STEM and non-STEM faculty assessment knowledge. In the literature reviewed in Chapter 2, Rohrbacher (2017) reported that humanities professors were enthusiastic about assessment in teaching and learning (pp. 29-30). Guadu and Boersma (2018) concluded that the English faculty had positive perceptions of using writing as a formative assessment (pp. 48-49). In the STEM discipline, Keller (2017) determined that formative assessment was beneficial to students regarding summative assessment success (p. 16). The results from the current study suggest that the knowledge may not differ based on whether the faculty member is classified as STEM or non-STEM.

The literature did not contain many direct assessment comparison studies. One of the studies had reported in the literature that assessment methodologies differed between disciplines (Simpson, 2016, p. 933). The current study comparing STEM and non-STEM participants measured only knowledge and self-efficacy in assessment and not the assessment tools used. The STEM and non-STEM faculty may not have had a significant difference because they may feel they have the knowledge and self-efficacy for their current assessment tools. The current study was not designed to determine if they have the knowledge or self-efficacy for tools outside their current norm. A second study in the literature that included a STEM and non-STEM comparison was designed to compare the use of creative process skills (Daly et al., 2016, p. 4). I did not detect a difference between the two groups in the current study.

The second research question for the study addressed faculty assessment self-efficacy. Chapter 2 included examples of previous self-efficacy research. Previous researchers found a higher self-efficacy in aspects of teaching, such as online modalities, and alternative practices correlated with a higher satisfaction and persistence among the faculty (Gosselin, 2016, p. 189; Hardy et al., 2017, p. 55; Lai et al., 2018, pp. 45-46). The current study did not examine a relationship between self-efficacy and assessment but compared the assessment self-efficacy of two groups: STEM, and non-STEM faculty. There was no statistically significant difference in assessment self-efficacy between the two groups. As stated previously, the frame of reference for each faculty member was their current assessment knowledge and self-efficacy. The data suggests that STEM and non-STEM faculty do not have a difference in means scores for assessment knowledge or self-efficacy.

Limitations of the Study

One of the limitations of this study was a larger response from one of the independent variable groups when compared to the other. The non-STEM group had a higher response (n = 35) compared to the STEM response (n = 28). Additionally, the general low number of responses was a second limitation. Distribution of the survey from an email address outside of the study site institution resulted in it being initially flagged, and it did not get adequately distributed. The second email attempt was distributed but still flagged as an outside email address. The label as an outside email may have discouraged faculty from participating in the study due to the study site institution's technology training that emphasizes not to open links that the faculty member does not trust. The lower number of responses from the STEM group may have been partially the result of the email distribution problems. Technology is one of the disciplines in the STEM group, and these faculty may have been more sensitive and concerned by the potential for the outside email address to be phishing.

Another limitation that influenced this study was the COVID-19 pandemic. Ideally, the best way to implement and collect survey information would have been to attend department meetings and physically collect the results. Due to the COVID-19 pandemic, the departments were all holding their meetings virtually, which prevented the physical collection of data. Although faculty had returned to the classroom, meetings were still held virtually. Additionally, the COVID-19 pandemic could have been a factor in how faculty answered the survey. The changes in teaching modalities as a result of the pandemic could have led to faculty teaching in a modality that is not their normal format. Although, at the time of the survey faculty had returned to the classroom, the changes and disruption to the normal education environment might have influenced their answers.

Recommendations

The recommendations included in this section are based on my experiences from the data collection process. Recommendations are also based on the limitations reported for the study. Due to the low sample size, my first recommendation would be to conduct this survey in a physical setting to improve the overall response rate and use a larger sample size that might allow for statistically significant results. Additionally, the in-person data collection method might increase the number of participants that were not motivated to complete the email survey. The inclusion of faculty with less of an interest in educational research could result in more a representative sample of 2-year college faculty. A second option would be to conduct this study using 2-year college faculty in general and not restrict the study to one geographic area or institution to get a larger and more representative sample.

Another consideration is that a larger sample size could allow the researcher to report out by discipline taught and not use general variables, such as STEM and non-

STEM. Future researchers might also consider collecting more demographic information from the participants. Factors such as length of time teaching and the completion of teacher training programs could influence the perceptions and answers of the individual faculty members.

Another recommendation would be to collect more demographic information. I did not collect any additional demographic information other than content area taught in this study. Additionally, information about the faculty members' background could be a factor in their assessment knowledge and self-efficacy. Some of the faculty taught PK-12 before becoming college faculty. This could indicate that they took courses to be a licensed teacher that would have included assessment training.

My final recommendation is to consider conducting the research using an assessment training program as an intervention. The Yin Assessment Survey could be used to measure responses pre- and postintervention scores after a professional development course (see Yin et al., 2015). The survey instrument used in this study was originally designed to examine perceptions pre- and postintervention. STEM and non-STEM faculty assessment knowledge and self-efficacy could be examined prior to the intervention to establish any baseline differences between the two groups. The postintervention scores could also be compared to determine if the two groups both benefited equally from the intervention.

Some of the interesting data points noted for this study were when comparing the means for each question for the entire group, I found the larger standard deviations were associated with formative assessment knowledge. An area that could warrant further investigation is faculty knowledge specific to formative assessment. If this is an area where faculty have diverse knowledge, assessment strategies might be improved upon by intervention through training. Assessment knowledge and self-efficacy in 2-year college faculty is an area where future research could be beneficial for improving the success of students.

Implications for Positive Social Change

The original goal of this study was to determine if STEM faculty had a difference in assessment knowledge and self-efficacy. The positive social change that could have resulted from this information is the potential to find an area in STEM education that could be improved upon to increase student success. An improvement in student success would not only be beneficial to the students and the study site institution, but it could be beneficial to society by increasing the amount of educated people available to the work force. Based on the lack of statistically significant data, I did not detect a difference in assessment knowledge and self-efficacy between STEM and non-STEM faculty in this study. The failure to reject the null hypotheses does suggest that assessment knowledge and self-efficacy of 2-year college faculty may not be a target area to improve STEM success. Although based on the small sample size and geographic isolation, this could be reassessed on a larger scale. Future studies could result in information that could be used for improving the assessment knowledge and self-efficacy of 2-year college faculty. This improvement in assessment knowledge and self-efficacy could result in more student success and increased student completion of 2-year college programs.

Conclusion

The purpose of this quantitative study was to determine the extent of differences in STEM and non-STEM instructor assessment knowledge and self-efficacy as measured by the Yin Assessment Survey at the 2-year college level. The results from the survey did not show a significant difference in mean assessment knowledge or self-efficacy scores between the STEM and non-STEM participants in the study. However, limitations of the study, including a small sample size and geographic isolation, might require additional research to definitively determine the lack of a significant difference. Examination of the mean scores and standard deviation from the individual questions on the survey showed that the most variance in responses was present in the questions that were specifically related to formative assessment knowledge. This data may suggest that faculty in the 2-year college setting might benefit from further training in formative assessment.

Two-year, open enrollment colleges often face challenges regarding student success and retention (Pratt, 2017). The success and retention are often lower in STEM courses (Soricone & Endel, 2019). I designed this study to examine one aspect that influences student success and retention. Part of the rationale was that assessment differences in the disciplines could be related to student success and retention. The mean difference in assessment knowledge or assessment self-efficacy scores between the STEM and non-STEM could be used to illustrate the variances in assessment that could relate to student success. However, I found no mean differences between STEM and non-STEM assessment knowledge or assessment self-efficacy scores, but there remains a need for improving the educational process in the 2-year college setting.

References

Abraham, R. R., Alele, F., Kamath, U., Kurien, A., Kiranmai, S. R., Bairy, I., Rao, M., Rao, G., Devi, V., Rao, Y., Gupta, T., & Malau-Aduli, B. (2018). Assessment for learning: A needs analysis study using formative assessment to evaluate the need for curriculum reform in basic sciences. *Advances in Physiology Education 42*, 482-486. https://doi.org/10.1152/advan.00093.2018

Akkaraju, S., Atamturktur, S., Broughton, L., & Frazer, T. (2019). Ensuring student success: Using formative assessment as the key to communication and compassion among faculty, student, and staff. *New Directions for Community Colleges 2019*, (186), 71-79. https://doi.org/10.1002/cc.20358

Alotaibi, K. A. (2019). Teachers' perceptions on factors influencing adoption of formative assessment. *Journal of Education and Learning 8*(1), 74-86 https://doi.org/10.5539/jel.v8n1p74

Baiduc, R. R., Linsenmeier, R. A., & Ruggeri, N. (2016). Mentored discussions of teaching: An introductory teaching development program for future STEM faculty. *Innovative Higher Education, 41*(3), 237-254. https://doi.org/10.1007/s10755-015-9348-1

Bandura, A. (1997). *Self-efficacy the exercise of control*. W.H. Freeman and Company.

Barton, G. M., Baguley, M., Kerby, M., & MacDonald, A. (2020). Investigating the assessment practices within an initial teacher education program in an Australian university: Staff perceptions and practices. *Australian Journal of Teacher Education, 45*(3), 34-47. https://doi.org/10.14221/ajte.2020v45n3.3

Benken, B. M., Ramirez, J., Li, X., & Wetendorf, S. (2015). Developmental mathematics Success: Impact of Stud4ents' Knowledge and Attitudes. *Journal of Developmental Education*, *38*(2), 14-22.

Bruck, L. B., & Bruck, A. D. (2018). Two-year community: Tools for success- A study of the resources and study habits of General Chemistry 1 students at two community colleges. *Journal of College Science Teaching*, *47*(3), 9-17. https://doi.org/10.2505/4/jcst18_047_03_9

Cohen, R., & Kelly, A. M. (2019). The impact of community college science and mathematics course taking on graduation, transfer, and non-completion. *The Review of Higher Education*, *42*(2), 595-617. https://doi.org/10.1353/rhe.2019.0008

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology class more equitable? *PLOS One*, *12*(12), 1-11. https://doi.org/10.1371/journal.pone.0189610

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative and mixed methods approaches. SAGE

Daly, S. R., Mosyjowski, E. A., Oprea, S. L., Huang-Saad, A., & Seifert, C. M. (2016). College students' views of creative process instruction across disciplines. *Thinking Skills and Creativity*, *22*(2016), 1-13. https://doi.org/10.1016/j.tsc.2016.07.002

Daou, B. M. N. (2016). The effect of using computer skills on teachers perceived self-efficacy beliefs towards technology integration, attitudes, and performance. *World*

Journal on Educational Technology, 8(3), 294-306. https://doi.org10.18844/WJET.813.1118

De, S., & Arguello, G. (2020). STEM education in college: An analysis of stakeholders' recent challenges and potential solutions. *FDLA Journal*, *5*(2020), 1-13.

Deci, E. L. (1995). *Why we do what we do.* Penguin Books.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. The University of Rochester Press.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology, 49*(3), 182-185. https://doi.org/10.1037/a0012801

Deci, E. L., & Ryan, R. M. (2016). Optimizing students' motivation in the era of testing and pressure: A self-determination theory perspective. In W. Liu, J. Wang, & R. Ryan (Eds.), *Building autonomous learners* (pp.9-29). Springer.

Doerschuk, P., Bahrim, C., Daniel, J., Kruger, J., Mann, J., & Martin, C. (2016). Closing the gaps and filling the STEM pipeline: A multidisciplinary approach. *Journal of Science Education and Technology, 25*(4), 682-695. https://doi.org/10.1007/s10956-016-9622-8

Dolin, J. & Evans, R. (2018). *Transforming assessment.* Springer International Publishing.

Eather, N., Riley, N., Miller, D., & Imig, S. (2019). Evaluating the impact of two dialogical feedback methods for improving pre-service teacher's perceived confidence and competence to teach physical education within the authentic learning environments. *Journal of Education and Training Studies, 7*(8), 32-46. https://doi.org/10.11114/jets.v7i8.4053

Frankfort-Nachmias, C., & Nachmias, D. (2008). *Research methods in the social sciences*. Worth Publishers.

Gaze, E. (2018). Quantitative reasoning: A guided pathway from two- to four-year colleges. *Scholar Commons, 11*(1), 1-4. https://doi.org/10.5038/1936-466.11.1.1

Geert van de Berg, R. (n.d.). *Effect size - A quick guide.* SPSS Tutorials. https://www.spss-tutorials.com/effect-size/

Gosselin, K. P. (2016). Development of an evidence-based professional learning program informed by online teachers' self-efficacy and threshold concepts. *Online Learning, 20*(3), 178-194.

Gray, J. S., Brown, M. A., & Connolly, J. P. (2017). Examining construct validity of the quantitative literacy VALUE rubric in college-level STEM assignments. *Research and Practice in Assessment, Summer*, 20-31.

Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research, 9*(1), 42-50. https://doi.org/10.17507/jltr.0901.06

Hanauer, D. I., & Baruerle, C. (2012). Facilitating innovation in science education through assessment reform. *Liberal Education*, *98*(3), 34-41. http://www.aacu.org/

Hardy, P., Shepard, M., & Pilotti, M. (2017). Does part-time faculty's self-efficacy predict critical dimensions of online college teaching? *College Teaching*, *65*(2), 50-57. https://doi.org/10.1080/87567555.2016.1232692

Hodgins, H. S., & Knee, C. R. (2002). The integrating self and conscious experiences. In Deci, E. l. & Ryan, E. M. (Eds), *Handbook of self-determination research* (pp. 87-100). University of Rochester Press.

Horn, A. S., Horner, O. G., & Lee, G. (2019). Measuring the effectiveness of two-year colleges: A comparison of raw and value-added performance indicators. *Studies in Higher Education*, *44*(1), 151-169. https://doi.org/10.1080/03075079.2017.1349741

Horvitz, B. S., Beach, A. L., Anderson, M. L., & Xia, J. (2014). Examination of faculty self-efficacy related to teaching online. *Innovation in Higher Education*, *40*(2015), 305-316. https://doi.org/10.1007/s10755-014-9316-1

Hu, P., & Zhang, J. (2017). A pathway to learner autonomy: A self-determination theory perspective. *Asian Pacific Education Review*, *18*(2017), 147-157. https://doi.org/10.1007/s12564-016-9468-z

Jaafar, R., Toce, A., & Polnariev, B. A. (2016). A multidimensional approach to overcoming challenges in leading community college math tutoring success.

Community College Journal of Research & Practice, *40*(6), 534-549. https://doi.org/10.1080/10668926.2015.1021406

Karaarslan, G., Ertepinar, H., & Sungur, S. (2013). Use of self-determination theory to support basic psychological needs of preservice teachers in an environmental science course. *Environmental Education Research, 19*(3), 342-369. https://doi.org/10.1080/13504622.2012.695013

Kaygisiz, S., Anagun, S. S., & Karahan, E. (2018). The predictive relationship between self-efficacy levels of English teachers and language teaching methods. *Euroasian Journal of Educational Research, 78*, 183-201.

Keller, C. (2017). Using formative assessment to improve microscope skills among urban community college general biology I lab students. *Journal of College Science Teaching*, *46*(3), 11-18. https://doi.org/10.2505/4/4jcst17_046_03_11

Kim, Y. A., Cox, J., Southard, K. M., Elfring, L., Blowers, P., & Talanquer, V. (2019). Learning Researchers: promoting formative assessment in STEM courses. *Journal of College Science Teaching 48*(5), 36-41.

Koretsky, M. D., Brooks, B. J., & Higgins, A. Z. (2016). Written justifications to multiple-choice concept questions during active learning in class. *International Journal of Sciences Education 38*(11), 1747-1765. https://doi.org/10.1080/09500693.2016.1214303

Krulatz, A. (2017). Promoting student success through active involvement in assessment. *College Teaching*, *65*(4), 207-208. https://doi.org/10.1080/87567555.2016.12049709

Laerd Statistics. (2020). *Independent samples t-test*. https://statistics.laerd.com/premium/stata/istt/independent-*t*-test-in-stata.php

Lai, H. M., Hsiao, Y. L., & Hsieh, P. J. (2018). The role of motivation, ability, and opportunity in university teacher's continuance use intention for flipped teaching. *Computers and Education, 124*(2018), 37-50. https://doi.org/10.1016/j.compedu.2018.05.013

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., & Caballero, M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *Plos ONE, 11*(9), 1-21. https://doi.org/10.1371/journal.pone.0162333

Lawrie, G. A., Gahan, L. R., Matthews, K. E., Weaver, G. C., Bailey, C., Adams, P., & Taylor, M. (2014). Technology supported facilitation and assessment of small group collaborative inquiry learning in large first-year classes. *Journal of Learning Design, 7*(2), 120-135. http://doi.org/10.5204/jld.v7i2.204

Lysne, S. J., & Miller, B. G. (2017). A comparison of long-term knowledge between two teaching approaches. *Journal of College Science Teaching, 46*(6), 100-107. https://doi.org/10.2505/4/jcst17_046_06_100

Malcom, S., & Feder, M. (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees: Systematic change to support students' diverse pathways*. National Academies Press.

Nortvedt, G. A., & Siqveland, A. (2018). Are beginning calculus and engineering students adequately prepared for higher education? An assessment of student's basic mathematical knowledge. *International Journal of Mathematical Education in Science and Technology*, *50*(3), 325-343. https://doi.org/10.1080/0020739X.2018.1501826

Offerdahl, E. G., McConnell, M., & Boyer, J. (2018). Can I have your recipe? Using a fidelity of implementation (FOI) framework to identify the key ingredients of formative assessment for learning. *CBE-Life Sciences Education, 17*(4), 1-9. https://doi.org/10.1187/cbe.18-02-0029

O'Gara, L., Mechur Karp, M., & Hughes, K. L. (2008). Student success courses in the community college: An exploratory study of student perspectives. *Community College Review*, *36*(3), 195-218. https://doi.org/10.1177/0091552108327186

Pratt, T. (2017). The open access dilemma: How can community colleges better serve underrepresented students?. *Education Next, Fall* 17, 34- 41. http://educationnext.org/

Rawn, C. D., & Fox, J. A. (2018). Understanding the work and perceptions of teaching focused faculty in a changing academic landscape. *Research in Higher Education*, *59*(5), 591-622. http://doi.org/1007/s11162-017-9479-6

Reeves, T. D., Hake, L. E., Chen, X., Fredrick, J., Rudenga, K., & Ludlow, L. H. (2018). Does context matter? Convergent and divergent findings in the cross-institutional evaluation of graduate teaching assistant professional development programs. *CBE Life Sciences Education, 17*(1), 1-13. https://doi.org/10.1187/cbe.17-03-0044

Rodriguez, M., Mundy, M. A., Kupczynski, L., & Challoo, L. (2018). Effects of teaching strategies on student success, persistence, and perceptions of course evaluations. *Research in Higher Education Journal*, *35*, 1-21.

Rohrbacher, C. (2017). Humanities professors' conceptions of assessment: A case study. *The Journal of General Education*, *66*(1-2), 17-41. https://doi.org/10.5325/jgeneeduc.6.1-2.0017

Sanchez-Olivia, D., Pulido-Gonzalez, J. J., Leo, F. M., Gonzalez-Ponce, I., & Garcia-Calvo, T. (2017). Effects of an intervention with teachers in the physical education context: A self-determination theory approach. *PLOS One*, *12*(2), 1-17. https://doi.org/10.1371/journal.pone.0189986

Schwartz, M. S., Hazari, Z., & Sadler, P. M. (2008). Divergent views: Teacher and professor perceptions about pre-college factors that influence college science success. *Science Educator*, *17*(1), 18-35.

Scott, A. N., McNair, D. E., Lucas, J. C., & Land, K. M. (2017). From gatekeeper to gateway: Improving student success in an introductory biology course. *Journal of College Science Teaching*, *46*(4), 93-99.

Shaffer, S. C., Eshbach, B. E., & Santiago-Blay, J. A. (2015). A dual approach to fostering under-prepared student success: Focusing on doing and becoming. *Insight: A Journal of Scholarly Teaching*, *10*79-91.

Shana, Z., & Alwaely, S. (2021). Does the flipped classroom boost student science learning and satisfaction? A pilot study from the UAE. *International Journal of Instruction*, *14*(4), 607-626. https://doi.org/10.29333/iji.2021.14435a

Shapiro, D., Dunda, A., Huie, F., Wakhunga, P. K., Yuan, X., Nathan, A., & National Student Clearing House Research, C. (2017). *A national view of student attainment rates by race and ethnicity- Fall 2010 cohort (*signature report no. 12B). National Student Clearing House. http://www.studentclearinghouse.org

Siegel, M. A., Roberts, T. M., Freyermuth, S. K., Witzig, S. B., & Izci, K. (2015). Aligning assessment to instruction: Collaborative group testing in large enrollment science classes. *Journal of College Science Teaching*, *44*(6), 74-82.

Simpson, A. (2016). Assessment and its outcomes: The influence of disciplines and institutions. *Assessment and Evaluation in Higher Education*, *41*(6), 917-937. https://doi.org/10.1080/02602938.2015.1052369

Sit, S. M., & Brudzinski, M. R. (2017). Creation and assessment of an active e-learning introductory geology course. *Journal of Science Education Technology*, *26*, 629-645. https://doi.org/10.1007/s10956-017-9703-3

Soricone, L., & Endel, B. (2019). Nudging to STEM success: Implementation report. *The Hemsley Charitable Trust*.

Stephan, J. L., Davis, E., Lindsay, J., & Miller, S. (2015). Who will succeed and who will struggle? Predicting early college success with Indiana's student information system. REL 2015-078. *Regional Educational Laboratory Midwest.*

Suwito, Budijanto, Handoyo, B., & Susilo, S. (2020). The effects of 5E learning cycle assisted with spatial based population geography textbook on students' achievement. *International Journal of Instruction*, *13*(1), 315-324. https://doi.org/10.29333/iji.2020.13121a

Taylor, J. L., & Giani, M. (2019). Modeling the effect of reverse credit transfer associate's degree: Evidence from two states. *The Review of Higher Education, 42*(2), 427-455. https://doi.org/10.1353/rhe.2019.0002

Torres, V., Hagedorn, L. S., & Heacock, L. T. (2018). Closing the academic and equity gaps: How achieving the dream redefined assessment. *New Directions for Institutional Research, 177,* 73-86. https://doi.org/10.1002/ir.20257

Varsavsky, C., & Rayner, G. (2013). Strategies that challenge: Exploring the use of differentiated assessment to challenge high-achieving students in large enrollment undergraduate cohorts. *Assessment and Evaluation in Higher Education, 38*(7), 789-802. http://doi.org/10.1080/02602935.2012.714739

Velthuis, C., Fisser, P., & Pieters, J. (2014). Teacher training and preservice primary teachers' self-efficacy for teaching. *Journal of Science Teacher Education, 25*(2014), 445-464. http://doi.org/10.1007/s10972-013-9363-y

Westat, J. A., Merlino, F. J., & Stephenson, K. (2015). The call to transform postsecondary STEM educational practices and institutional policies. *American Journal of Educational Studies, 7*(1), 27-42.

White, E., & Shakibnia, A. F. (2019). State of STEM: Defining the landscape to determine high-impact pathways for the future workforce. *Proceedings of the Interdisciplinary STEM teaching and Learning Conference*, *3*(4), 22-56. https://doi.org/10.20429/stem.2019.030104

Wright, A. M., & Meyer, K. R. (2017). Exploring the relationship between students needing accommodations and instructor self-efficacy in complying with accommodations. *Higher Learning Research Communications, 7*(1), 65-83.

Yin, Y., Olson, J., Olson, M., Solvin, H., & Brandon, P. R. (2015). Comparing two versions of professional development for teachers using formative assessment in networked mathematics classrooms, *Journal of Research on Technology in Education*, *47*(1), 41-70, https://doi.org/10.1080/15391523.2015.967560

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., & Young, D. B. (2008). On the impact of formative assessment on students' motivation, achievement, and conceptual change. *Applied Measurement in Education*, *21*(4), 335–359. https://doi.org/10.1080/08957340802347845

Zwart, R. C., Korthagen, F. A. J., & Attema-Noordewier, S. (2015). A strength-based approach to teacher professional development. *Professional Development in Education*, *41*(3), 579-596. https://doi.org/10.1080/19415257.2014.919341

Appendix A: Demographic Survey

Demographics regarding Experience

1. Please select the content area that best describes your current teaching assignment.

    A. Art
    B. English
    C. Engineering
    D. Foreign Language
    E. History
    F. Humanities
    G. Mathematics
    H. Science
    I. Social Science
    J. Student Life Skills
    K. Technology
    L. Other, please describe.

Appendix B: Permission to Use Survey

Re: Comparing two versions of professional development for teachers using formative assessment in networked mathematics classrooms

Yue Yin <XXXXXXXX>

Sun 4/26/2020 9:30 PM

To:

- Karen Braley

Hi Karen,

Sure, feel free to use the survey. Thank you for your interest. I left the University of Hawaii before the team finished the project, so I did not continue to work on the project after I published that paper. You may search for other coauthors on the paper to see whether they published or presented other papers.

All the best,

Yue

On Sun, Apr 26, 2020 at 12:42 PM Karen Braley < XXXXXXXX> wrote:

Dr. Yin,

I am writing regarding your paper Comparing two versions of professional development for teachers using formative assessment in networked mathematics classrooms. I am interested in the survey focused on assessment located in appendix E. I am completing a dissertation in assessment and faculty self-efficacy and would like to use your survey. Of course, I am willing to credit your work in any publications that could be the result of my dissertation.

I am a graduate student at Walden University, I am working on a doctorate in educational assessment. I have an MS in Pharmacology from Dartmouth College. I am also currently a professor of science at XXXXXXX in Florida. My goal after completing my dissertation is to work with professors in STEM to improve the use of assessment data and explore diverse assessment methodology. I noticed that you also have a science background and wanted to share my goals with you.

If you are willing to allow me to use the survey, I would be grateful. I also wanted to inquire if you knew of any publications that might have followed up on your 2015 work. Thank you for your time and consideration.

Karen Braley